Kennzahlen nach § 48a SGB II
Übersicht zur Datenimputation

Version 5.2

Bundesagentur für Arbeit
Statistik

Impressum

Titel: Kennzahlen nach § 48a SGB II,
Übersicht zur Datenimputation

Veröffentlichung: 15.05.2020

Version: 5.2

Download: https://www.sgb2.info/ > Kennzahlen > Hilfe und Erläuterungen > Materialien und Downloads

Herausgeberin: Bundesagentur für Arbeit
Statistik/Arbeitsmarktberichterstattung

Rückfragen an: Zentraler Statistik-Service
Regensburger Straße 104

90478 Nürnberg

E-Mail: Zentraler-Statistik-Service@arbeitsagentur.de

Telefon: 0911 179-3632

Fax: 0911 179-1131

Weiterführende statistische Informationen:

Internet: https://statistik.arbeitsagentur.de

Zitierhinweis: Statistik der Bundesagentur für Arbeit,

Kennzahlen nach § 48a SGB II,

Übersicht zur Datenimputation, Nürnberg, Version 5.2

Nutzungsbedingungen: © Statistik der Bundesagentur für Arbeit

Sie können Informationen speichern, (auch auszugsweise) mit Quellenangabe weitergeben, vervielfältigen und verbreiten. Die Inhalte dürfen nicht verändert oder verfälscht werden. Eigene Berechnungen sind erlaubt, jedoch als solche kenntlich zu machen.

Im Falle einer Zugänglichmachung im Internet soll dies in Form einer Verlinkung auf die Homepage der Statistik der Bundesagentur für Arbeit erfolgen.

Die Nutzung der Inhalte für gewerbliche Zwecke, ausgenommen Presse, Rundfunk und Fernsehen und wissenschaftliche Publikationen, bedarf der Genehmigung durch die Statistik der Bundesagentur für Arbeit.

Änderungshistorie

Tabelle 1: Änderungshistorie

Datum	Version	Änderungen
August 2012	1.0	Erstellung
		Imputationsverfahren der ersten Entwicklungsstufe, vgl. Abschnitt 1.5
Mai 2013	2.0	Imputationsverfahren der zweiten Entwicklungsstufe, vgl. Abschnitt 1.5
		Neuer Standardvergleich: für die einheitliche quantitative Beurteilung wurden die Imputationsverfahren anfänglich nur mit dem gleitenden Mittelwert aus den letzten drei Monatswerten verglichen. Nun erfolgt der Standardvergleich mit einem neuen Set an Referenzmodellen: (1) Vormonatswertersetzung, (2) Interpolationsverfahren und (3) Fortschreibung anhand der SGB-II-Typisierung, vgl. Abschnitt 2.1.3
Mai 2014	3.0	Imputationsverfahren der dritten und letzten Entwicklungsstufe, vgl. Abschnitt 1.5
Nov 2014	3.1	Redaktionelle Anpassungen infolge der revidierten Zugangsmessung für Langzeitleistungsbeziehenden ab September 2014, vgl. Fußnote 122 in Anhang 0
Feb 2016	4.0	Ergänzung um Kapitel 7, Analysen zu Mehrfachausfällen
Mai 2016	5.0	Redaktionelle Anpassungen infolge der Revision der Grundsicherungsstatistik SGB II ab April 2016 (vgl. insb. Abschnitt 2.1.1)
Mai 2017	5.1	Hinweis zu Auswirkungen des 9. Änderungsgesetzes SGB II zur vermittlerischen Betreuung von Aufstocker/innen im SGB III auf die Imputation von Integrationen, Imputationen von LZB und Integrationen von Alleinerziehenden (vgl. insb. Fußnote 74 im Abschnitt 6.1.2)
Mai 2020	5.2	Anpassungen infolge der geänderten Rechtsverordnung zu Kennzahlen nach § 48a SGB II (15. März 2019):
		• Neufassung von Abschnitt 6.7: Für die geänderten Zähler-Grunddaten der K2E3 „kontinuierliche Beschäftigungen nach Integration“ wurde ein neues Imputationsverfahren entwickelt und implementiert. • Hieraus ergaben sich weitere redaktionelle Anpassungen über das gesamte Dokument hinweg.
		• Ersetzung des bisherigen Begriffs „Langzeitleistungsbezieher“ durch den neuen Begriff „Langzeitleistungsbeziehende“ sowie weitere geschlechtergerechte Formulierung über das gesamte Dokument hinweg.
		• Anpassungen in Abschnitt 4.4 infolge geänderter Definition der „aktivierten LZB“ im Sinne der K3E2.
		• Anpassungen in Abschnitt 6.5 infolge geänderter Definition von „öffentlich geförderten Beschäftigungen“ im Sinne der K2E2.
		Weitere redaktionelle Anpassungen über das gesamte Dokument hinweg, u. a. Aufnahme eines Textbausteins zu Abweichungen vom üblichen Muster bei der Geschlechterschätzung in Abschnitt 2.3.

Inhaltsverzeichnis

1 Einleitung 14

1.1 Hintergründe für fehlende Werte bei den Kennzahlen nach § 48a SGB II 14
1.2 Konsequenzen aus fehlenden Werten ohne Datenimputation 14
1.3 Datenimputationsbedarf 15
1.4 Ansätze für Datenimputationen 16
1.5 Stufenweise Entwicklung von Imputationsverfahren 17
1.6 Kurz-Charakterisierung aller verwendeten Datenimputationsverfahren 21

1.6.1 Fortschreibungsverfahren 21
1.6.2 Rückgriff auf Lieferinformationen im zeitlichen Umfeld 23
1.6.3 Bestandszahlenbasiertes Rekonstruktionsverfahren 24

2 Einheitliche Vorgehensweisen bei den Schätzungen und bei deren Bewertung 26

2.1 Vergleich und Bewertung der Imputationsverfahren 26

2.1.1 Datengrundlage und Stabilität der Ergebnisse 26
2.1.2 Vorgehen bei der Simulation und Bewertungskriterien 27
2.1.3 Standardvergleich und Standardvergleichsverfahren 28

2.2 Verwendung der SGB-II-Typisierung 30
2.3 Schätzung fehlender Werte nach Geschlecht 31

3 Datenimputation zum Themenbereich „Leistungen“ 34

3.1 Leistungen zum Lebensunterhalt 34

3.1.1 Analyse fehlender Werte 34
3.1.2 Schätzmodell 35
3.1.3 Bewertung 37
3.1.4 Fazit, Ausblick 38

3.2 Leistungen für Unterkunft und Heizung 39

3.2.1 Analyse fehlender Werte 39
3.2.2 Schätzmodell 39
3.2.3 Bewertung 41
3.2.4 Fazit, Ausblick 42

4 Datenimputation zum Themenbereich „Bestandsdaten“ 43

4.1 Bestand an erwerbsfähigen Leistungsberechtigten 43

4.1.1 Analyse fehlender Werte 43
4.1.2 Schätzmodell 44

4.1.3 Bewertung 45
4.1.4 Fazit, Ausblick 47
4.2 Bestand an alleinerziehenden erwerbsfähigen Leistungsberechtigten 47
4.2.1 Analyse fehlender Werte 47
4.2.2 Schätzmodell 47
4.2.3 Bewertung 48
4.2.4 Fazit, Ausblick 50
4.3 Bestand an Langzeitleistungsbeziehenden 50
4.3.1 Analyse fehlender Werte 50
4.3.2 Schätzmodell 52
4.3.2.1 Schätzmodell bei isolierten Datenlücken 52
4.3.2.2 Schätzmodell bei Langzeitlücken 53
4.3.3 Bewertung 54
4.3.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken 54
4.3.3.2 Bewertung des Schätzmodells bei Langzeitlücken 56
4.3.4 Fazit, Ausblick 56
4.4 Bestand an aktivierten Langzeitleistungsbeziehenden 57
4.4.1 Analyse fehlender Werte 57
4.4.2 Schätzmodell 58
4.4.2.1 Schätzmodell bei isolierten Datenlücken 58
4.4.2.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren 59
4.4.2.1.2 Variante b) Fortschreibung mit Hilfe der Deutschlandentwicklung 61
4.4.2.2 Schätzmodell bei Langzeitlücken 62
4.4.3 Bewertung 62
4.4.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken 62
4.4.3.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren 62
4.4.3.1.2 Variante b) Fortschreibung mit Hilfe der Deutschlandentwicklung 63
4.4.3.2 Bewertung des Schätzmodells bei Langzeitlücken 65
4.4.4 Fazit, Ausblick 65
5 Datenimputation zum Themenbereich „Bewegungsdaten“ 66
5.1 Zu- und Abgänge von erwerbsfähigen Leistungsberechtigten 66

5.1.1 Analyse fehlender Werte 66
5.1.2 Schätzmodell 67
5.1.3 Bewertung 68
5.1.4 Fazit, Ausblick 70
5.2 Zu- und Abgänge von Langzeitleistungsbeziehenden 71
5.2.1 Analyse fehlender Werte 71
5.2.2 Schätzmodell 72
5.2.3 Bewertung 73
5.2.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken 73
5.2.3.2 Bewertung des Schätzmodells bei Langzeitlücken 75
5.2.4 Fazit, Ausblick 77
6 Datenimputation zum Themenbereich „Beschäftigungsaufnahmen“ 78
6.1 Integrationen von erwerbsfähigen Leistungsberechtigten 78
6.1.1 Analyse fehlender Werte 78
6.1.2 Schätzmodell 79
6.1.3 Bewertung 80
6.1.4 Fazit, Ausblick 82
6.2 Integrationen von alleinerziehenden erwerbsfähigen Leistungsberechtigten 82
6.2.1 Analyse fehlender Werte 82
6.2.2 Schätzmodell 83
6.2.3 Bewertung 85
6.2.4 Fazit, Ausblick 87
6.3 Integrationen von Langzeitleistungsbeziehenden 87
6.3.1 Analyse fehlender Werte 87
6.3.2 Schätzmodell 88
6.3.2.1 Schätzmodell bei isolierten Datenlücken 89
6.3.2.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren 89
6.3.2.1.2 Variante b) Regressionsbasiertes Schätzverfahren 90
6.3.2.2 Schätzmodell bei Langzeitlücken 91
6.3.3 Bewertung 91
6.3.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken 91
6.3.3.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren 91
6.3.3.1.2 Variante b) Regressionsbasiertes Schätzverfahren 91

6.3.3.2 Bewertung des Schätzmodells bei Langzeitlücken 93
6.3.4 Fazit, Ausblick 93
6.4 Eintritte in geringfügige Beschäftigung 94
6.4.1 Analyse fehlender Werte 94
6.4.2 Schätzmodell 95
6.4.3 Bewertung 95
6.4.4 Fazit, Ausblick 97
6.5 Eintritte in öffentlich geförderte Beschäftigung 97
6.5.1 Analyse fehlender Werte 97
6.5.2 Schätzmodell 98
6.5.3 Bewertung 99
6.5.4 Fazit, Ausblick 100
6.6 Integrationen in sozialversicherungspflichtige Beschäftigung 101
6.6.1 Analyse fehlender Werte 101
6.6.2 Schätzmodell 101
6.6.3 Bewertung 102
6.6.4 Fazit, Ausblick 103
6.7 Kontinuierliche Beschäftigung nach Integration in sozialversicherungspflichtige Beschäftigung 103
6.7.1 Analyse fehlender Werte 103
6.7.2 Schätzmodell 104
6.7.3 Bewertung 108
6.7.4 Fazit, Ausblick 109
7 Analysen zu Mehrfachausfällen 110
7.1 Einleitung 110
7.2 Kennzahl K1 „Veränderung der Summe der Leistungen zum Lebensunterhalt“ 111
7.2.1 Kennzahl K1, ihre Grunddaten und dazugehörigen Imputationsverfahren 111
7.2.2 Simulation von Mehrfachausfällen 111
7.2.3 Ergebnisse 112
7.2.4 Fazit, Ausblick 113
7.3 Ergänzungsgröße K1E2 „Veränderung der Zahl der erwerbsfähigen Leistungsberechtigten“ 113
7.3.1 Ergänzungsgröße K1E2, ihre Grunddaten und dazugehörigen Imputationsverfahren 113
7.3.2 Simulation von Mehrfachausfällen 113

7.3.3 Ergebnisse 114
7.3.4 Fazit, Ausblick 115
7.4 Kennzahl K2 „Integrationsquote“ 115
7.4.1 Kennzahl K2, ihre Grunddaten und dazugehörigen Imputationsverfahren 115
7.4.2 Simulation von Mehrfachausfällen 116
7.4.3 Ergebnisse 117
7.4.4 Fazit, Ausblick 118
7.5 Kennzahl K3 „Veränderung des Bestands an Langzeitleistungsbeziehenden “ 119
7.5.1 Kennzahl K3, ihre Grunddaten und dazugehörigen Imputationsverfahren 119
7.5.2 Simulation von Mehrfachausfällen 119
7.5.3 Ergebnisse 120
7.5.4 Fazit, Ausblick 121
7.6 Abschätzung des Schätzfehlers bei gleichzeitigen Mehrfachausfällen in Zähler und Nenner 121
7.7 Fazit aus den Mehrfachausfallanalysen....122

8 Anhang....124

8.1 Lesehilfe zu Boxplots 124
8.2 Lesehilfe zu kombinierten Boxplot-Dichte-Grafiken 125
8.3 Schrittweise Erweiterung des Regressionsmodells für die LLU-Schätzung 127
8.4 Datenimputation mit dem Optimierungsmodell für Bewegungsdaten (Zu- und Abgänge von ELB bzw. LZB) 129
8.5 Fortschreibung auf Basis des k-nächste-Nachbarn-Verfahrens am Beispielthema Integrationen 134
8.6 Modellauswahl und -vergleich für ein regressionsbasiertes Schätzverfahren bei Integrationen von Alleinerziehenden 138
8.7 Modellauswahl und -vergleich für ein regressionsbasiertes Schätzverfahren bei Integrationen von LZB....142

I. Abkürzungsverzeichnis

Tabelle 2: Abkürzungsverzeichnis

Abkürzung	Langfassung
A	
AMP-Daten	Daten zur Arbeitsmarktpolitik
B	
BA	Bundesagentur für Arbeit
BG	Bedarfsgemeinschaft
BBSR	Bundesinstitut für Bau-, Stadt- und Raumforschung
BL	Bundesland
BST	Beschäftigungsstatistik
E	
ELB	erwerbsfähige Leistungsberechtigte
F	
FW	Fortschreibungswert
FST	Förderstatistik
G	
gE	gemeinsame Einrichtung
geB	geringfügige Beschäftigung
I	
IAB	Institut für Arbeitsmarkt- und Berufsforschung
J	
JC	Jobcenter
K	
K1	Veränderung der Summe der Leistungen zum Lebensunterhalt (LLU)
K1E1	Veränderung der Summe der Leistungen für Unterkunft und Heizung (LUH)
K1E2	Veränderung der Zahl der erwerbsfähigen Leistungsberechtigten (ELB)
K1E3	Durchschnittliche Zugangsrate der erwerbsfähigen Leistungsberechtigten (ELB)
K1E4	Durchschnittliche Abgangsrate der erwerbsfähigen Leistungsberechtigten (ELB)
K2	Integrationsquote
K2E1	Quote der Eintritte in geringfügige Beschäftigung (geB)
K2E2	Quote der Eintritte in öffentlich geförderte Beschäftigung (ögB)
K2E3	Kontinuierliche Beschäftigung nach Integration
K2E4	Integrationsquote der Alleinerziehenden
K3	Veränderung des Bestands an Langzeitleistungsbeziehenden
K3E1	Integrationsquote der Langzeitleistungsbeziehenden
K3E2	Aktivierungsquote der Langzeitleistungsbeziehenden
K3E3	Durchschnittliche Zugangsrate der Langzeitleistungsbeziehenden
K3E4	Durchschnittliche Abgangsrate der Langzeitleistungsbeziehenden
KBNI	kontinuierliche Beschäftigung nach Integration
KBNÜ	kontinuierliche Beschäftigung nach Übergängen von erwerbsfähigen Leistungsberechtigten (ELB) in sozialversicherungspflichtige Beschäftigung (svB)
knN-Verfahren	k-nächste-Nachbarn-Verfahren
L	
LLU	Leistungen zum Lebensunterhalt
LUH	Leistungen für Unterkunft und Heizung
LZB	Langzeitleistungsbeziehende
N	
NEF	nicht erwerbsfähige Leistungsberechtigte
O	
ögB	öffentlich geförderte Beschäftigung

Abkürzung	Langfassung
R	
RVO	Rechtsverordnung
S	
svB	sozialversicherungspflichtige Beschäftigung
T	
TgAw	Träger in getrennter Aufgabenwahrnehmung
V	
VJM	Vorjahresmonat
VM	Vormonat
VT	SGB-II-Vergleichstyp
W	
WZ	Wartezeit
WZ<3	Daten mit verkürzter Wartezeit von unter drei Monaten
Z	
zkT	zugelassene kommunale Träger

II. Abbildungsverzeichnis

Abbildung 1: Abweichungen der Männeranteilswerte vom endgültigen WZ3-Anteil – Bestand an ELB 32
Abbildung 2: LLU – fehlende Werte 34
Abbildung 3: LLU – Methodenvergleich Fehlerverteilungen 37
Abbildung 4: LUH – Methodenvergleich Fehlerverteilungen 41
Abbildung 5: ELB – fehlende Werte 43
Abbildung 6: ELB – Methodenvergleich Fehlerverteilungen 46
Abbildung 7: Alleinerziehende ELB – Methodenvergleich Fehlerverteilungen 49
Abbildung 8: LZB – fehlende Werte 51
Abbildung 9: LZB – Methodenvergleich Fehlerverteilungen 55
Abbildung 10: Aktivierte LZB – Schätzmodell/Entscheidungsbaum 59
Abbildung 11: Aktivierte LZB – Methodenvergleich Fehlerverteilungen bei der Fortschreibung mit Hilfe der Deutschlandentwicklung 64
Abbildung 12: ELB-Bewegungen – fehlende Werte 66
Abbildung 13: Abgänge von ELB – Methodenvergleich Fehlerverteilungen 69
Abbildung 14: Zugänge von ELB – Methodenvergleich Fehlerverteilungen 69
Abbildung 15: LZB-Bewegungen – fehlende Werte 71
Abbildung 16: Abgänge von LZB – Methodenvergleich Fehlerverteilungen 74
Abbildung 17: Zugänge von LZB – Methodenvergleich Fehlerverteilungen 74
Abbildung 18: Zu- und Abgänge LZB – Methodenvergleich bei Langzeitlücken: Test-Rekonstruktion versus Optimierungsmodell 76
Abbildung 19: Integrationen – fehlende Werte 78
Abbildung 20: Integrationen – Methodenvergleich Fehlerverteilungen 81
Abbildung 21: Integrationen von Alleinerziehenden – Methodenvergleich Fehlerverteilungen 86
Abbildung 22: Integrationen von LZB – fehlende Werte 88
Abbildung 23: Integrationen von LZB – Schätzmodell/Entscheidungsbaum 89
Abbildung 24: Integrationen von LZB – Methodenvergleich Fehlerverteilungen beim regressionsbasierten Schätzverfahren 92
Abbildung 25: Eintritte in geB – fehlende Werte 94
Abbildung 26: Eintritte in geB – Methodenvergleich Fehlerverteilungen 96
Abbildung 27: Eintritte in ögB – Methodenvergleich Fehlerverteilungen 99
Abbildung 28: Eintritte in ögB – Schätzung auf Basis der Daten der FST – Methodenvergleich Fehlerverteilungen 99
Abbildung 29: Integrationen in svB – Methodenvergleich Fehlerverteilungen 102
Abbildung 30: KBNI bzw. KBNÜ (Stufen 1 bis 3) – Überblick über die Messmodelle 106
Abbildung 31: KBNI – Methodenvergleich Fehlerverteilungen 108

Abbildung 32: KBNI – Methodenvergleich Fehlerverteilungen der Kaskaden 108
Abbildung 33: Mehrfachausfälle: Relativer Fehler bei LLU und die daraus resultierende Abweichung zur Kennzahl K1 112
Abbildung 34: Mehrfachausfälle: Relativer Fehler bei ELB und die daraus resultierende Abweichung zur Ergänzungsgröße K1E2.....114
Abbildung 35: Mehrfachausfälle: Relativer Jahresfehler bei Integrationen und ELB und die daraus resultierende Abweichung zur Kennzahl K2.....117
Abbildung 36: Mehrfachausfälle: Relativer Fehler bei LZB und die daraus resultierende Abweichung zur Kennzahl K3 120
Abbildung 37: Kombinierte Boxplot-Dichte-Grafik für den Methodenvergleich am Beispiel LLU 125
Abbildung 38: Modellvergleich – schrittweise Erweiterung des Regressionsmodells für die LLU-Schätzung 127

III. Tabellenverzeichnis

Tabelle 1: Änderungshistorie 3
Tabelle 2: Abkürzungsverzeichnis 9
Tabelle 3: Erste Entwicklungsstufe für Imputationsverfahren 18
Tabelle 4: Zweite Entwicklungsstufe für Imputationsverfahren 18
Tabelle 5: Dritte Entwicklungsstufe für Imputationsverfahren 18
Tabelle 6: Mit Datenimputationen versorgte Grunddaten und Kenngrößen 20
Tabelle 7: Verwendete Datenimputationsverfahren nach Entwicklungsstufen 21
Tabelle 8: Auswirkungen von Mehrfachausfällen nur im Zähler gegenüber gleichzeitigen Mehrfachausfällen in Zähler und Nenner bei ausgewählten Kennzahlen 122
Tabelle 9: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen mit Daten mit zwei Monaten Wartezeit 139
Tabelle 10: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen mit Daten mit einem Monat Wartezeit 140
Tabelle 11: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen mit Daten ohne Wartezeit 140
Tabelle 12: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen, wenn keine Daten mit verkürzter Wartezeit vorliegen 141
Tabelle 13: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen: Ersatzwertverfahren (Stufe 2) versus regressionsbasiertes Schätzverfahren (Stufe 3) 141
Tabelle 14: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen mit Daten mit zwei Monaten Wartezeit 143
Tabelle 15: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen mit Daten mit einem Monat Wartezeit 144
Tabelle 16: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen mit Daten ohne Wartezeit 144
Tabelle 17: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen, wenn keine Daten mit verkürzter Wartezeit vorliegen 145

1 Einleitung

1.1 Hintergründe für fehlende Werte bei den Kennzahlen nach § 48a SGB II

Im Rahmen der Stichtagsvalidierung für die amtliche Statistik der Grundsicherung für Arbeitsuchende nach dem SGB II (Grundsicherungsstatistik SGB II) kommt es vor, dass **einzelne Monatswerte** für Jobcenter (JC) als **nicht plausibel** eingestuft werden müssen. Oft liegt der Grund für unplausibel hohe oder niedrige JC-Werte in Verarbeitungs- oder Zuordnungsproblemen nach einer Änderung des Trägergebietes oder der Trägerform. Ursächlich sind häufig auch Softwareumstellungen und Versionsänderungen beim Datenstandard XSozial-BA-SGB II, auf dem die Datenübermittlung aus den Liefersystemen der zugelassenen kommunalen Träger (zkT) beruht.

Die Plausibilitätsprüfungen fußen auf statistischen Verfahren. Sie umfassen unter anderem einfache und kombinierte Vormonats- und Vorjahresvergleiche sowie Stock-Flow-Prüfungen[1]. Daneben erfolgen geregelte Rückkopplungsprozesse: JC, die von bestimmten Datenauffälligkeiten betroffen sind, werden kontaktiert und deren Einschätzung eingeholt.[2]

Werte, die als unplausibel eingestuft wurden, werden in der Grundsicherungsstatistik SGB II für die betroffenen JC und Berichtsmonate bis auf einige durch Schätzung ermittelte Eckwerte (vgl. Abschnitt 4.1) nicht ausgewiesen. Entsprechend kommt es auch bei den **Grunddaten** der Kennzahlen und Ergänzungsgrößen nach § 48a SGB II zu **Datenlücken**. Diese Lücken treten im Wesentlichen vereinzelt auf und weisen keine systematischen Muster auf. Insbesondere sind die fehlenden Werte nicht von der Höhe der unbeobachteten (aber interessierenden) Daten abhängig. Es gilt demnach die „Missing at Random"-Annahme. Das bedeutet zum Beispiel, dass bei JC, bei denen in einem bestimmten Berichtsmonat die Anzahl der erwerbsfähigen Leistungsberechtigten (ELB) nicht plausibel vorliegt, die ELB-Anzahl weder systematisch niedriger noch höher ist als bei anderen JC.

1.2 Konsequenzen aus fehlenden Werten ohne Datenimputation

Bis Oktober 2012 wurden für Datenlücken bei der Berichterstattung für Kennzahlen und Ergänzungsgrößen nach § 48a SGB II nicht explizit Schätzwerte erstellt. Bis zu diesem Zeitpunkt führten Datenlücken häufig dazu, dass Kenngrößen für das betreffende JC nicht ausgewiesen werden konnten – mit folgenden Konsequenzen:

[1] Das „Stock-Flow-Modell" beschreibt allgemein den Zusammenhang zwischen Stromgrößen und Bestand. Der Bestand eines Berichtsmonats ergibt sich rechnerisch aus dem Bestand des Vormonats minus der Abgänge und plus der Zugänge zwischen den Stichtagen. Stock-Flow-Prüfungen sind gebräuchliche statistische Konsistenztests für Strom- und Bestandsgrößen im Zeitablauf.

[2] Nähere Informationen zu den Prüfkriterien befinden sich im „Handbuch – Grundlagen der Datenübermittlung" sowie im „Handbuch – Grundsicherung für Arbeitsuchende". Beide Dokumente werden von der Statistik der BA zum Download zur Verfügung gestellt, vgl. https://statistik.arbeitsagentur.de/ > Grundlagen > Datenquellen > Datenstandard XSozial-BA-SGB II > Handbuch.

- Bei Kenngrößen, die auf einem Vorjahresvergleich beruhen (K1, K1E1, K1E2, K3) oder einen Anteilswert darstellen (K3E2), führten fehlende Zähler- oder Nennergrunddaten **direkt zu Datenlücken bei den entsprechenden Kenngrößen**.
- Bei allen anderen Kenngrößen und fast allen Jahresfortschrittskennzahlen wurden Datenlücken bis zu einem gewissen Maß über **ausfalltolerante Berechnungsregeln** für die zugrundeliegenden Zwölfmonatswerte in Zähler und Nenner aufgefangen: Solange im betreffenden Zwölfmonatszeitraum wenigstens zehn plausible Werte in Zähler und zugleich Nenner vorlagen, wurde auf die zwölf Werte umgerechnet. Dieses Vorgehen stellte implizit bereits eine Ersetzung fehlender Werte (Datenimputation)[3] dar, bei der fehlende Grunddaten durch gewichtete Mittelwerte ersetzt wurden. Die betreffende Kenngröße wurde dann geklammert ausgewiesen. Die Klammer verdeutlichte, dass der auf Grundlage der zehn oder elf verfügbaren Berichtsmonate ausgewiesene Wert vom „wahren", aber unbekannten Wert, abwich.

Aufgrund der engen Verwandtschaft zwischen den einzelnen Kennzahlen und Ergänzungsgrößen bedingten Datenlücken bei einem Grundwert häufig auch Datenlücken bei anderen Grundwerten: So setzten sich fehlende Werte in einem bestimmten Grundwert bei dessen Differenzierung nach Geschlecht natürlich fort. Datenlücken bei Langzeitleistungsbeziehenden bedingten Datenlücken bei deren Integrationszahl, Aktivierungsquoten sowie Zu- und Abgangsraten. Fehlende Integrationswerte brachten fehlende Werte für Integrationen in sozialversicherungspflichtige Beschäftigung sowie bei Alleinerziehenden oder Langzeitleistungsbeziehenden mit sich. Insofern **setzten sich Datenlücken** bei Grundwerten **über die verschiedenen Themenbereiche hinweg fort**, auch wenn sie über die beschriebenen Umrechnungsregeln (bzw. impliziten Datenimputationsregeln) für Kenngrößen und Jahresfortschrittskennzahlen in erheblichem Maße aufgefangen werden konnten.

1.3 Datenimputationsbedarf

Aus dem Blickwinkel der Steuerung richtet sich der Fokus jedoch nicht nur auf die Kenngrößen und Jahresfortschrittskennzahlen, sondern gleichrangig auch auf die zugrundeliegenden Grunddaten und Monatswerte. **Aus Steuerungssicht** sind **lückenlose Zeitreihen** für die vergleichende Bewertung von operativen Ergebnissen der JC **unabdingbar**. Deshalb wurde die Statistik der Bundesagentur für Arbeit (BA) gebeten, die Lücken bei Grunddaten für Kenngrößen und Jahresfortschrittskennzahlen mit **Schätzwerten** für die betroffenen JC und Berichtsmonate zu schließen, fehlende Grunddaten für JC also zu **imputieren**.

Die Datenimputation auf der Ebene von Grunddaten, JC und Berichtsmonaten bietet **Vorteile**. Wo statt fehlender Werte geschätzte Imputationswerte vorliegen, müssen die beschriebenen Umrechnungs- und Ausfallregeln nicht mehr angewendet werden. Dadurch wird die bisherige implizite Imputation **explizit** durch **sichtbare Daten** ersetzt. Folglich werden bei den verschiedenen Kennzahlenberechnungen stets

[3] Imputation bedeutet wörtlich „Unterstellung", „Annahme", „Vermutung". Bei der Imputation geht es darum, Ersatzwerte zu finden, die als sinnvoll und angemessen, also als „wahrscheinlich" oder „mutmaßlich" zutreffend unterstellt werden können.

die gleichen Datengrundlagen aus der Imputation der Grunddaten verwendet, was eine Voraussetzung für konsistente Berechnungen ist.

Die Berichterstattung der Kennzahlen nach § 48a SGB II ist die Grundlage für den Leistungsvergleich der Träger der Grundsicherung für Arbeitsuchende nach den in der Rechtsverordnung (RVO) zu § 48a SGB II festgelegten Zielgrößen. **Zusammenfassend** gesagt bieten die geschätzten Imputationswerte somit eine **explizite und sichere Vervollständigung** der Datengrundlage für den gesetzlichen Leistungsvergleich **im Sinne der RVO**.

1.4 Ansätze für Datenimputationen

Grundsätzlich geht es bei der Datenimputation im vorliegenden Fall um eine systematische Schätzung einzelner Werte in Zeitreihen für JC. Es wird angenommen, dass Informationen „aus dem Umfeld“ der Datenlücke gute Grundlagen für die Schätzung fehlender Werte bieten. Die hier vorliegenden Zeitreiheninformationen für die einzelnen Grunddaten und die rund 400 JC stellen von der Struktur her Matrizen mit Zeitreihen von Querschnittsdaten (‚cross sectional time series‘) dar, in denen einzelne Elemente fehlen.[4] Es gibt **in der Praxis verschiedene Ansätze** zur systematischen Ermittlung fehlender Werte in diesen Matrizen, bei denen vorhandene Informationen unterschiedlich intensiv genutzt werden, zum Beispiel:

- Subjektive Expertenschätzung
- Vormonatswert oder Vorjahreswert verwenden („Cold Deck“)
- Vormonatswert mit vorangegangener Veränderung fortschreiben
- Mittelwert der jüngeren Vergangenheit einsetzen oder zwischen zwei benachbarten Zeitreihenwerten interpolieren
- Zeitreihenanalytisches Fortschreibungsmodell verwenden
- Auch jene Zeitreiheninformationen verwenden, die zeitlich auf die Datenlücke folgen („Leads“ berücksichtigen)
- Aktuelle Veränderungsraten oder Relationen „ähnlicher“ JC verwenden („Hot Deck“, „Nearest-Neighbour-Algorithmus“)

[4] Der Ausfallanteil in den hier betrachteten Zeitreihen ist mit unter 1 % bei meist rund 30.000 zugrundeliegenden Datenpunkten sehr gering.

- Einbeziehung „verwandter“ Variablen in Regressionsmodellen und Nutzung der Modelle für die Schätzung der Ersatzwerte
- Einbeziehung „verwandter“ Variablen und Informationen aus anderen Regionen und über den Zeitverlauf durch multiples Imputationsmodell (zum Beispiel AMELIA II von Honaker, King und Blackwell)[5]

Die weiter oben in der Liste aufgeführten Ansätze sind einfacher durchzuführen, ziehen aber in der Regel auch größere Fehler nach sich. Die Ansätze weiter unten in der Liste erfordern in der Regel komplexe Berechnungen und häufig auch eine intensivere Suche nach geeigneten Modellierungen. Zudem stehen der häufig höheren Schätzsicherheit höhere Entwicklungs- und höhere monatliche Produktionsaufwände gegenüber.

Insbesondere die Einbeziehung von Daten des Folgemonats („Leads“) hätte im Produktivbetrieb große Nachteile: Das Warten auf zeitlich nachgelagerte Informationen hätte zur Folge, dass auch die Imputationswerte erst zeitverzögert zur Verfügung stünden. Zöge man zum Beispiel den festgeschriebenen Nachfolgemonatswert des fraglichen Grunddatums heran, erhielte man endgültige imputierte Werte am aktuellen Rand (das heißt im vorliegenden Fall mit Wartezeit von drei Monaten für die jüngste Vergangenheit) erst nach mehr als dreimonatiger Wartezeit.[6] Übergangsweise könnten am aktuellen Rand zum Beispiel einfachere vorläufige Imputationen erfolgen. Sie müssten allerdings in den darauffolgenden Monaten unter Berücksichtigung der Lead-Informationen revidiert werden. Insbesondere diesen Nachteilen ist es geschuldet, dass bei den bislang vorgeschlagenen Imputationsverfahren auf die Einbeziehung von Lead-Variablen verzichtet wird.

1.5 Stufenweise Entwicklung von Imputationsverfahren

Bei der Entwicklung von Verfahren zur Datenimputation für Kennzahlen nach § 48a SGB II wurden für die verschiedenen Sachverhalte jeweils mehrere sachlogisch sinnvolle Ansätze realisiert und die Ergebnisse verglichen. Es hat sich gezeigt, dass nicht über alle Grunddaten hinweg ein einheitliches Modell sinnvoll anwendbar ist. Vielmehr ist es erforderlich, möglichst **individuell angepasste Verfahren und Prädiktoren** für die verschiedenen Grunddaten zu verwenden. Nur so lassen sich Verfahren finden, die auch in den Einzelfällen, also für die einzelnen Berichtsmonate und JC, jeweils optimale Schätzergebnisse versprechen.

[5] AMELIA II ist ein Software-Paket für die multiple Datenimputation, das mit der freien Statistiksoftware R oder eigenständig angewendet werden kann. Der Schätzalgorithmus wurde bereits 2001 erstmals veröffentlicht, vgl. King, G.; Honaker, J.; Joseph, A.; Scheve, K. (2001), Analyzing Incomplete Political Science Data – An Alternative Algorithm for Multiple Imputation, in: American Political Science Review, 95(1), 49-69. Ein aktueller Methodenbeitrag dazu ist Honaker, J.; King, G.; Blackwell, M. (2011), Amelia II: A Program for Missing Data, in: Journal of Statistical Software, 45(7).

[6] Üblicherweise beruht die amtliche Grundsicherungsstatistik SGB II auf Auswertungen der Geschäftsdaten nach dreimonatiger Wartezeit (WZ3), da sie sonst erfahrungsgemäß zu stark von operativen Unter- und gelegentlich auch Übererfassungen betroffen wäre. Die Auswertung nach Wartezeit stellt eine erneute Betrachtung desselben Berichtszeitraums dar. Durch die Wartezeit erhöht sich dabei der Wissensstand. Erst mit dreimonatiger Wartezeit werden die Daten endgültig festgeschrieben. Die bisherigen Erkenntnisse ergeben, dass dann in der Regel alle Bearbeitungsrückstände aufgelöst sind, sodass man von einer vollständigen Erfassung ausgehen kann.

Die Entwicklung von Imputationsansätzen ist bei manchen Grunddaten einfacher, bei anderen schwieriger. Die **Entwicklungsarbeit** und **Primärimplementierung** (Februar 2012 bis Mai 2014) von Imputationsverfahren wurde deswegen **stufenweise** vorangetrieben.

In einer **ersten Entwicklungsstufe** kommen seit dem Liefermonat Oktober 2012 Schätzverfahren für folgende Grunddaten zum Einsatz:

Tabelle 3: Erste Entwicklungsstufe für Imputationsverfahren

Stufe 1	Leistungen zum Lebensunterhalt (LLU)
	Bestand an erwerbsfähigen Leistungsberechtigten (ELB)
	Bestand an alleinerziehenden ELB
	Bestand an Langzeitleistungsbeziehenden (LZB) bei isolierten Datenlücken
	Integrationen von ELB

In einer **zweiten Entwicklungsstufe** kommen seit dem Liefermonat Mai 2013 Schätzverfahren für folgende Grunddaten zum Einsatz:

Tabelle 4: Zweite Entwicklungsstufe für Imputationsverfahren

Stufe 2	Leistungen für Unterkunft und Heizung (LUH)
	Bestand an LZB bei Langzeitlücken (Vervollständigung des Verfahrens aus der 1. Stufe, das nur für isolierte Datenlücken angewendet werden kann)
	Bestand an aktivierten LZB bei Langzeitlücken
	Integrationen von alleinerziehenden ELB (erster Aufschlag)
	Integrationen von LZB bei Langzeitlücken
	Eintritte in geringfügige Beschäftigung (geB)
	Eintritte in öffentlich geförderte Beschäftigung (ögB)
	Integrationen in sozialversicherungspflichtige Beschäftigung (svB)

Anlass, die Schätzverfahren der vorangegangenen ersten Stufe zu ändern, gab es in der zweiten Stufe nicht. In der **dritten Entwicklungsstufe** hat die Statistik der BA Schätzverfahren für die restlichen Grunddaten entwickelt:

Tabelle 5: Dritte Entwicklungsstufe für Imputationsverfahren

Stufe 3	Bestand an aktivierten LZB bei isolierten Datenlücken (Vervollständigung des Verfahrens aus der 2. Stufe, das nur für Langzeitlücken angewendet werden kann)
	Zu- und Abgänge von ELB
	Zu- und Abgänge von LZB
	Integrationen von LZB bei isolierten Datenlücken (Vervollständigung des Verfahrens aus der 2. Stufe, das nur für Langzeitlücken angewendet werden kann)
	Nachhaltige Integrationen[7]

Bei drei ausstehenden Datenpunkten für Eintritte in öffentlich geförderte Beschäftigung wurde das Verfahren aus der zweiten Stufe vervollständigt. Für Integrationen von alleinerziehenden ELB wurde das Schätzverfahren aus der zweiten Stufe grundlegend verbessert. Mit der dritten Stufe konnten alle Kenngrößen durch Imputationen vervollständigt werden. Die Verfahren der dritten Stufe fließen für die betroffenen

[7] Zählergröße der K2E3 bis zur geänderten RVO zu Kennzahlen nach § 48a SGB II (15. März 2019), vgl. vorangehende Versionen dieser Methodenübersicht vor Version 5.2 (Stand Liefermonat Mai 2020).

Grunddaten seit Liefermonat Mai 2014 zusätzlich zu jenen der ersten beiden Stufen in die Kennzahlenberichterstattung nach § 48a SGB II ein. Dabei werden Grunddaten, die imputiert sind, sowie daraus abgeleitete Werte, Kennzahlen und Ergänzungsgrößen, die Datenimputationen enthalten geklammert ausgewiesen.

Nach diesen drei Entwicklungsstufen (Primärimplementierung) entstand erst 2019 neuerlicher Entwicklungsbedarf: Mit der geänderten RVO zu Kennzahlen nach § 48a SGB II (15. März 2019) wurde die bisherige Ergänzungsgröße K2E3 „Nachhaltigkeit der Integration“ ersetzt durch die „neue“ K2E3 „Kontinuierliche Beschäftigung nach Integration“. Die Schätzwerte für die Zähler-Grunddaten „kontinuierliche Beschäftigungen nach Integration in sozialversicherungspflichtige Beschäftigung“ fließen ab Liefermonat Mai 2020 in die Kennzahlenberichterstattung nach § 48a SGB II ein und vervollständigen die gesamten Zeitreihen.

Auch in Zukunft gilt: wenn zweckmäßige Ideen zur Konsolidierung oder Weiterentwicklung des Imputationssystems entstehen, wird fachlich geprüft, ob sie mit vertretbarem Aufwand umzusetzen sind.[8] Die nachfolgende Tabelle fasst die Inhalte der Gesamtentwicklung in Bezug auf Grunddaten und Kenngrößen übersichtlich zusammen.

[8] Die Revision der amtlichen Grundsicherungsstatistik SGB II im April 2016 gab keinen Anlass, die Imputationsmodelle zu verändern. Näheres hierzu in Abschnitt 2.1.1.

Tabelle 6: Mit Datenimputationen versorgte Grunddaten und Kenngrößen

Nr.	Kürzel	Kennzahl	Beteiligte Monatswerte: Zähler / Nenner	Datenimputationen in Entwicklungsstufe …	
1	K1	Veränderungsrate bei LLU	LLU / LLU	1 1	
2	K1E1	Veränderungsrate bei LUH	LUH / LUH	2 2	
3	K1E2	Veränderungsrate bei ELB	ELB / ELB	1 1	
4	K1E3	Zugangsrate bei ELB	Zugänge von ELB / ELB	3 1	
5	K1E4	Abgangsrate bei ELB	Abgänge von ELB / ELB	3 1	
6	K2	Integrationsquote	Integrationen / ELB	1 1	
7	K2E1	Quote der Eintritte in geringfügige Beschäftigung (geB)	Eintritte in geB / ELB	2 1	
8	K2E2	Quote der Eintritte in öffentlich geförderte Beschäftigung (ögB)	Eintritte in ögB / ELB	2 1	
9	K2E3	Kontinuierliche Beschäftigung nach Integration	Kontinuierliche Beschäftigung nach Integration / Integrationen in svB	3 3[1)]	
10	K2E4	Integrationsquote bei Alleinerziehenden	Integrationen alleinerziehender ELB / Alleinerziehende ELB	3[2)] 1	
				Isolierte Datenlücken	Langzeitlücken
11	K3	Veränderungsrate bei LZB	LZB / LZB	1 1	2 2
12	K3E1	Integrationsquote bei LZB	Integrationen von LZB / LZB	3 1	2 2
13	K3E2	Aktivierungsquote bei LZB	Aktivierte LZB / LZB	3 1	2 2
14	K3E3	Zugangsrate bei LZB	Zugänge von LZB / LZB	3 1	3 2
15	K3E4	Abgangsrate bei LZB	Abgänge von LZB / LZB	3 1	3 2

[1)] Unabhängig von den drei Entwicklungsstufen der Primärimplementierung: Neu entwickeltes Schätzverfahren für „kontinuierliche Beschäftigung nach Integration“ infolge der geänderten RVO zu Kennzahlen nach § 48a SGB II, im Einsatz ab Mai 2020
[2)] Verbesserung des Schätzverfahrens aus Entwicklungsstufe 2

1.6 Kurz-Charakterisierung aller verwendeten Datenimputationsverfahren

Die verwendeten Datenimputationsverfahren sind grundlegend in drei Gruppen zu unterscheiden: (1) Fortschreibungsverfahren, (2) Verfahren, die auf Lieferinformationen im zeitlichen Umfeld zurückgreifen und (3) bestandszahlenbasierte Rekonstruktionsverfahren, vgl. nachfolgende tabellarische Übersicht:

Tabelle 7: Verwendete Datenimputationsverfahren nach Entwicklungsstufen

Art des Verfahrens			Anwendung für Grunddaten in der Entwicklungsstufe…		
			… 1	… 2	… 3
1 Fortschreibungsverfahren					
	a	basierend auf Bundesentwicklung			Aktivierte LZB[4)]
	b	basierend auf SGB-II-Typisierung	ELB LZB[2)]		
	c	basierend auf linearer Regression	LLU		
	d	basierend auf knN-Verfahren[1)]		Eintritte in geB	
	e	basierend auf Kombination: lineare Regr./knN-Verfahren[1)]	Integrationen	LUH	
	f	basierend auf Optimierungsmodell mit Stock-Flow-Gleichung			Zu- und Abgänge ELB Zu- und Abgänge LZB
2 Rückgriff auf Lieferinformationen im zeitlichen Umfeld					
	a	Anteilswertverfahren	alle Geschlechterschätzungen Alleinerziehende ELB	Eintritte in svB	Aktivierte LZB[4)] Integrationen von LZB[4)]
	b	Ersatzwertverfahren		Eintritte in ögB	
	c	Regressionsbasiertes Schätzverfahren			Integrationen von Alleinerziehenden Integrationen von LZB[4)] Kontinuierliche Beschäftigungen nach Integration[5)]
3 Bestandszahlenbasierte Rekonstruktionsverfahren					
		(ohne weitere Untergliederung)		LZB[3)] Aktivierte LZB[3)] Integration von LZB[3)]	

[1)] k-nächste-Nachbarn-Verfahren, ausführlich in Abschnitt 6.1.2 und 0
[2)] Bei isolierten Datenlücken
[3)] Bei Langzeitlücken
[4)] Bei einem Teil der isolierten Datenlücken
[5)] Unabhängig von den drei Entwicklungsstufen der Primärimplementierung: Neu entwickeltes Schätzverfahren für „kontinuierliche Beschäftigung nach Integration“ infolge der geänderten RVO zu Kennzahlen nach § 48a SGB II, im Einsatz ab Mai 2020

Die genannten Verfahren werden in den Kapiteln 3 bis 6 für die jeweiligen Grunddaten der Kenngrößen ausführlich erläutert. Im Folgenden werden die verwendeten Datenimputationsverfahren knapp charakterisiert.

1.6.1 Fortschreibungsverfahren

Bei den Fortschreibungsverfahren werden fehlende Werte durch das Produkt des Vormonatswertes mit einem Fortschreibungsfaktor imputiert. Der Fortschreibungsfaktor ist dabei die mittlere zeitliche Verände-

rung des entsprechenden Grunddatums bei einer gegebenen Referenzgruppe. Das heißt die monatsspezifische Entwicklung der jeweiligen Referenzgruppe wird auf das betreffende JC übertragen. Diesem Verfahren liegt die Annahme zugrunde, dass sich die Daten der Referenzgruppe und des betreffenden JC im Zeitverlauf ähnlich entwickeln. Das Fortschreibungsverfahren wird in sechs verschiedenen Varianten angewendet. Sie unterscheiden sich lediglich darin, auf welche Weise die Fortschreibungsfaktoren bzw. die Referenzgruppen ermittelt werden. Diese Referenzgrößen variieren zugunsten einer im konkreten Anwendungsfall optimierten Schätzgüte.

a) Eine **Fortschreibung**, die **auf der Entwicklung auf Bundesebene basiert**, wird als „Rückfall-Lösung“ für Bestandsdaten der **aktivierten LZB** bei isolierten Datenlücken verwendet (vgl. Kapitel 4.4). Als Referenzgruppe dienen alle JC im Bundesgebiet. Der Fortschreibungsfaktor ist die mittlere Veränderungsrate der aktivierten LZB zum Vormonat. Hier ist die Schätzgüte – mangels anderer geeigneter Prädiktoren – nur mäßig, was angesichts der seltenen Anwendung als „Rückfall-Lösung“ akzeptabel erscheint, zumal die Schätzgüte mit dem ansonsten als „Bestlösung“ verwendeten Anteilswertverfahren gut ist.

b) Eine **Fortschreibung**, die **auf der SGB-II-Typisierung basiert**, wird für die Datenimputation bei Bestandsdaten der **ELB und der LZB** (bei isolierten Datenlücken) angewendet (vgl. Kapitel 4.1 für ELB bzw. 4.3 für LZB). Als Referenzgruppe dienen die JC des SGB-II-Vergleichstyps, dem das betreffende JC angehört. Somit ist der Fortschreibungsfaktor die mittlere Veränderungsrate der ELB bzw. LZB zum Vormonat im gemeinsamen Vergleichstyp. Somit werden Veränderungsmuster in wirtschaftsstrukturell ähnlichen Gebieten gut abgebildet. Für die Berichterstattung zentraler Eckwerte der amtlichen Grundsicherungsstatistik wird das Verfahren bereits seit 2005 angewendet. Mit der neuerlichen Verwendung im Rahmen von Datenimputationen für Kennzahlen nach § 48a SGB II stimmen amtliche Statistik und Kennzahleninformationsquellen überein. Zudem bietet das Verfahren in den Simulationstests eine hervorragende Schätzgüte.

c) Eine **Fortschreibung**, die **auf einer linearen Regression basiert**, wird für Imputationen bei **LLU** angewendet (vgl. Kapitel 3.1). Das regressionsbasierte LLU-Modell kann als Erweiterung des einfachen Fortschreibungsmodells für ELB bzw. LZB verstanden werden. Außer den mittleren Veränderungsraten für LLU im gemeinsamen SGB-II-Vergleichstyp werden als Prädiktoren noch die mittleren Veränderungsraten im gemeinsamen Bundesland sowie die JC-spezifischen Veränderungsraten der ELB, nicht erwerbsfähigen Leistungsberechtigten und Bedarfsgemeinschaften einbezogen. Somit lassen sich JC-übergreifende und JC-spezifische Sondereffekte gut abbilden. Zudem verbessert sich die Schätzgüte gegenüber einem einfachen Vergleichstyp-basierten Fortschreibungsmodell merklich auf ein ebenfalls sehr gutes Niveau.

d) Eine **Fortschreibung**, die **auf einem k-nächste-Nachbarn-Verfahren (knN) basiert**, wird für Imputationen bei **Eintritten in geringfügige Beschäftigung** angewendet (vgl. Kapitel 6.4). Als Referenzgruppe dienen möglichst ähnliche Nachbarn bzw. JC. Ähnlichkeit drückt sich über ähnliche saisonale Verlaufsmuster aus und wird anhand der monatlichen Veränderungsraten im zeitlichen Vorfeld zur Datenlücke gemessen. Die durchschnittliche Veränderungsrate der k nächsten Nachbarn wird auf das betreffende JC übertragen. In den Simulationstests zeigt sich ein merklicher Schätzfeh-

ler, zudem mit vielen deutlichen Ausreißern in Einzelfällen. Doch angesichts der häufig kleinen Fallzahlen und der starken Datensprünge in den Originaldaten ist dies erwartbar. Dementsprechend und mangels anderer geeigneter Modelle wird die Schätzgüte als befriedigend bewertet.

e) Eine **Fortschreibung**, die **auf einer Kombination aus einer linearen Regression und einer knN-Fortschreibung basiert**, wird für Imputationen bei **Integrationen** (vgl. Kapitel 6.1) und für Imputationen bei **LUH** (vgl. Kapitel 3.2) angewendet. Die Vereinigung der beiden Verfahren läuft technisch jeweils unterschiedlich ab. Das hat für die Bewertung jedoch keine Bedeutung. Die Vorteile beider Teilverfahren werden miteinander vereint: Das Teilverfahren der linearen Regression (analog LLU-Modell) ist ein guter Schätzer in der mittleren Tendenz – das Teilverfahren der knN-Fortschreibung fängt die Integrations- und LUH-typischen Sondereffekte gut ein, die JC-übergreifend in der Zeitreihe zum Tragen kommen. Die Schätzgüte des Verfahrens ist bei LUH auf einem guten Niveau, wenn auch derjenigen bei LLU-Imputationen unterlegen. Letzteres ist bei den LUH-typischen Sondereffekten durch einmalige Nachzahlungen erwartbar und akzeptabel. Der merkliche Schätzfehler bei Integrationen wird mangels anderer geeigneter Modelle und angesichts der starken Datensprünge in den Originaldaten als befriedigend bewertet.

f) Das **Optimierungsmodell**, wie es für Imputationen bei beiden Bewegungsarten von ELB und LZB angewendet wird (vgl. Kapitel 5.1 für **ELB-Bewegungen** bzw. 5.2 für **LZB-Bewegungen**), kann als **Fortschreibungsverfahren** interpretiert werden: die in einem Zwischenschritt geschätzten Bewegungsraten kommen ähnlich wie Fortschreibungsfaktoren zum Einsatz. Dabei ist das Modell maßgeblich geprägt von der Stock-Flow-Gleichung, die als Nebenbedingung in die Optimierungsfunktion eingeht. Angesichts der teilweise kleinen Fallzahlen und der starken Datensprünge in den Originaldaten wird die Schätzgüte für alle vier Bewegungsgrößen als gut, in Teilbereichen als befriedigend bewertet.

1.6.2 Rückgriff auf Lieferinformationen im zeitlichen Umfeld

a) Ein auf imputierte Gesamtgrößen **aufsetzendes Anteilswertverfahren**, das **Anteilswertinformationen im zeitlichen Umfeld** (zum Beispiel frühere Wartezeitstände) des Schätzpunktes nutzt, wird für alle **Geschlechterschätzungen** (vgl. jeweiliges Kapitel) sowie für Imputationen bei **alleinerziehenden ELB** (vgl. Kapitel 4.2), bei **Integrationen in sozialversicherungspflichtige Beschäftigung** (vgl. Kapitel 6.6), im Falle isolierter Datenlücken bei **aktivierten LZB** (vgl. Kapitel 4.4) und schließlich bei **Integrationen von LZB** (vgl. Kapitel 6.3) als „Bestlösung“ angewendet. In den beiden letztgenannten Fällen werden die Anteilswertinformationen allerdings nicht aus dem zeitlichen Vorfeld, sondern aus den Daten gezogen, die das JC für den Ausfallmonat geliefert hat.

   In allen genannten Anwendungsfällen setzt das Anteilswertverfahren auf bereits imputierte Gesamt- oder Teilgrößen auf. In der Konsequenz kommt es zu kumulierten Schätzunsicherheiten – die Schätzgüte kann immer nur so gut sein wie bei der zugrundeliegenden geschätzten Gesamt- oder Teilgröße (zum Beispiel ELB, LZB, Integrationen oder sozialversicherungspflichtige Integrationen).

   Die jeweilige Schätzgüte variiert somit von Größe zu Größe: Sie ist bei der Geschlechterschätzung für ELB sehr gut, da sowohl die Schätzgüte für die Gesamtgröße der ELB als auch die Güte der

Anteilswertinformation nach Geschlecht sehr gut ist. Aber sie ist zum Beispiel bei Imputationen für kontinuierliche Beschäftigungen nach Integration nur befriedigend, weil bereits die Schätzgüte für die zugrundeliegenden Integrationen nur befriedigend ist. Durch die Qualität der nachfolgend verwendeten Anteilswertinformationen wird die Schätzgüte allenfalls stabil gehalten, mit kleinen Fallzahlen aber tendenziell verschlechtert. Besonders ausgeprägt dürfte dieses Phänomen bei allen Geschlechterschätzungen für Grunddaten mit kleinen Fallzahlen sein, zum Beispiel bei der Schätzung der Integrationen von männlichen Alleinerziehenden.

Für die jeweiligen Gesamtgrößen ohne Geschlechterdifferenzierung wird die kumulierte Schätzunsicherheit jedoch in keinem Anwendungsfall als bedrohlich angesehen.

b) Beim **Ersatzwertverfahren** werden **Informationen aus dem zeitlichen Vorfeld** des Schätzpunktes genutzt. Doch wird hier auf **Absolutwerte** statt auf Anteilswerte zurückgegriffen. Das Verfahren wird bei **Eintritten in öffentlich geförderte Beschäftigung** (vgl. Kapitel 6.5) angewendet. Neben Werten mit verkürzter Wartezeit (hier mit zweimonatiger Wartezeit) wird vor allem auf Werte aus der Förderstatistik zurückgegriffen. Die Schätzgüte des Verfahrens ist bei den Simulationstests in der mittleren Tendenz auf einem sehr guten Niveau, allerdings mit einigen deutlichen Ausreißern in Einzelfällen. Doch angesichts der häufig kleinen Fallzahlen und der starken Datensprünge ist dies erwartbar. Dementsprechend wird die Schätzgüte mindestens als gut bewertet.

   Für Imputationen bei Integrationen von alleinerziehenden ELB wird seit der dritten Entwicklungsstufe nicht mehr auf das Ersatzwertverfahren zurückgegriffen, sondern auf das nachfolgend beschriebene verbesserte Verfahren:

c) Ein **regressionsbasiertes Schätzverfahren** wird für Imputationen bei **Integrationen von Alleinerziehenden** (vgl. Kapitel 6.2), für Imputationen bei **Integrationen von LZB** (als „Rückfall-Lösung“, vgl. Kapitel 6.3) sowie für kontinuierliche Beschäftigungen nach Integration (vgl. Kapitel 6.7) angewendet. Es beruht bei den beiden zuerst genannten Größen insbesondere auf Informationen mit verkürzter Wartezeit, die mit Nacherfassungsquoten und anderen geeigneten Prädiktoren „hochgerechnet“ werden, bei der zuletzt genannten Größe auf Anteilswerten aus der Beschäftigungsstatistik. In der jeweiligen Auswahl der Prädiktoren gibt es Unterschiede, die fachlich und teststatistisch begründet sind. Bei Integrationen von LZB (in der „Rückfall-Lösung“) ist von einer guten Schätzgüte, bei Integrationen von Alleinerziehenden und kontinuierlichen Beschäftigungen nach Integration von einer befriedigenden Schätzgüte auszugehen.

1.6.3 Bestandszahlenbasiertes Rekonstruktionsverfahren

a) Im Falle von **Langzeitlücken** wird für Imputationen bei **LZB ein bestandszahlenbasiertes Rekonstruktionsverfahren** (vgl. Kapitel 4.3 für LZB) angewendet. Es basiert auf kombinatorischen Abfragen im Auswertungssystem der Statistik der BA, mit denen auf Einzelfallebene die vorliegende Leistungsbezugshistorie untersucht wird. Die Abfragen werden für das betreffende JC spezifisch auf dessen Ausfallkonstellation hin erstellt und sind auch immer nur für das betreffende JC anwendbar. Ziel ist die näherungsweise Rekonstruktion der LZB-Zahl für das JC, so als hätte es keinen dauermessungsschädlichen Ausfall von Personenbestandsdaten gegeben. Konkret wird für ausgewählte

Personengruppen der Schwellwert für die Zugehörigkeit zu den LZB von 21 Monaten um die bei dem JC infolge technischer Lieferprobleme angefallenen Ausfallmonate herabgesetzt. Voraussetzung ist, dass für diese Personen in den teilweisen Ausfallmonaten tatsächlich keine Daten geliefert wurden. Umgekehrt ausgedrückt: Die Personen erhalten einen Aufschlag auf ihre gemessene, aber untererfasste Bezugsdauer in Höhe ihrer individuell anzunehmenden Zahl an Ausfallmonaten. In fallindividuellen Simulationen zeigt sich insbesondere angesichts des extrem langen Schätzzeitraums eine sehr gute Schätzgüte. Würde man stattdessen ein statistisches Schätzverfahren – welcher Art auch immer – anwenden, wäre es dem Rekonstruktionsverfahren vor allem am Ende des langen Schätzzeitraums deutlich unterlegen.

b) Ein großer Vorteil des Rekonstruktionsverfahrens ist, dass das JC-individuelle Verfahren nach seiner initiierenden Konzeption ohne großen Aufwand bei Langzeitlücken **auch für Teilgruppen** angewendet werden kann. Neben der Auswertung **nach Geschlecht** sind also auch Auswertungen für die Ergänzungsgrößen Bestand an **aktivierten LZB** (vgl. Kapitel 4.4) und **Integrationen von LZB** (vgl. Kapitel 6.3) – wiederum insgesamt und nach Geschlecht getrennt – relativ einfach möglich.

Das Rekonstruktionsverfahren ergänzt für den LZB-Bestand das Fortschreibungsverfahren, das nur für isolierte Datenlücken angewendet werden kann. Die Schätzverfahren für isolierte Datenlücken bei den Ergänzungsgrößen sind weiter oben beschrieben, vgl. 1.6.1a) und 1.6.2.a) für aktivierte LZB sowie 1.6.2.a) und 1.6.2.c) für Integrationen von LZB.

2 Einheitliche Vorgehensweisen bei den Schätzungen und bei deren Bewertung

2.1 Vergleich und Bewertung der Imputationsverfahren

Dieser Abschnitt beschreibt die Datengrundlage und geeignete Bewertungskriterien, mit denen die Datenimputationsverfahren gemessen und beurteilt werden.

2.1.1 Datengrundlage und Stabilität der Ergebnisse

Die **quantitative Bewertung der Schätzverfahren** wird **anhand statistischer Simulationsanalysen** durchgeführt, die im nächsten Abschnitt näher erläutert werden. Herangezogen wurden

- bei Imputationsverfahren der ersten und zweiten Entwicklungsstufe die Daten von August 2009 bis August 2012 mit Gebietsstand November 2012,
- bei Imputationsverfahren der dritten Entwicklungsstufe die Daten von August 2009 bis August 2013 mit Gebietsstand November 2013.

Aus verschiedenen Gründen beginnen die Zeitreihen mancher Grunddaten erst nach August 2009 – für (sozialversicherungspflichtige) Integrationen und geringfügige Beschäftigungsaufnahmen zum Beispiel erst im Januar 2011. Trotz stabiler Ausfallmuster wurden die auf diese Grunddaten bezogenen Simulationsanalysen aus der zweiten Stufe in der dritten Stufe sicherheitshalber mit den um zwölf Monate längeren Zeitreihen bis inklusive August 2013 wiederholt. Die Ergebnisse blieben stabil.

Im April 2016 hat die Statistik der BA eine umfassende Revision der amtlichen Grundsicherungsstatistik SGB II durchgeführt.[9] Alle ab Mai 2016 veröffentlichten Kennzahlen und Ergänzungsgrößen nach § 48a SGB II basieren auf den revidierten Daten der Grundsicherungsstatistik SGB II. Mit der Revision wurden statistische Definitionen geschärft und die einzelnen Personengruppen konsistenter voneinander abgegrenzt, insbesondere in den Randbereichen der Grundsicherungsstatistik SGB II. Quantitativ kam es zu geringfügigen Veränderungen. Allerdings blieben die grundlegenden Schätzzusammenhänge für die hier verwendeten Datenimputationsverfahren durch die Revision unbeeinflusst. Es bestand kein Anlass, die Schätzmodelle erneuten Analysen zu unterziehen oder sie zu verändern. Somit wurden im April 2016 – basierend auf vollständig revidierten Zeitreihen, aber unveränderten Schätzverfahren – alle Imputationswerte neu geschätzt.

Die Einschätzungen zur Schätzgüte der Imputationsverfahren beruhen auf der **Annahme stabiler Ausfallhäufigkeiten und -muster**. Deswegen wurden die **Lückenanalysen** für die Grunddaten bereits zwei Mal wiederholt:

[9] Nähere Informationen finden sich in einer Reihe von Methodenberichten zur Revision der Grundsicherungsstatistik SGB II. Die Dokumente werden von der Statistik der BA zum Download zur Verfügung gestellt, vgl. https://statistik.arbeitsagentur.de/ > Grundlagen > Methodik und Qualität > Methodenberichte und Hintergrundinfos > Grundsicherung für Arbeitsuchende (SGB II). Einen guten Überblick bietet insbesondere der erste Methodenbericht: Bergdolt, Breuer, Harsch, Noll (2015): Revision der Statistik der Grundsicherung für Arbeitsuchende – Grundlagen. Methodenbericht. Nürnberg: Statistik der BA.

- Für die Grunddaten, die in der ersten und zweiten Stufe mit Imputationsverfahren versehen wurden, wurden die Lückenanalysen auf Basis der um zwölf Monate verlängerten Zeitreihen in der dritten Stufe wiederholt. Die Ausfallhäufigkeiten und -muster blieben stabil. Die Simulationsanalysen der ersten und zweiten Stufe wurden deshalb in der dritten Stufe nicht wiederholt.
- Mit der Revision der Grundsicherungsstatistik SGB II kamen vereinzelt Datenlücken hinzu. Deswegen wurden die Lückenanalysen für alle Grunddaten auf Basis der Daten bis Oktober 2015 erneut wiederholt – weitestgehend mit stabilen Ergebnissen. Auch in dieser Hinsicht bestand kein Anlass, die Schätzmodelle erneuten Simulationsanalysen zu unterziehen oder die Schätzmodelle zu verändern.

Die aktuellen Ergebnisse und Bewertungen zu diesen Lückenanalysen finden sich jeweils in Abschnitt „Analyse fehlender Werte“ in den Kapiteln 3.1 bis 6.7.

2.1.2 Vorgehen bei der Simulation und Bewertungskriterien

Bei der Simulation werden die tatsächlich vorhandenen Ursprungswerte einzeln als fehlend angenommen, Schätzwerte simuliert und die Abweichungen zwischen tatsächlichen Werten und simulierten Schätzwerten bestimmt – je kleiner die Abweichungen sind, umso besser ist das Imputationsverfahren.[10] Zur Beurteilung der Verfahren werden in einem **ersten Schritt mittlere Abweichungen** sowie die **Streuung der Abweichungen** herangezogen. Insbesondere die **Häufigkeit** größerer Abweichungen, also von sogenannten **Ausreißern**, ist von Belang.

In einem **zweiten Schritt** werden beobachtete **Ausreißer** näher analysiert. Die jeweilige Konstellation wird im Einzelfall einer **qualitativen** Bewertung unterzogen und geprüft, ob es Gründe dafür gibt. Typischerweise liegt die Erklärung in spezifischen JC-individuellen Mustern im Zeitablauf wie zum Beispiel untypische Schwankungen in den Ursprungsdaten im nahen zeitlichen Umfeld der Schätzung. In solchen Konstellationen kann man die zwangsläufige Beeinträchtigung der Schätzgüte nicht dem Imputationsverfahren anlasten. Wenn es jedoch in der Mehrzahl der Fälle nicht gelingt, plausible Gründe für beobachtete große Einzelabweichungen zu identifizieren, muss man das Verfahren als ungeeignet zurückweisen.

Ergibt sich aus der quantitativen und qualitativen Bewertung keine eindeutige Entscheidung für ein bestimmtes Schätzverfahren, werden **möglichst einfache Verfahren** bevorzugt – einfach im Sinne von nachvollziehbar und effizient umsetzbar. Im Zweifel fällt somit die Entscheidung tendenziell auf diejenigen Modelle, die

- bei der Einführung leichter zu kommunizieren sind,
- zum Einstieg in die monatliche Produktion einfacher zu implementieren sind und/oder

[10] Diese sich jeweils über alle Berichtsmonate und JC erstreckenden Simulationen sind a) bei dem in der zweiten Stufe entwickelten Rekonstruktionsverfahren für Langzeitlücken bei LZB und deren Ergänzungsgrößen „aktivierte LZB“ und „Integrationen von LZB“ sowie b) bei dem in der dritten Stufe entwickelten Anteilswertverfahren für isolierte Datenlücken bei den selben Ergänzungsgrößen verfahrensbedingt nicht möglich. Die Begründung und Bewertung dieser Imputationsansätze erfolgt in den jeweiligen Textabschnitten (4.3, 4.4 und 6.3) in erster Linie anhand von Einzelfallanalysen und ergänzenden sachlogischen Erwägungen.

- im laufenden Produktivbetrieb einfacher zu pflegen sind.

Während der Primärentwicklung von Verfahren zur Datenimputation für Kennzahlen nach § 48a SGB II (Februar 2012 bis Mai 2014) wurden bei der Simulation zunächst nur Einzelwerte als fehlend behandelt („leave-one-out“). In der Realität liegen in seltenen Fällen aber auch Mehrfachausfälle vor, sodass sich Datenlücken bei einzelnen JC häufen. Hierzu wurden im Anschluss an die Primärentwicklungsarbeit ergänzende Analysen durchgeführt (September 2014 bis Ende 2015). Die ab Version 4.0 in dieser Methodenübersicht in Kapitel 7 dokumentierten Analysen zu Mehrfachausfällen beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Vor dieser Revision gab es bei den dort analysierten Grunddaten[11] höchstens eine Häufung von bis zu drei, nach Revision von bis zu fünf direkt aufeinander folgenden Mehrfachausfällen je JC, die mit statistischen Schätzverfahren imputiert werden. Bei allen anderen Grunddaten stieg diese Zahl von vier auf fünf. Angesichts dieser geringfügigen Ausweitung von direkt aufeinander folgenden Mehrfachausfällen wurde die Analyse nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden – zumal die Wiederholung der Lückenanalysen auf Basis der revidierten, bis Oktober 2015 verlängerten Zeitreihen ansonsten weitgehend stabile Ausfallmuster zeigte. Somit gilt vor wie nach der Revision der Grundsicherungsstatistik SGB II: Mehrfachausfälle gehen erwartungsgemäß mit einer erhöhten Schätzunsicherheit einher. Ausgehend von den bisher beobachteten Ausfallmustern ist jedoch von akzeptablen Unsicherheiten auszugehen. Insgesamt können die Ergebnisse der Imputationsverfahren – auch bei Mehrfachausfällen – als robust angesehen werden.

2.1.3 Standardvergleich und Standardvergleichsverfahren

Die vorgeschlagenen Datenimputationsverfahren werden mittels der beschriebenen Simulationsanalysen immer demselben **Standardvergleich** unterworfen. Die Idee ist, dem jeweils entwickelten Modell immer dasselbe oder dieselben Vergleichsverfahren entgegen zu halten. Es sollen einfache, intuitiv eingängige und leicht umsetzbare Referenzmodelle sein, die aufzeigen, wie die Ergebnisse ausfielen, wenn man keinen oder nur wenig Aufwand bei der Modellierung betreiben würde. Sie müssen nicht zwangsläufig für den Produktionsbetrieb geeignet sein, da sie nur dem simulativen Standardvergleich dienen. Die Wahl fällt auf (1) Vormonatswertersetzung, (2) Interpolationsverfahren und (3) Fortschreibungsmodell anhand der SGB-II-Typisierung.

Die **drei Standardvergleichsverfahren** funktionieren wie folgt:

[11] Leistungen zum Lebensunterhalt, Integrationen von ELB, Bestand an ELB sowie LZB.

1. **Vormonatswertersetzung:**
   Hier wird als Schätzwert für eine simulierte Datenlücke der Vormonatswert eingesetzt. Falls dieser fehlt, wird der letzte verfügbare Vormonatswert herangezogen.[12]

2. **Interpolationsverfahren:**
   Hier wird als Schätzwert für eine simulierte Datenlücke der Mittelwert aus den beiden umliegenden Datenpunkten, also der lineare Interpolationswert, eingesetzt. Falls umliegend Datenlücken vorliegen, werden die beiden nächsten verfügbaren Datenpunkte verwendet.[13] Bei diesem Standardvergleichsverfahren wird demnach der Nachfolgemonatswert des jeweiligen Grunddatums und damit eine zeitlich nachgelagerte Lead-Information herangezogen.

3. **Fortschreibungsmodell:**
   Hier wird als Schätzwert ein Fortschreibungswert eingesetzt. Er ergibt sich als Produkt des Vormonatswertes und der mittleren prozentualen Veränderung zum Vormonat in jenem SGB-II-Vergleichstyp (vgl. nachfolgender Abschnitt), dem das JC angehört. Das heißt, dass die Entwicklung der Daten aus dem Vergleichstyp auf das betreffende JC übertragen wird. Die zugrundeliegende Annahme ist, dass sich die Daten von JC desselben Vergleichstyps im Zeitverlauf ähnlich entwickeln (im Gegensatz zur Entwicklung in anderen Vergleichstypen).

Bei allen drei Standardvergleichsverfahren kann der erste simulierte Schätzwert einer Zeitreihe erst berechnet werden, wenn ein Vormonatswert verfügbar ist. Außerdem endet beim zweitgenannten Standardvergleichsverfahren – der Interpolation – infolge der Verwendung des Nachfolgemonatswerts die Zeitreihe der simulierten Schätzwerte spätestens beim vorletzten Datenpunkt. Beim Methodenvergleich werden nur die Datenpunkte eines Grunddatums herangezogen, wo über alle Vergleichsverfahren hinweg Schätzwerte vorliegen. Somit verkürzen sich die Zeitreihen links und rechts um je einen Monat. Für den hier vorliegenden Zweck des Standardvergleichs ist diese Kürzung unschädlich, da die Zeitreihen ausreichend lang sind.

Im Produktivbetrieb hätte ein Interpolationsverfahren den bereits erwähnten großen Nachteil, dass endgültige imputierte Werte erst nach mehr als dreimonatiger Wartezeit zur Verfügung stünden (vgl. Abschnitt 1.4). Für den jeweils aktuellen Berichtsmonat (das heißt im vorliegenden Fall mit Wartezeit von drei Monaten für die jüngste Vergangenheit) könnten allenfalls vorläufige Imputationen erfolgen, die dann rollierend revidiert werden müssten.

Vom ersten bis zum dritten Vergleichsmodell steigt die Komplexität der Berechnungen an. Dennoch produzieren auch die beiden einfacheren Standardvergleichsverfahren zumindest intuitiv brauchbare Ergeb-

[12] Die Vormonatswertersetzung produziert bei den hier vorliegenden trendbehafteten Zeitreihen zwangsläufig systematische Unter- bzw. Überschätzungen und somit keine erwartungstreuen Ergebnisse. Doch ist dieses Defizit erheblich geringer ausgeprägt als bei der gleitenden Mittelwertersetzung aus den letzten drei Monatswerten, die vor Version 2.0 dieser Methodenübersicht (Stand Mai 2013) als einziges Standardvergleichsmodell fungierte. Da noch zwei weitere Vergleichsmodelle im Einsatz sind, ist dieses geringfügige Defizit für den Zweck des Standardvergleichs akzeptabel.

[13] Im Falle umliegender Datenlücken wird keine Gewichtung vorgenommen. In diesen Fällen handelt es sich streng genommen nicht um eine lineare Interpolation. Diese geringfügige Unschärfe kann für den hier vorliegenden Zweck des Standardvergleichs in Kauf genommen werden.

nisse. Jedes Modell hat seine Stärken. Das erste Standardvergleichsverfahren der Vormonatswertersetzung ist insbesondere bei stabilen Zeitreihen ein harter Prüfstein. Das zweite Modell der linearen Interpolation nivelliert gut die Wirkung von Ausreißern im zeitlichen Umfeld eines Schätzpunktes. Das dritte Modell der Fortschreibung kann überdies Saisoneffekte gut einfangen. Die vorgeschlagenen Imputationsverfahren müssen sich somit **gegen ein starkes und variables Set an Standardvergleichsverfahren** behaupten. Die einheitliche Anwendung bietet Anhaltspunkte für die übergreifende Einschätzung der jeweiligen Schätzgüte.

Der erläuterte Standardvergleich erfolgt unter anderem anhand kombinierter Boxplot-Dichte-Grafiken, die in der Regel im jeweiligen Bewertungs-Abschnitt zu den Imputationsverfahren abgebildet sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots sowie Anhang 8.2, Lesehilfe zu kombinierten Boxplot-Dichte-Grafiken).

2.2 Verwendung der SGB-II-Typisierung

Sowohl bei den hier vorgeschlagenen Imputationsverfahren als auch bei deren Bewertung kommt immer wieder die SGB-II-Typisierung des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) zum Einsatz. Diese Typisierung trägt dem Umstand Rechnung, dass ein sinnvoller Vergleich von JC mit unterschiedlichen regionalwirtschaftlichen Rahmenbedingungen nicht möglich ist. Deshalb wird jedes JC einem der **Vergleichstypen mit möglichst ähnlichen Rahmenbedingungen** zugeordnet. Die Identifizierung der SGB-II-Vergleichstypen erfolgt mit Hilfe eines kombinierten Modells aus Regressions- und Clusteranalyse. Dabei werden unter anderem typische und relativ zeitstabile Aspekte wie zum Beispiel die Aufnahmefähigkeit des jeweiligen lokalen Arbeitsmarktes berücksichtigt. Der Rückgriff auf die SGB-II-Typisierung vereinfacht die Modellierung: Gemeinsamkeiten in der Entwicklung bestimmter Größen werden effizient genutzt.

Für die Arbeiten an allen drei Entwicklungsstufen wurde die SGB-II-Typisierung aus dem Jahr 2011 verwendet.[14] Zum Jahresbeginn 2014 wurde diese von einer aktualisierten Typisierung abgelöst.[15] Tests auf Grundlage der neuen Typisierung führten zu keinen nennenswerten Unterschieden bei den Imputationsergebnissen, deuteten aber auf eine Verbesserung der Schätzgüte hin.

[14] Die Dokumentation zur Berechnung der SGB-II-Vergleichstypen 2011 kann unter https://www.iab.de/ > Publikationen > IAB-Stellungnahme > 2011 heruntergeladen werden (Nr. 08/2011).

[15] Die Dokumentation zur Neukonzeption der SGB-II-Vergleichstypen 2014 kann unter https://www.iab.de/ > Publikationen > IAB Forschungsbericht > 2013 heruntergeladen werden (Nr. 11/2013).

Im Rahmen von Datenimputationen für Kennzahlen nach § 48a SGB II erfolgte im Liefermonat Mai 2014 mit der Produktivsetzung der dritten Entwicklungsstufe der Umstieg auf die neue SGB-II-Typisierung. Zeitgleich wurde auch die amtliche Grundsicherungsstatistik SGB II darauf umgestellt. Schätzrevisionen sind aus diesem Anlass nicht erforderlich. Bei zukünftigen Aktualisierungen der SGB-II-Typisierung wird ebenso vorgegangen. Es ist davon auszugehen, dass immer die jeweils aktuelle Typisierung die besten Imputationsergebnisse erbringt.

2.3 Schätzung fehlender Werte nach Geschlecht[16]

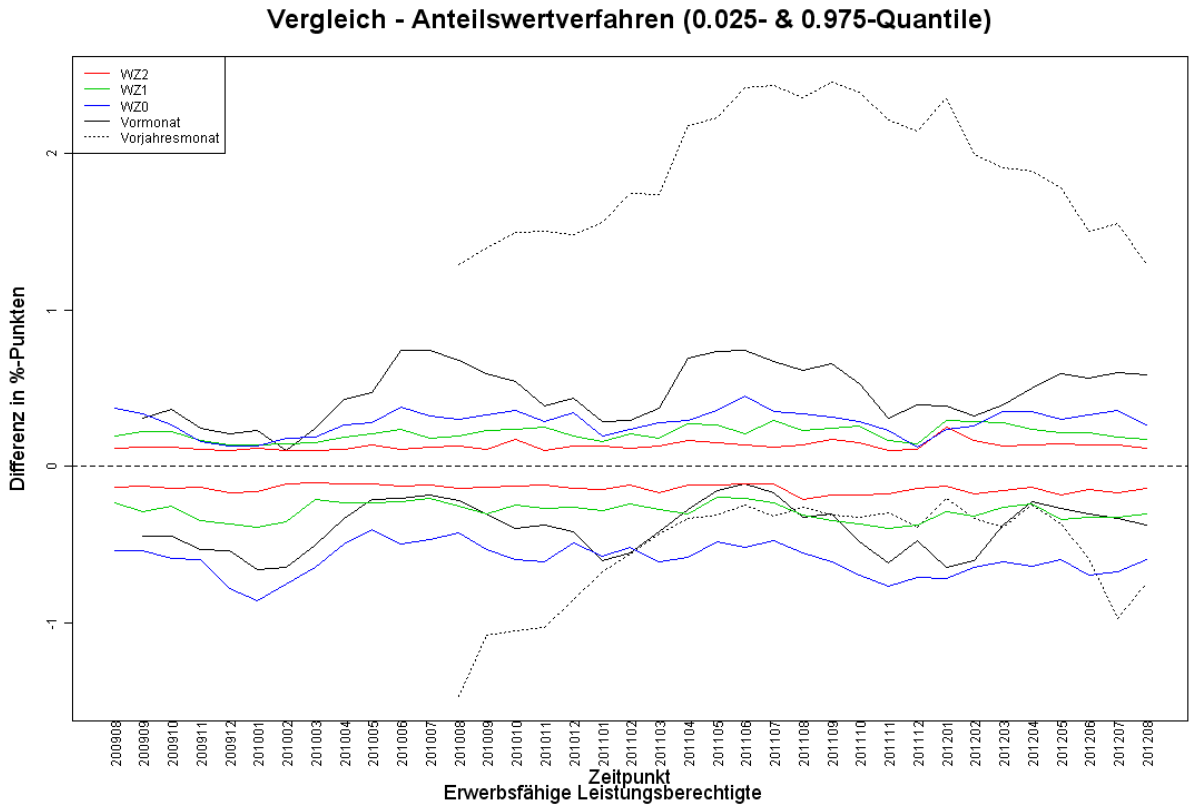
Die Berichterstattung der Kennzahlen und Ergänzungsgrößen nach § 48a SGB II erfolgt bei allen personenbezogenen Größen auch nach Geschlecht getrennt. Analysen zeigen, dass der jeweilige **Anteil von Männern und Frauen** bei den verschiedenen Grunddaten in allen JC **über die Zeit recht stabil** ist. Deswegen erfolgt die Schätzung von Datenimputationswerten nach Geschlecht per Anteilswertmethode, die die Verteilungsinformation aus der Vergangenheit nutzt. Nutzbar sind sowohl Informationen aus den vorangehenden Berichtsmonaten als auch aus den vorläufigen Daten mit verkürzter Wartezeit (WZ<3), also mit einer Wartezeit von zwei Monaten (WZ2), von einem Monat (WZ1) oder ohne Wartezeit (WZ0).

Die Liniengrafik zeigt am Beispiel der ELB-Bestandsdaten, wie nah sich die verschiedenen Anteilsinformationen um die endgültige Information mit dreimonatiger Wartezeit (WZ3, Nulllinie) scharen. Dargestellt sind die 2,5-%- und die 97,5-%-Perzentil-Linien der Abweichungen der Männeranteilswerte vom endgültigen WZ3-Anteil (a) im Vergleich zu den Daten mit WZ2, WZ1 und WZ0 desselben Berichtsmonats, (b) aus dem Vormonat (VM) mit WZ3 sowie (c) aus dem Vorjahresmonat (VJM) mit WZ3. Trotz Unter- bzw. gelegentlicher Übererfassung der Daten mit verkürzter Wartezeit desselben Berichtsmonats ist deren Geschlechterverteilung jener vorangehender Berichtsmonate in Punkto Stabilität erfahrungsgemäß überlegen:[17] Die Perzentil-Linien scharen sich nach außen hin, zunächst jene für denselben Berichtsmonat – je kürzer die Wartezeit, desto weiter außen, dann jene vorangehender Berichtsmonate. Die geschlechtsspezifischen Anteilswerte variieren also zwischen den verschiedenen Wartezeitständen eines Berichtsmonats weniger als zwischen den Berichtsmonaten.

[16] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[17] Um die Abbildung nicht zu überladen, sind für die vorangehenden Berichtsmonate die Perzentil-Linien für die verkürzten Wartezeiten (WZ2, WZ1 und WZ0) nicht dargestellt. Sie würden sich ähnlich wie die dargestellten Perzentil-Linien für denselben Berichtsmonat nach außen hin scharen – je kürzer die Wartezeit, desto weiter außen.

Abbildung 1: Abweichungen der Männeranteilswerte vom endgültigen WZ3-Anteil – Bestand an ELB



Die Abstände zwischen den jeweiligen Linienpaaren können als **Variabilitätsspannen** zwischen dem 2,5-%- und dem 97,5-%-Quantil über alle JC hinweg verstanden werden. Die Interpretation dieser verdichteten Information lautet: je enger die Variabilitätsspanne ist, desto näher liegt die Anteilsinformation an der endgültigen WZ3-Information und ist somit besser für eine Imputation mit der Anteilswertmethode geeignet. Für die ELB-Bestandsdaten der Berichtsmonate August 2009 bis einschließlich August 2012 liegt diese Variabilitätsspanne zwischen den Berichtsmonaten bei durchschnittlich 0,9 bis 1,0 Prozentpunkten (je nach Wartezeitstand). Die entsprechende Variabilitätsspanne zwischen den einzelnen Wartezeitständen desselben Berichtsmonats liegt – noch enger – bei 0,3 Prozentpunkten (WZ2 zu WZ3) bis 0,9 Prozentpunkten (WZ0 zu WZ3). Für die Schätzung der Datenimputationswerte nach Geschlecht sind also beide Arten der Anteilsinformation sehr gut geeignet, diejenige aus den geringeren Wartezeitständen meist ein wenig besser. Für die anderen Grunddaten ergeben sich die gleichen Muster. Somit folgt die **Schätzung der Verteilung nach Geschlecht über die meisten Grunddaten** hinweg dem **Schema**:

Sind die Daten im Referenzmonat unplausibel, so wird der Anteil geschätzt anhand von

1. WZ2-Daten, falls diese auch fehlen
2. WZ1-Daten, falls diese auch fehlen
3. WZ0-Daten, falls diese auch fehlen

4. Vormonatsdaten mit WZ3.[18]

Somit ergibt sich der Schätzwert beispielsweise für Frauen (*f*) eines JC *i* zum Zeitpunkt *t* mit dreimonatiger Wartezeit ($\hat{y}_{f,i,t,WZ3}$) als der entsprechende Schätzwert über beide Geschlechter (*g*) ($\hat{y}_{g,i,t,WZ3}$) multipliziert mit dem Frauenanteil zu einem früheren Wartezeitstand bzw. Zeitpunkt:

$$\hat{y}_{f,i,t,WZ3} = \hat{y}_{g,i,t,WZ3} \cdot \frac{\hat{y}_{f,i,t-j,WZ<3}}{\hat{y}_{g,i,t-j,WZ<3}} \quad (1)$$

Im besseren Fall findet man die Anteilsinformation im nächstkleineren Wartezeitstand desselben Berichtsmonats (WZ2, WZ1 oder WZ0 und j=0), im geringfügig schlechteren Fall im Vormonat mit dreimonatiger Wartezeit (WZ3 und j=1). Der auf Personenanzahl gerundete Wert wird genutzt, um den Wert für Männer (*m*) als Differenz zum Wert über beide Geschlechter zu ermitteln.

$$\hat{y}_{m,i,t,WZ3} = \hat{y}_{g,i,t,WZ3} - \hat{y}_{f,i,t,WZ3} \quad (2)$$

Die Reihenfolge des Vorgehens nach der Anteilswertmethode – zuerst Schätzung für Frauen oder für Männer – hat keinen Einfluss auf die Schätzung.

Bei einigen Grunddaten wird für die Ermittlung der Werte für Frauen und Männer vom oben beschriebenen Schema abgewichen. Das betrifft:

- Bestand an LZB bei Langzeitlücken (vgl. Abschnitt 4.3.2.2),
- Bestand an aktivierten LZB bei isolierten Lücken und bei Langzeitlücken (vgl. Abschnitt 4.4.2),
- Integrationen von LZB bei isolierten Lücken, sofern das Anteilswertverfahren zur Anwendung kommt (vgl. Abschnitt 6.3.2.1.1), und bei Langzeitlücken (vgl. Abschnitt 6.3.2.2),
- Eintritte in öffentlich geförderte Beschäftigung (vgl. Abschnitt 6.5.2) sowie
- kontinuierliche Beschäftigungen nach Integration (vgl. Abschnitt 6.7.2).

[18] Die Vorjahresinformationen wurden interessehalber mit geprüft und sind in der Abbildung ergänzend dargestellt.

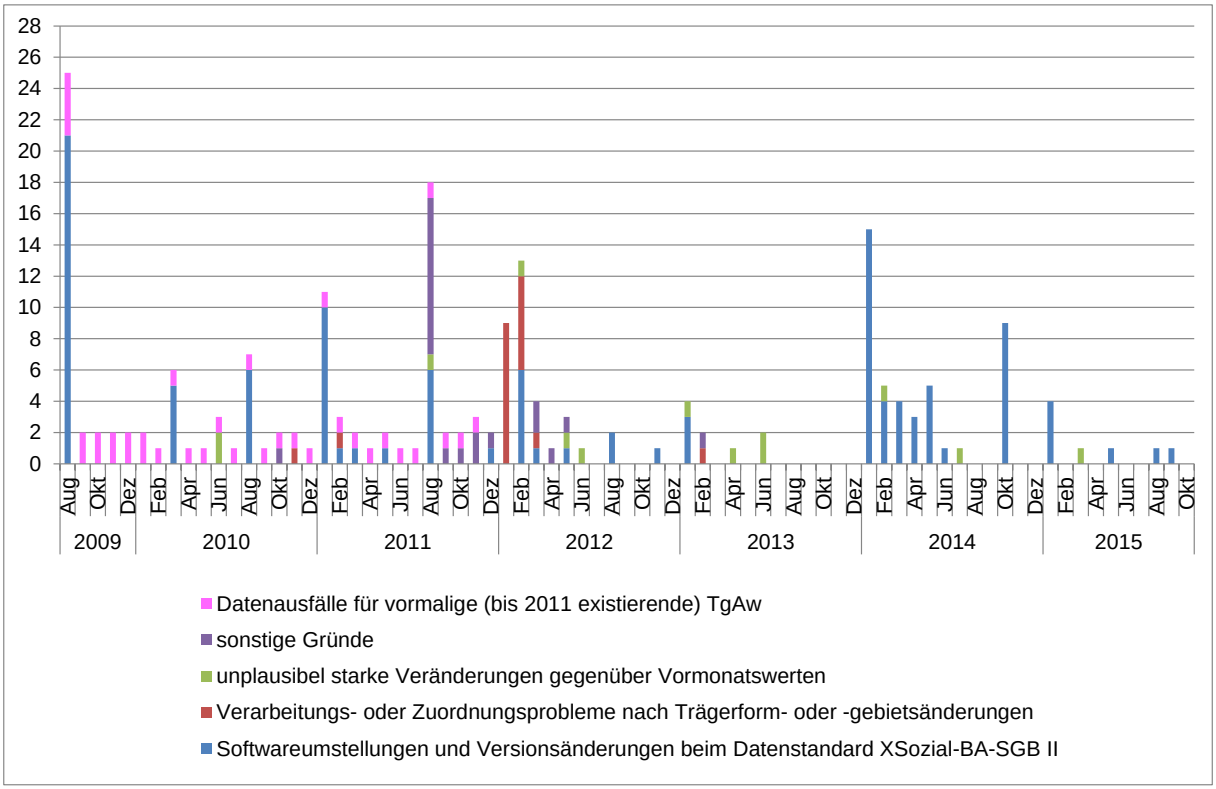
3 Datenimputation zum Themenbereich „Leistungen“

3.1 Leistungen zum Lebensunterhalt

3.1.1 Analyse fehlender Werte

Abbildung 2: LLU – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei LLU
Alle JC
Berichtsmonate August 2009 bis Oktober 2015



Bei Leistungen zum Lebensunterhalt (LLU) kommt es von August 2009 bis Oktober 2015 zu insgesamt 203 Datenlücken. Bezogen auf 75 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,66 %.

Mit der Revision der Grundsicherungsstatistik SGB II (April 2016) gibt es bei LLU gegenüber vorangehenden Versionen dieser Methodenübersicht ab Version 5.0 (Stand Mai 2016) mehr Datenlücken:

- Bis Ende 2011 wurde bei den bis dahin existierenden Trägern in getrennter Aufgabenwahrnehmung (TgAw) die Plausibilität der Datenlieferungen zu Bundesleistungen (Leistungen zum Lebensunterhalt, LLU) und kommunalen Leistungen (Leistungen für Unterkunft und Heizung, LUH, vgl. Abschnitt 3.2) getrennt geprüft. Diese Plausibilitätsinformationen wurden bis zur Revision im April 2016 getrennt weiterverarbeitet.
- Seit der Revision wird die Plausibilität für LLU und LUH – wie für andere Trägerformen auch – gemeinsam geprüft und weiterverarbeitet. Die Datenaufbereitung wird hierdurch an mehreren Stellen

deutlich erleichtert und abgesichert. Für einzelne dieser vormaligen TgAw liegen in Bezug auf LUH zu Beginn der Zeitreihe lange Datenausfälle vor: im Extremfall fehlen die Datenlieferungen für 28 direkt aufeinander folgende Berichtsmonate bis November 2011. Mit der Übertragung von Datenlücken von LUH- auf LLU-Daten kommt es somit zu einer Ausweitung der Datenlücken bei LLU-Daten bis Ende 2011.

- Dieses Problem wird für die Kennzahlenberichterstattung nach § 48a SGB II pragmatisch gelöst: Da die Datenlücken bei den betroffenen vormaligen TgAw ausschließlich auf fehlende Daten zu LUH (kommunale Leistungen) zurückgehen, können die Daten zu LLU (Bundesleistungen) als Schätzwerte eingesetzt werden.

Davon abgesehen häuften sich die Datenlücken in den vor allem Anfang 2012, was auf die zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen eine untergeordnete Rolle. Die Lücken verteilen sich auf 71 JC (etwa 17 % der JC). Sieht man von der beschriebenen Sonderkonstellation bei TgAw ab, fehlen bei gut der Hälfte aller JC mehrfach Daten für LLU, wobei die maximale Häufigkeit bei fünf hintereinander gereihten Datenlücken für ein JC liegt.

3.1.2 Schätzmodell

Im Gegensatz zu den meisten anderen Grunddaten der Kennzahlen nach § 48a SGB II verhält sich die Größe der LLU je JC über den Zeitverlauf vergleichsweise stabil. Sie verändert sich von Monat zu Monat oft nur geringfügig. Veränderungen der LLU sind dabei stark an Veränderungen der Zahl an ELB sowie nicht erwerbsfähigen Leistungsberechtigten (NEF) bzw. an Veränderungen der Zahl an Bedarfsgemeinschaften (BG) gekoppelt. Die Veränderungsmuster folgen überregionalen konjunkturellen Einflüssen und ähneln sich in wirtschaftsstrukturell ähnlichen Gebieten, hängen aber auch vom Antrags- und Erfassungsverhalten infolge von Ferien- und Feiertagen in den Bundesländern ab. Als Prädiktoren bieten sich deshalb zum einen Veränderungen der LLU im SGB-II-Vergleichstyp an, zum anderen aber auch Veränderungen der LLU im Bundesland. Mithilfe dieser Prädiktoren können auch Sondereffekte eingebunden werden wie zum Beispiel die jährliche Erhöhung des Regelbedarfs. Insgesamt erfolgt die Auswahl der Prädiktoren aufgrund sachlogischer Überlegungen und der engen bivariaten Zusammenhänge mit der Zielgröße. Alle Prädiktoren sind im laufenden Produktivbetrieb verfügbar.

Bei den Analysen erwiesen sich ein regressionsbasiertes Fortschreibungsmodell (Methode der kleinsten Quadrate) sowie ein multiples Imputationsmodell im Mittel ähnlich gut geeignet, fehlende Werte zu schätzen. Bei der Schätzung besonders atypischer Verläufe ist das multiple Imputationsmodell jedoch unterlegen. Aus diesem Grund und aus Effizienzgesichtspunkten fällt die Entscheidung auf das regressionsbasierte Fortschreibungsmodell.

Als abhängige Variable fungiert die prozentuale Veränderung der LLU zum Vormonat. Sie kann im zweiten Schritt auf den Vormonatswert aufgeschlagen werden, um fehlende LLU-Werte im Berichtsmonat zu schätzen. Die Zeitpunkte werden systematisch aufsteigend abgearbeitet. Dies bedingt – vorteilhaft im

Rahmen amtlicher Berichterstattung – die Festschreibung der Schätzwerte.[19] Die Schätzgleichung der prozentualen Veränderung der LLU eines JC *i* im SGB-II-Vergleichstyp *j(i)* (vgl. Abschnitt 2.2) im Bundesland *k(i)* zum Zeitpunkt *t* ergibt sich als:[20]

$$\Delta^p(LLU)_{t,i} = \beta_{0,i} + \beta_1 \cdot \Delta^p(ELB)_{t,i} + \beta_2 \cdot \Delta^p(NEF)_{t,i} + \beta_3 \cdot \Delta^p(BG)_{t,i} + \beta_4 \cdot \Delta^p(LLU_VT)_{t,j(i)} + \beta_5 \cdot \Delta^p(LLU_BL)_{t,k(i)} + \varepsilon_{t,i} \quad (3)$$

mit:

$\Delta^p(LLU)$	Veränderungen der LLU zum Vormonat in Prozent
β_0 bis β_5	Regressionskoeffizienten
$\Delta^p(ELB)$	Veränderung der ELB in Prozent
$\Delta^p(NEF)$	Veränderung der NEF in Prozent
$\Delta^p(BG)$	Veränderung der Anzahl an BG in Prozent
$\Delta^p(LLU_VT)$	Veränderung der LLU in einem SGB-II-Vergleichstyp (VT) in Prozent
$\Delta^p(LLU_BL)$	Veränderung der LLU in einem Bundesland (BL) in Prozent
i	JC
t	Zeitpunkt
$j(i)$	Nummerierung des SGB-II-Vergleichstyps j zu i
$k(i)$	Nummerierung des Bundeslandes k zu i
ε	Fehlerterm

Wenn die zur Bestimmung der jeweiligen Regressoren erforderlichen Werte ELB, NEF und BG fehlende Werte aufweisen, werden die Werte des Fortschreibungsmodells aus der amtlichen Grundsicherungsstatistik[21] verwendet. In die Berechnung der prozentualen Veränderung der LLU im SGB-II-Vergleichstyp *j(i)* und die prozentuale Veränderung der LLU im Bundesland *k(i)* fließen nur jene JC ein, für die im aktuellen Monat und im Vormonat Daten vorliegen. Das hier vorgeschlagene regressionsbasierte Fortschreibungsmodell für LLU kann als Erweiterung des Fortschreibungsmodells aus der amtlichen Grundsicherungsstatistik verstanden werden. Letzteres nutzt die Veränderungsraten im jeweiligen Vergleichstyp. Für die LLU-Schätzung wird das Modell um die genannten sinnvollen und verfügbaren Prädiktoren erweitert.[22]

Es erfolgt keine Differenzierung der Datenimputationswerte nach dem Personenmerkmal Geschlecht, da die Summe der LLU nicht über Personen, sondern über BG ermittelt wird.

[19] Bei fehlenden Startwerten zu Beginn der Zeitreihe verschiebt sich die erste Schätzung in der Zeitreihe entsprechend.

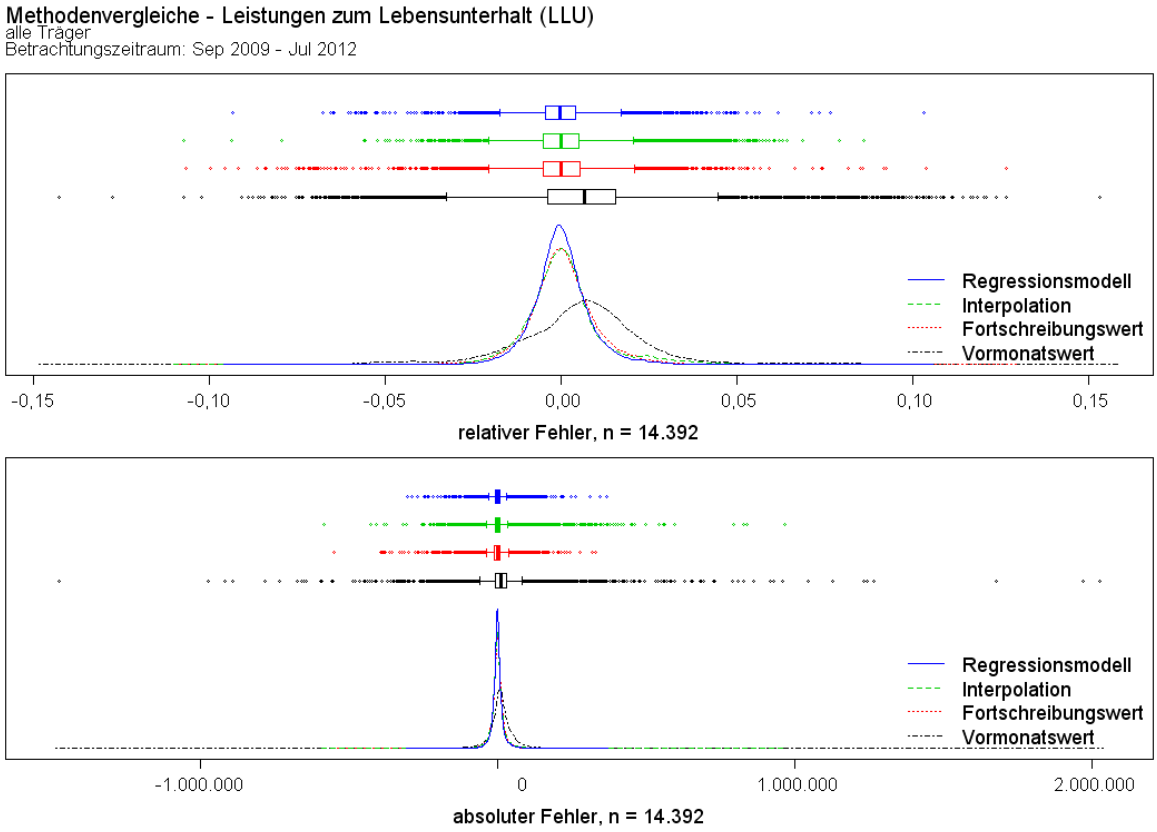
[20] Es erfolgt pro SGB-II-Vergleichstyp eine Regressionsschätzung über alle Zeitpunkte. Es wird angenommen, dass der Fehlerterm $\varepsilon_{t,i}$ die Annahmen für eine Regression nach der Methode der kleinsten Quadrate erfüllt. Das Modell lässt sich als „Fixed-Effect"-Panelregession beschreiben; der Term $\beta_{0,i}$ erfasst JC-spezifische, zeitkonstante Heterogenität. Inhaltlich lässt sich diese JC-spezifische Konstante $\beta_{0,i}$ als Trend in der Entwicklung der LLU des jeweiligen JC interpretieren.

[21] Vgl. Abschnitt 4.1.2 in diesem Dokument sowie Kapitel 4.1 „Eckwerte" im „Handbuch – Grundsicherung für Arbeitsuchende", das die Statistik der BA zum Download zur Verfügung stellt, vgl. https://statistik.arbeitsagentur.de/ > Grundlagen > Datenquellen > Datenstandard XSozial-BA-SGB II > Handbuch.

[22] Die Auswahl des Schätzmodells zur Fortschreibung der LLU erfolgte dabei schrittweise. Für eine ausführliche Erläuterung der schrittweisen Erweiterung siehe Anhang 8.3.

3.1.3 Bewertung[23]

Abbildung 3: LLU – Methodenvergleich Fehlerverteilungen



Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von August 2009 bis August 2012.[24] Im beschriebenen Schätzmodell lassen sich die prozentualen Veränderungen der LLU gut durch die unabhängigen Variablen erklären.[25] Im Ergebnis ist das Regressionsmodell den Standardvergleichsmodellen deutlich überlegen: die Verteilung der absoluten und relativen Fehler weist hier die deutlichste Konzentration bei null auf und starke Ausreißer sind seltener. Das vorgeschlagene Regressionsmodell ist sogar besser als die Interpolation, einem starken Konkurrenten, der als einziges Vergleichsmodell zukünftige, zum Zeitpunkt des Datenausfalls noch unbekannte, Informationen in die Schätzung einbezieht. Die teils enge Verwandtschaft der Regressoren befördert zwar Multikollinearität,[26] zum Beispiel

[23] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[24] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von September 2009 bis Juli 2012.

[25] Der Anteil der erklärten Varianz (R-Quadrat) liegt im Durchschnitt der zwölf Teilmodelle bei über 85 % und ist in keinem Teilmodell geringer als 80 %.

[26] Eine gute Erläuterung des Problems der Multikollinearität im Zusammenhang mit multivariaten Regressionsanalysen findet sich zum Beispiel in Backhaus, K; Erichson, B; Plinke, W.; Weiber, R. (2018): Multivariate Analysemethoden, 15. Auflage, Springer, Seite 98ff.

infolge hoher Korrelation zwischen der ELB- und der BG-Zahl. Etwaige Effekte wie resultierende Schätzungenauigkeiten für die Koeffizienten können in Kauf genommen werden. Denn im vorliegenden Anwendungsfall interessiert weniger, wie gut einzelne Koeffizienten geschätzt werden. Das Verfahren wird hier als prädiktives Modell eingesetzt, um im Einzelfall Schätzwerte möglichst nah an den Ursprungswerten zu erhalten. Folglich werden die gewählten Regressoren aufgrund ihrer hohen inhaltlichen Relevanz im Modell belassen.

Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1, Lesehilfe zu Boxplots) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/- 1,7 %. Dies spricht für einen sehr geringen Schätzfehler im Allgemeinen. Im Median liegt der betragsmäßige Fehler der Regressionsschätzung bei 7.560 Euro bzw. 0,43 %. Insbesondere bei der Betrachtung von Ausreißern zeigt sich die vorgeschlagene Regressionsschätzung überlegen. In lediglich 0,90 % aller Fälle von September 2009 bis Juli 2012 hätte man den tatsächlichen Wert im Betrag um mehr als 3 % verfehlt; in 31 Fällen wäre die betragsmäßige Abweichung größer als 5 %. Beim ansonsten besten Standardvergleichsmodell, der Interpolation, hätte man den tatsächlichen Wert mehr als dreimal so oft, nämlich in 3,06 % der Fälle, im Betrag um mehr als 3 % falsch geschätzt.[27]

3.1.4 Fazit, Ausblick

Mit dem vorgeschlagenen Datenimputationsverfahren werden LLU mit einem sehr geringen mittleren Fehler geschätzt. Nur in vereinzelten Fällen mit oft atypischen Verlaufsmustern, in denen die Entwicklung der LLU offenbar wenig mit den ausgewählten Prädiktoren korreliert, zeigen sich Abweichungen von 5 % oder mehr.

Das Modell bietet mit den vorliegenden Mitteln das bestmögliche Verfahren für die Schätzung fehlender LLU-Werte. Trotz vereinzelter Ausreißer ist von einer noch sehr guten Schätzgüte auszugehen. Mit anwachsender Zeitreihe ist zu prüfen, ob zum Beispiel durch ein rein JC-spezifisches Modell (ohne Berücksichtigung des SGB-II-Vergleichstyps) oder durch eine Verwendung von Daten der k ähnlichsten JC eine verbesserte Schätzung erreicht werden kann.

[27] Auch bei der Betrachtung des Medians des betragsmäßigen Fehlers erweist sich die vorgeschlagene Regressionsschätzung der Interpolation (8.958 Euro bzw. 0,52 %) überlegen.

3.2 Leistungen für Unterkunft und Heizung

3.2.1 Analyse fehlender Werte

Bei Leistungen für Unterkunft und Heizung (LUH) liegen dieselben vereinzelten Datenlücken und derselbe sehr geringe Ausfallanteil von 0,66 % wie bei Leistungen zum Lebensunterhalt (LLU) vor (vgl. Abschnitt 3.1.1).

3.2.2 Schätzmodell

Insgesamt verhält sich die Größe der LUH sehr ähnlich wie die Größe der LLU. Zudem treten dieselben Koppelungen an potenzielle Prädiktoren auf (Veränderungen der Zahl an ELB, NEF und BG sowie Veränderungen der LUH im SGB-II-Vergleichstyp bzw. im Bundesland[28]). Es läge nahe, das LLU-Regressionsmodell für LUH zu übernehmen. Allerdings treten bei LUH Sondereffekte auf: die hier häufigen einmaligen Nachzahlungen führen zu heftigen Ausreißern. LUH liegen bei einzelnen JC in einzelnen Berichtsmonaten deutlich oberhalb des üblichen Niveaus (Veränderungsraten über 30 %). Bei den Analysen erwies sich ein mehrstufiges Schätzverfahren für die fehlenden Werte am besten, das ein Fortschreibungsmodell mit einer linearen Regression kombiniert.[29]

Schritt 1:

Analog zum Schätzverfahren bei LLU (vgl. Abschnitt 3.1.2) wird für LUH ein **regressionsbasiertes Fortschreibungsmodell** (Methode der kleinsten Quadrate) angewendet. Als abhängige Variable fungiert die prozentuale Veränderung der LUH zum Vormonat. Sie wird dann auf den Vormonatswert aufgeschlagen, um fehlende LUH-Werte im Berichtsmonat zu schätzen. Die Zeitpunkte werden systematisch aufsteigend abgearbeitet. Dies bedingt – vorteilhaft im Rahmen amtlicher Berichterstattung – die Festschreibung der Schätzwerte. Die Schätzgleichung lautet:

$$\Delta^p(LUH)_{t,i} = \beta_{0,i} + \beta_1 \cdot \Delta^p(ELB)_{t,i} + \beta_2 \cdot \Delta^p(NEF)_{t,i} + \beta_3 \cdot \Delta^p(BG)_{t,i} + \beta_4 \cdot \Delta^p(LUH_VT)_{t,j(i)} + \beta_5 \cdot \Delta^p(LUH_BL)_{t,k(i)} + \varepsilon_{t,i} \quad (4)$$

[28] Für die Regionaleinteilung böte sich auch die siedlungsstrukturelle Kreistypisierung des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR, 2011) an, hier zusätzlich unterteilt nach Ost- und Westdeutschland. Diese Typisierung dient dem inter- und intraregionalen Vergleich. Kreise bzw. Kreisregionen werden nach Einwohnerzahl, Bevölkerungsdichte und der Lage im siedlungsstrukturellen Regionstyp typisiert. Damit wird berücksichtigt, dass die Lebensbedingungen in den Kreisen bzw. Kreisregionen sowie ihre Entwicklung wesentlich auch von der Entwicklung und der Struktur der jeweiligen Region bzw. des Regionstyps abhängig sind. Die Analysen zeigen jedoch keinen wesentlichen Informationsgewinn gegenüber der sonst verwendeten SGB-II-Typisierung des IAB (vgl. Abschnitt 2.2) – sicherlich weil Informationen zu Kreisstrukturen ja auch in die SGB-II-Typisierung einfließen. Die siedlungsstrukturelle Kreistypisierung des BBSR wird deshalb nachfolgend nicht berücksichtigt. Für eine aktuelle Version der siedlungsstrukturellen Kreistypisierung siehe: https://www.bbsr.bund.de/ > Themen > Raumbeobachtung > Raumabgrenzungen > Deutschland > Siedlungsstrukturelle Kreistypen. Ursprünglich nach: Görmar, W.; Irmen, E. (1991): Nichtadministrative Gebietsgliederungen und -kategorien für die Regionalstatistik. Die siedlungsstrukturelle Gebietstypisierung der BfLR, in: Raumforschung und Raumordnung, 49(6).

[29] Bei fehlenden Startwerten zu Beginn der Zeitreihe verschiebt sich die erste Schätzung in der Zeitreihe entsprechend. Unter anderem liegt bei einem vormaligen Träger in getrennter Aufgabenwahrnehmung (TgAw) zu Beginn der Zeitreihe mit 28 direkt aufeinander folgenden fehlenden Werten ein extrem langer Datenausfall vor (vgl. Abschnitt 3.1.1). Hier wurde sowohl auf die Simulationsanalyse als auch auf die Schätzung verzichtet.

mit:

$\Delta^p(LUH)$	Veränderungen der LUH zum Vormonat in Prozent
β_0 bis β_5	Regressionskoeffizienten
$\Delta^p(ELB)$	Veränderung der ELB in Prozent
$\Delta^p(NEF)$	Veränderung der NEF in Prozent
$\Delta^p(BG)$	Veränderung der Anzahl an BG in Prozent
$\Delta^p(LUH_VT)$	Veränderung der LUH in einem SGB-II-Vergleichstyp (VT) in Prozent
$\Delta^p(LUH_BL)$	Veränderung der LUH in einem Bundesland (BL) in Prozent
i	JC
t	Zeitpunkt
$j(i)$	Nummerierung des SGB-II-Vergleichstyps j zu i
$k(i)$	Nummerierung des Bundeslandes k zu i
ε	Fehlerterm

In der alleinigen Anwendung wäre dieses erste Teilverfahren (bezeichnet mit a) aufgrund der insgesamt relativ stabilen Zeitreihen ein guter Schätzer in der mittleren Tendenz. Im Hinblick auf Ausreißer wäre es dem nachfolgend erläuterten Teilverfahren allerdings unterlegen.

Parallel zu Teilverfahren (a) wird – analog zum Schätzverfahren bei Integrationen (vgl. Abschnitt 6.1.2) – ein **spezielles Fortschreibungsverfahren, ein k-nächste-Nachbarn-Verfahren** (knN-Verfahren), angewendet (bezeichnet mit Teilverfahren b). Hiermit wird die Entwicklung der LUH ähnlicher Nachbarn auf das betreffende JC übertragen. Der Schätzwert ergibt sich als Produkt des Vormonatswertes und der geschätzten mittleren Veränderung zum Vormonat der k ähnlichsten JC. Die Ähnlichkeit wird anhand der Entwicklung der letzten drei verfügbaren Monate und dem Vorjahresmonat – sofern vorhanden – ermittelt. Die optimale Anzahl k der ähnlichsten JC wird dabei monatsweise so bestimmt, dass der mittlere quadratische Fehler (auf Basis des prozentualen Fehlers) minimiert wird.[30] Die Zeitpunkte werden systematisch aufsteigend abgearbeitet. Dies bedingt – wiederum vorteilhaft im Rahmen amtlicher Berichterstattung – die Festschreibung der Schätzwerte.

In der alleinigen Anwendung und in der mittleren Tendenz wäre dieses zweite Teilverfahren (b) dem erstgenannten Teilverfahren (a) unterlegen. Doch fängt die knN-Fortschreibung LUH-typische Sondereffekte einmaliger Nachzahlungen gut ein, die JC-übergreifend in der Zeitreihe zum Tragen kommen. Im Hinblick auf Ausreißer ist sie dem ersten Teilverfahren (a) somit überlegen.

Schritt 2:

Die im ersten Schritt getrennt gewonnenen Schätzwerte für LUH fließen im zweiten Schritt **erneut** in eine **lineare Regression** ein – quasi eine Regression „on top“. Als abhängige Variable fungiert nun die Größe der LUH (und nicht die prozentuale Veränderung der LUH zum Vormonat). Die beiden Prädiktoren sind (a) die mit der regressionsbasierten Fortschreibung gewonnenen LUH-Schätzwerte sowie (b) die mit der knN-Fortschreibung gewonnenen LUH-Schätzwerte. Um möglichen JC-spezifischen Unterschieden gerecht zu werden und zugleich eine ausreichende Zahl von Datenpunkten zur Verfügung zu haben, erfolgt die Regressionsschätzung wiederum pro SGB-II-Vergleichstyp.

Der zweite Schritt produziert tendenziell ähnliche Werte, wie sie bei der regressionsbasierten Fortschreibung in Schritt 1 erzeugt werden. Zugleich werden mit Hilfe des zweiten Prädiktors die Schätzfehler an

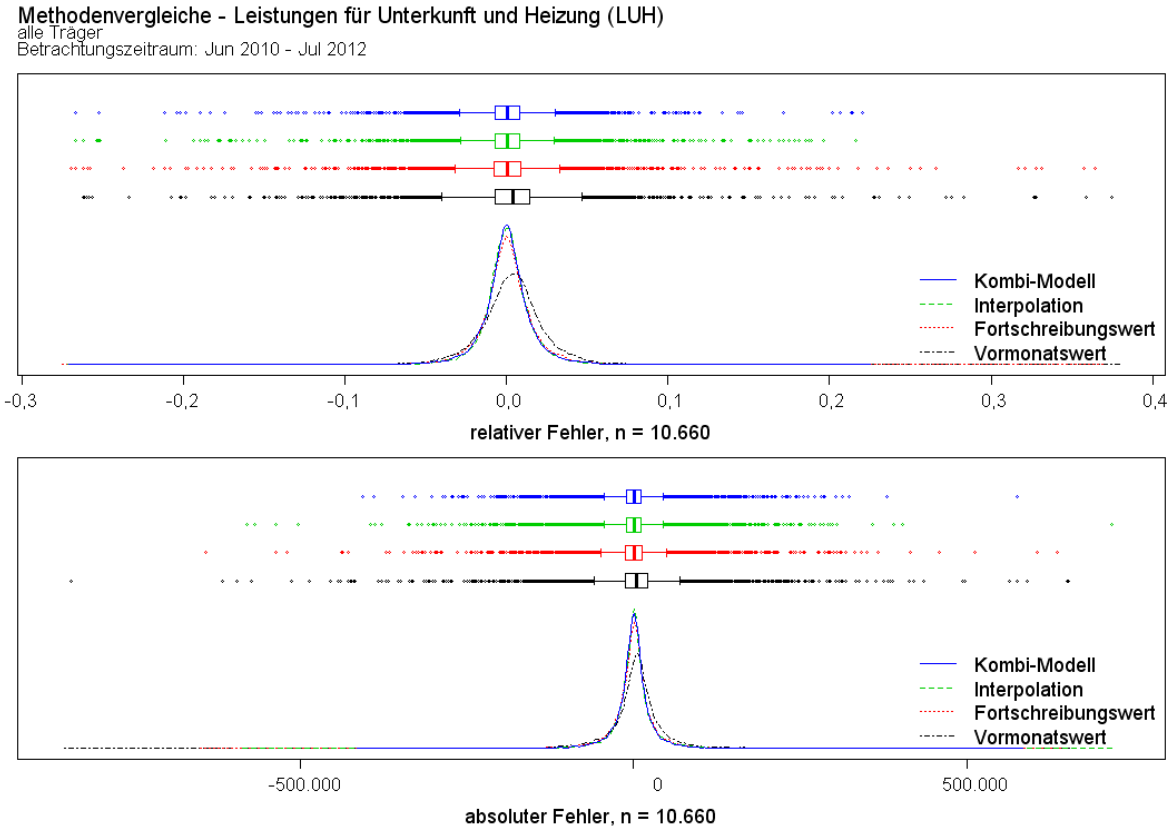
[30] Für eine ausführliche Erläuterung des k-nächste Nachbarn-Verfahrens am Beispielthema Integrationen siehe Anhang 0.

den Extremstellen in den Zeitreihen nivelliert. Somit werden die positiven Aspekte beider Teilverfahren vereint.

Es erfolgt keine Differenzierung der Datenimputationswerte nach dem Personenmerkmal Geschlecht, da die Summe der LUH nicht über Personen, sondern über BG ermittelt wird.

3.2.3 Bewertung[31]

Abbildung 4: LUH – Methodenvergleich Fehlerverteilungen



Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von Januar 2010 bis August 2012.[32] Das kombinierte zweistufige Verfahren (aus der regressionsbasierten und der knN-Fortschreibung zum Einen sowie aus der Regression „on top“ zum Anderen) wird mit den Standardvergleichsmodellen verglichen, die in Abschnitt 2.1.3 beschrieben sind.

[31] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[32] Einheitlich vollständige Daten liegen erst ab Februar 2010 vor. Da das knN-Verfahren aus konzeptuellen Gründen einen 4-monatigen Vorlauf benötigt und die verglichenen Verfahren einen 1-monatigen Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Juni 2010 bis Juli 2012.

Im Ergebnis ist das vorgeschlagene Schätzverfahren lediglich der Interpolation geringfügig unterlegen. Damit bewegt sich das vorgeschlagene Kombimodell in punkto Schätzgüte auf einem sehr hohen Niveau gegenüber einem starken Konkurrenten, der als einziges Vergleichsmodell sogar zukünftige, zum Zeitpunkt des Datenausfalls noch unbekannte, Information in die Schätzung einbezieht. Zudem hat das vorgeschlagene Schätzverfahren gegenüber der Interpolation den großen Vorteil, dass im Produktivbetrieb endgültige imputierte Werte sofort zur Verfügung stehen und nicht erst zukünftige Informationen abgewartet werden müssen, um die Interpolation anwenden zu können.

Die Verteilung der absoluten und der relativen Fehler weist beim vorgeschlagenen Fortschreibungsmodell eine sehr starke Konzentration bei null auf. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1, Lesehilfe zu Boxplots) der relativen Abweichungen enden bei ca. +/- 3,0 %. Dies spricht für einen sehr geringen Schätzfehler im Allgemeinen. Im Median liegt der betragsmäßige Fehler des Modells bei 11.072 Euro bzw. 0,74 %. Insbesondere bei der Betrachtung von Ausreißern zeigt sich die Stärke des vorgeschlagenen Kombimodells. Lediglich in 232 Fällen (2,1 % aller Fälle von Juni 2010 bis August 2012) hätte man den tatsächlichen Wert im Betrag um 5 % oder mehr verfehlt; in 54 Fällen (0,5 % aller Fälle) wäre die betragsmäßige Abweichung größer als 10 %. Somit treten hier höhere Abweichungen vom tatsächlichen Wert auf als bei der vorgeschlagenen LLU-Schätzung. Dies ist bei den LUH-typischen Sondereffekten durch einmalige Nachzahlungen (Veränderungsraten über 30 %) erwartbar und akzeptabel.

3.2.4 Fazit, Ausblick

Aufgrund von LUH-typischen Sondereffekten durch einmalige Nachzahlungen ist es nicht sinnvoll, das vorgeschlagene LLU-Regressionsmodell für LUH unverändert zu übernehmen. Ein mehrstufiges Schätzverfahren, das ein regressionsbasiertes mit einem knN-Fortschreibungsverfahren kombiniert, erweist sich als gut geeignet, fehlende Werte bei LUH zu imputieren. Mit anwachsenden Zeitreihen ist das Modell insbesondere bei Ausreißern auf Saisoneffekte zu überprüfen.

4 Datenimputation zum Themenbereich „Bestandsdaten“

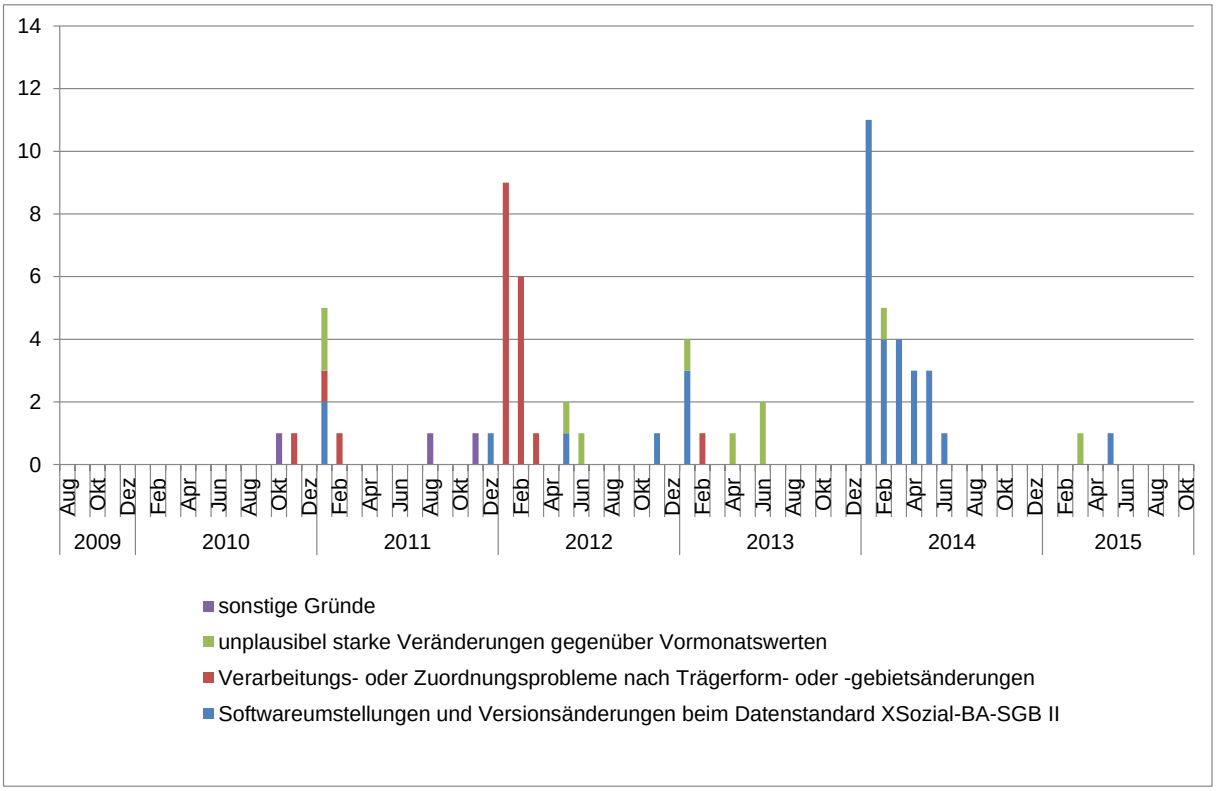
4.1 Bestand an erwerbsfähigen Leistungsberechtigten

4.1.1 Analyse fehlender Werte

Bei der Zahl der erwerbsfähigen Leistungsberechtigten (ELB) kommt es von August 2009 bis Oktober 2015 zu insgesamt 68 Datenlücken. Bezogen auf 75 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,22 %.

Abbildung 5: ELB – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei ELB
Alle JC
Berichtsmonate August 2009 bis Oktober 2015



Die Datenlücken häufen sich vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen eine untergeordnete Rolle. Die Lücken verteilen sich auf 42 JC (etwa 10 % der JC). Nur bei 15 JC fehlen mehrfach Daten für ELB, wobei die maximale Häufigkeit bei fünf hintereinander gereihten Datenlücken für ein JC liegt.

Auch wenn das Ausmaß der Datenlücken bei ELB insgesamt gering scheint – die Konsequenzen sind weitreichend: Die Datenlücken setzen sich ausnahmslos bei allen anderen Grunddaten fort. Wenn für ein JC zu wenig Personendatensätze geliefert werden und die Zahl der ELB nicht plausibel ausgewiesen

werden kann, können auch deren Teilgrößen (zum Beispiel Alleinerziehende, LZB oder Männer und Frauen), abgeleitete Bewegungsgrößen (Zu- und Abgänge), abgeleitete Geldleistungskennzahlen (zum Beispiel Leistungen zum Lebensunterhalt) und Integrationen (definiert als ELB aus dem Vormonat, die eine Erwerbstätigkeit aufgenommen haben) nicht mehr zuverlässig ermittelt werden.

4.1.2 Schätzmodell

Bei der Zahl der ELB kommt es zu Datenlücken, wenn der Wert eines JC in einem Berichtsmonat im Rahmen der regulären Stichtagsvalidierung für die amtliche Grundsicherungsstatistik SGB II als nicht plausibel eingestuft wird. Zur Feststellung der Plausibilität wird die Differenz zwischen dem tatsächlichen ELB-Wert und dem entsprechenden Fortschreibungswert des Fortschreibungsmodells aus der Grundsicherungsstatistik SGB II[33] betrachtet. Kommt es zu der Einstufung „unplausibel“, wird in der amtlichen Berichterstattung stattdessen der Fortschreibungswert als gültige Information verwendet. Dieses Modell wird bereits seit Mitte 2005 im Rahmen der Grundsicherungsstatistik als Imputationsverfahren für die ELB-Zahl angewendet und für die Imputation bei den Kennzahlen nach § 48a SGB II übernommen.

Der Fortschreibungswert ergibt sich als Produkt des Vormonatswertes und der geschätzten mittleren Veränderung zum Vormonat in jenem SGB-II-Vergleichstyp (vgl. Abschnitt 2.2), dem das JC angehört. Das heißt die Entwicklung der ELB-Zahl aus dem Vergleichstyp wird auf das betreffende JC übertragen. Die zugrundeliegende Annahme ist, dass sich die ELB-Zahl von JC desselben Vergleichstyps im Zeitverlauf ähnlich entwickeln. Für die Entwicklung innerhalb eines Vergleichstyps werden nur jene JC berücksichtigt, deren Daten sich im Vergleich zum Vormonat nicht „zu stark“ verändern (Ausschluss von Ausreißern).

Überschreitet die Differenz (zwischen Veränderung für das JC und die Veränderung im Vergleichstyp) einen statistisch definierten Korridor,[34] dann haben sich die Daten des jeweiligen JC anders entwickelt als erwartet. Für die Plausibilitätsprüfung wird geprüft, ob die Veränderungen auf technische bzw. lieferbedingte Fehler zurückzuführen sind. Ist dies der Fall, werden die Werte des JC als nicht plausibel eingestuft. Für Imputationszwecke wird in der Grundsicherungsstatistik – und so auch hier – der entsprechende Fortschreibungswert eingesetzt. Formalisiert lautet das Schätzmodell für fehlende ELB-Werte:

$$FW_i^{VT_{j(i)}} = \frac{\sum_{k \epsilon VT_j} ELB_{k,t} \cdot D_{k,t}}{\sum_{k \epsilon VT_j} ELB_{k,t-1} \cdot D_{k,t}} \cdot ELB_{i,t-1}$$

$$\text{mit } D_{k,t} = \begin{cases} 1 \; wenn \left(Q_{25,t} - 2 \cdot IQA_t\right) \leq \frac{ELB_{i,t}}{ELB_{i,t-1}} \leq \left(Q_{75,t} + 2 \cdot IQA_t\right) \\ 0 \; sonst \end{cases} \tag{5}$$

mit:

FW	Fortschreibungswert
VT	SGB-II-Vergleichstyp
i	JC
t	Zeitpunkt
j	Nummerierung des SGB-II-Vergleichstyps

[33] Vgl. Kapitel 4.1 „Eckwerte“ im „Handbuch – Grundsicherung für Arbeitsuchende“, das die Statistik der BA zum Download zur Verfügung stellt, vgl. https://statistik.arbeitsagentur.de/ > Grundlagen > Datenquellen > Datenstandard XSozial-BA-SGB II > Handbuch

[34] Ausreißer sind in Anlehnung an die Boxplotmethode definiert (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Als Ausreißer gelten hier jene Punkte, die von den Quartilen der Fehlerverteilung weiter als der 2-fache Interquartilsabstand entfernt sind.

$k \epsilon VT_{j(i)}$	Nummerierung der JC aus dem Vergleichstyp j zu i
ELB	erwerbsfähige Leistungsberechtigte
$D_{k,t}$	Dummy; 1 wenn Veränderung im plausiblen Bereich, 0 sonst
Q_{25} bzw Q_{75}	25-%- bzw. 75-%-Quantil (Quartile)
IQA	Interquartilsabstand

Sollte bereits der Vormonatswert nicht plausibel sein, so wird der dafür ermittelte Fortschreibungswert anstelle des für diesen Monat gelieferten Wertes herangezogen. Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

4.1.3 Bewertung[35]

Die Verwendung der SGB-II-Typisierung als Entwicklungsindikator beruht auf der Konstruktion und Ermittlung des SGB-II-Vergleichstyps, der JC mit ähnlicher Arbeitsmarktlage zusammenfasst (vgl. Abschnitt 2.2). Durch die Spannbreite, die der Originalwert vom Fortschreibungswert abweichen kann, wurde das Modell in der Vergangenheit datentechnisch hinreichend abgesichert und der Fortschreibungswert als sehr guter Ersatzwert in der amtlichen Grundsicherungsstatistik SGB II bestätigt.

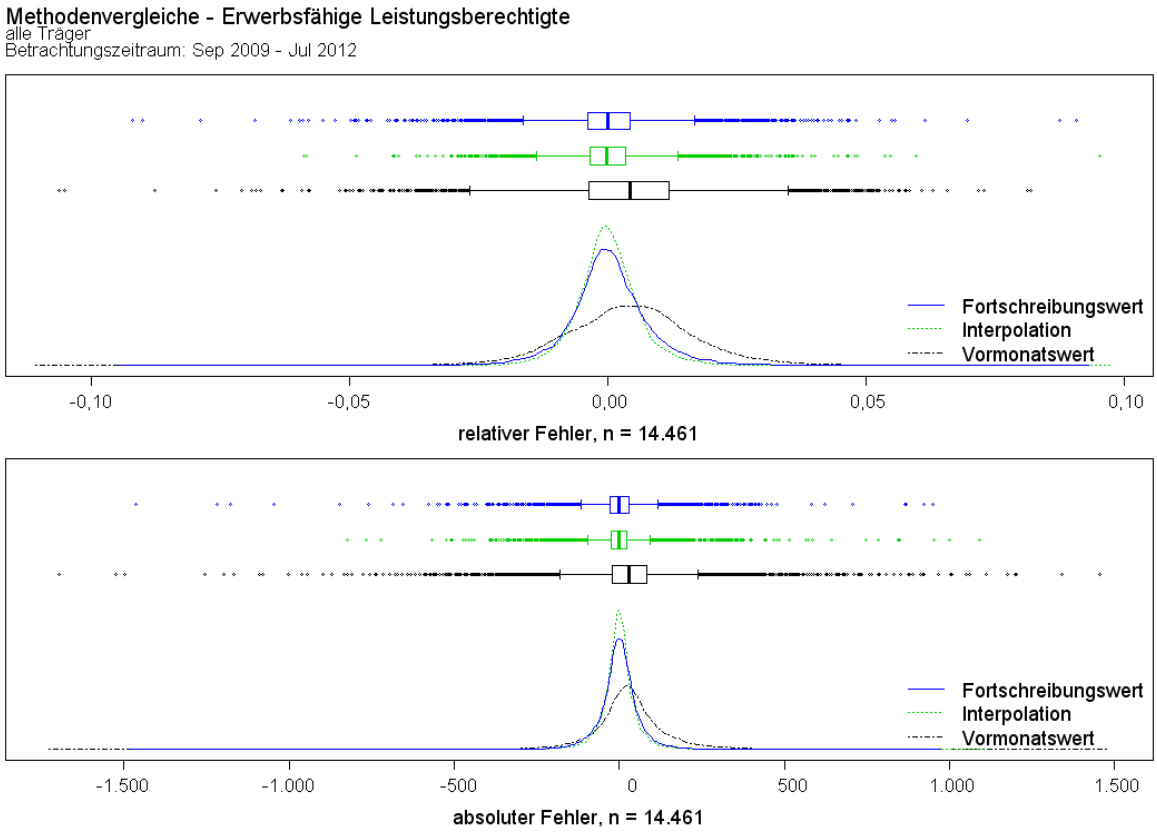
Zwar besteht durchaus das Risiko, dass einmalige Sondereffekte mit diesem Schätzverfahren nicht korrekt ermittelt werden. Jedoch liegt in den meisten Fällen der Anwendung des Verfahrens eine Bestätigung nicht plausibler Werte durch das JC vor. Insofern ist der Fortschreibungswert den mutmaßlich fehlerhaften Originalwerten vorzuziehen.

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von August 2009 bis August 2012.[36] Für die quantitative Bewertung wird das Fortschreibungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Das vorgeschlagene Schätzmodell ist dabei identisch mit dem Vergleichsmodell der Fortschreibung. Im Ergebnis ist das Fortschreibungsmodell lediglich der Interpolation geringfügig unterlegen. Damit bewegt sich das vorgeschlagene Fortschreibungsmodell in punkto Schätzgüte auf einem sehr hohen Niveau gegenüber einem starken Konkurrenten, der als einziges Vergleichsmodell sogar zukünftige, zum Zeitpunkt des Datenausfalls noch unbekannte, Information in die Schätzung einbezieht. Zudem hat das vorgeschlagene Schätzverfahren gegenüber der Interpolation den großen Vorteil, dass im Produktivbetrieb endgültige imputierte Werte sofort zur Verfügung stehen und nicht erst zukünftige Informationen abgewartet werden müssen, um die Interpolation anwenden zu können.

[35] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[36] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von September 2009 bis Juli 2012.

Abbildung 6: ELB – Methodenvergleich Fehlerverteilungen



Dieser Vorteil setzt sich bei Imputationen für andere Grunddaten (zum Beispiel LLU, alleinerziehende ELB) direkt fort, da die Größe der ELB dort als Prädiktor dient. Die Verteilung der absoluten und der relativen Fehler weist beim vorgeschlagenen Fortschreibungsmodell eine sehr starke Konzentration bei null auf. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1, Lesehilfe zu Boxplots) der relativen Abweichungen enden bei ca. +/- 1,6 %. Dies spricht für einen sehr geringen Schätzfehler im Allgemeinen.

Bei der Betrachtung getrennt nach SGB-II-Vergleichstyp ergeben sich für JC des Vergleichstyps 11 (vorwiegend ländliche Gebiete in Ostdeutschland mit unterdurchschnittlicher Arbeitsmarktlage und niedrigem BIP pro Kopf) vergleichsweise leicht erhöhte Schätzfehler. Mit der vorgeschlagenen Schätzmethode treten jedoch nie Schätzfehler über +/- 10,0 % auf. In nur 18 Fällen zeigen sich Abweichungen von über fünf Prozent.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei ELB von sehr guter Qualität ist. Denn die geschlechtsspezifischen Anteilswerte variieren von August 2009 bis August 2012 nur geringfügig. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 0,9 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats sogar nur zwischen 0,3 und 0,9 Prozentpunkten. Die Variabilitätsspannen sind somit als eng anzusehen.

4.1.4 Fazit, Ausblick

Für die Imputation der Bestandsdaten für ELB wird das bisher schon in der amtlichen Grundsicherungsstatistik verwendete und bewährte Verfahren der Fortschreibung genutzt. Damit wird die Übereinstimmung der Daten zwischen amtlicher Statistik und den Kennzahleninformationsquellen sichergestellt. Eine Anpassung würde erfolgen, sobald eine Änderung des Verfahrens in der Grundsicherungsstatistik SGB II vollzogen werden würde.

Mit dem vorgeschlagenen Datenimputationsverfahren wird die Zahl der ELB mit einem sehr geringen mittleren Fehler geschätzt. Nur in vereinzelten Fällen zeigen sich Abweichungen von über 5 %.

4.2 Bestand an alleinerziehenden erwerbsfähigen Leistungsberechtigten

4.2.1 Analyse fehlender Werte

Die Zahl der alleinerziehenden ELB ist eine Teilgröße der ELB insgesamt. Hier liegen dieselben vereinzelten Datenlücken und derselbe sehr geringe Ausfallanteil von 0,22 % vor (vgl. Abschnitt 4.1.1).

4.2.2 Schätzmodell

Analysen zeigen, dass der Anteil von alleinerziehenden ELB an allen ELB in den JC über die Zeit sehr stabil ist. Deswegen erfolgt die Schätzung fehlender Werte bei alleinerziehenden ELB per Anteilswertmethode, die die Anteilsinformation aus der Vergangenheit nutzt.[37] Nutzbar sind sowohl die Informationen aus den vorangehenden Berichtsmonaten als auch aus den geringeren Wartezeitständen als drei Monate. Trotz Unter- bzw. gelegentlicher Übererfassung bei Daten mit verkürzter Wartezeit (WZ<3) ist der Anteil Alleinerziehender bei WZ<3-Daten jenem vorangehender Berichtsmonate in Punkto Stabilität erfahrungsgemäß überlegen. Somit folgt die Schätzung der Alleinerziehenden dem Schema:

Sind die Daten im Referenzmonat unplausibel, so wird der Anteil geschätzt anhand von

1. WZ2-Daten, falls diese auch fehlen
2. WZ1-Daten, falls diese auch fehlen
3. WZ0-Daten, falls diese auch fehlen
4. Vormonatsdaten mit WZ3.

Für die Schätzung der alleinerziehenden ELB wird der letzte plausible Anteilswert (*AW*) am Gesamtbestand je JC *i* herangezogen und mit dem Schätzwert für ELB – also dem jeweiligen Fortschreibungswert der ELB – multipliziert. Der Schätzwert ist der auf Personenanzahl gerundete Wert.

$$FW(allein)_i = AW(allein)_i \cdot FW(ELB)_i \qquad (6)$$

mit:

FW	Fortschreibungswert
AW	Anteilswert

[37] Vgl. Vorgehen nach der Anteilswertmethode bei der Schätzung fehlender Werte nach Geschlecht in Abschnitt 2.3.

$allein$	alleinerziehende erwerbsfähige Leistungsberechtigte
i	JC
ELB	erwerbsfähige Leistungsberechtigte

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

4.2.3 Bewertung[38]

Die Qualität der imputierten Werte hat zwei Facetten: zum einen die Güte des geschätzten Wertes für ELB, die im entsprechenden Abschnitt 4.1.3 als sehr gut und als in der amtlichen Grundsicherungsstatistik SGB II bewährt beschrieben wird. Der zweite Qualitätsaspekt wird durch die Stabilität der Anteilswerte der alleinerziehenden an allen ELB bestimmt. Die Anteilswerte variieren von August 2009 bis August 2012 nur geringfügig. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 0,8 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats sogar nur zwischen 0,3 und 0,9 Prozentpunkten. Die Variabilitätsspannen sind somit als eng anzusehen.

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von August 2009 bis August 2012.[39] Für die quantitative Bewertung wird das Fortschreibungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Im Ergebnis ist das vorgeschlagene Modell den anderen Modellen überlegen: die Verteilung der absoluten und der relativen Fehler weist hier und bei der Interpolation die stärkste Konzentration bei null auf. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1, Lesehilfe zu Boxplots) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/- 1,8 %. Der Median der betragsmäßigen relativen Abweichung beträgt gerade einmal 0,4 %. Dies spricht für einen sehr geringen Schätzfehler im Allgemeinen.

Bei der Einzelbetrachtung stechen einzelne JC als Ausreißer hervor. Mit der vorgeschlagenen Anteilswertmethode treten in 15 Fällen betragsmäßige prozentuale Schätzfehler von über 10 % auf – obwohl die engen Variabilitätsspannen für die Anteilsinformationen solche heftigen Ausreißer nicht erwarten lassen. Hintergrund ist in vielen Fällen, dass infolge einer Gesetzesänderung im Sommer 2010 einzelne JC ihr Meldeverhalten veränderten. In der Folge stiegen hier Anzahl sowie Anteil alleinerziehender ELB.[40] Die Verschiebungen waren jedoch plausibel erklärbar und gingen mit einer Qualitätsverbesserung der gemel-

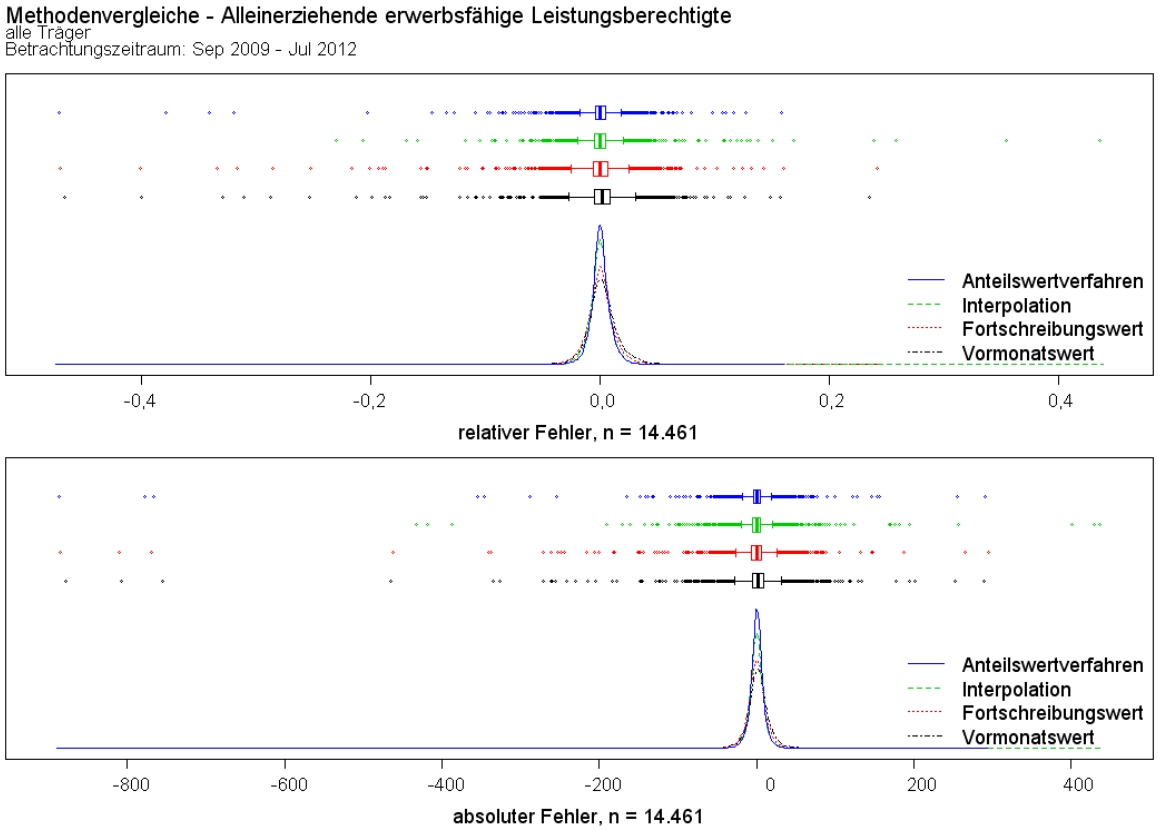
[38] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[39] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von September 2009 bis Juli 2012.

[40] Durch die Einführung des § 24a SGB II (in der bis zum 01.01.2011 geltenden Fassung) bekamen unter anderem auch Kinder, die individuell keinen Leistungsanspruch hatten, einen Anspruch auf zusätzliche Leistungen für die Schule. Infolge der Meldung dieser Kinder verschob sich bei den betroffenen JC die Zahl von Singles hin zu alleinerziehenden ELB.

deten Personendatensätze einher. Da getroffene Plausibilitätsentscheidungen in der amtlichen Grundsicherungsstatistik fixiert sind, kommt es in diesen Fällen zu deutlichen Sprüngen in den Ursprungsdaten. Hier stoßen statistische Imputationsverfahren in der Modellierung zwangsläufig an ihre Grenzen.

Abbildung 7: Alleinerziehende ELB – Methodenvergleich Fehlerverteilungen



Ähnliche Anteilsveränderungen ergaben sich Anfang 2011 im Rahmen einer Trägerfusion, im März 2011 infolge eines Softwarefehlers und im April 2011 infolge einer verschmutzten Datenlieferung mit WZ2. Allen beschriebenen Fallkonstellationen ist gemeinsam, dass sie willkürliche Schwankungen in den Ursprungsdaten hervorrufen, die nicht modellierbar sind. Auch im Bereich extremer Ausreißer sind das vorgeschlagene aufsetzende Anteilswertverfahren und die Interpolation etwa gleich gut. Damit bewegt sich das vorgeschlagene Fortschreibungsmodell in punkto Schätzgüte auf einem sehr hohen Niveau gegenüber einem starken Konkurrenten, der als einziges Vergleichsmodell sogar zukünftige, zum Zeitpunkt des Datenausfalls noch unbekannte, Information in die Schätzung einbezieht. Zudem hat das vorgeschlagene Schätzverfahren gegenüber der Interpolation den großen Vorteil, dass im Produktivbetrieb endgültige imputierte Werte sofort zur Verfügung stehen und nicht erst zukünftige Informationen abgewartet werden müssen, um die Interpolation anwenden zu können.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei alleinerziehenden ELB von sehr guter Qualität ist. Denn die geschlechtsspezifischen Anteilswerte variieren von August 2009 bis August 2012 nur geringfügig. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten

mit WZ3 rund 1,2 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats sogar nur zwischen 0,4 und 1,1 Prozentpunkten. Die Variabilitätsspannen sind somit als eng anzusehen.

4.2.4 Fazit, Ausblick

Für die Datenimputation bei alleinerziehenden ELB wird mit Hilfe eines Anteilswertverfahrens auf das in der amtlichen Grundsicherungsstatistik bewährte Fortschreibungsmodell für die Schätzung der ELB aufgesetzt. Um die Qualität der Schätzgüte bei der Zahl der alleinerziehenden ELB abzusichern, ist die Stabilität der Anteilswerte der alleinerziehenden an allen ELB fortlaufend zu beobachten.

Obwohl das Modell einige Ausreißer produziert, bietet es mit den vorliegenden Mitteln das bestmögliche Verfahren für die Schätzung fehlender Werte. Auf Basis dieses Schätzkonzepts ist von einer guten Schätzgüte auszugehen.

4.3 Bestand an Langzeitleistungsbeziehenden

4.3.1 Analyse fehlender Werte

Langzeitleistungsbeziehende (LZB) sind erwerbsfähige Leistungsberechtigte (ELB), die in den vergangenen 24 Monaten mindestens 21 Monate ELB waren, also Arbeitslosengeld II bezogen haben (Nettodauerbetrachtung). Für die Ermittlung der Personengruppe der LZB werden folglich die Leistungsbezugsdauern der letzten 24 Monate benötigt. Wenn für ein JC mehrfach zu wenig ELB-Datensätze geliefert werden und dadurch zu große Lücken in der Dauermessung entstehen, dann kann die Zahl der LZB nicht mehr zuverlässig ermittelt werden.

Vor diesem Hintergrund erklärt sich, dass es zwei Arten von Datenlücken bei der Zahl der LZB gibt: isolierte Datenlücken einerseits, Langzeitlücken andererseits. Isolierte Datenlücken treten in Berichtsmonaten auf, in denen Datenlücken bei der Zahl der ELB vorliegen. Langzeitlücken entstehen, wenn zurückliegende isolierte Datenlücken infolge ausgefallener Personenbestandsdaten für die Ermittlung der Bezugsdauern und somit für die Zählung der LZB nicht mehr zu kompensieren sind. Dies ist der Fall, wenn während der vergangenen 24 Monate der geschätzte Ausfall von ELB-Bestandsdaten in mehr als einem Monat über 20 % oder in mehr als drei Monaten unter 20 % lag. Diese Toleranzschwelle wurde von der Statistik der BA aus den Analysen der Wirkung von Datenausfällen auf Dauerverteilungen abgeleitet.[41] Unabhängig davon, ob der ELB-Bestand des betroffenen JC wieder vorliegt, kann dann der LZB-Bestand bei einer Langzeitlücke nicht ausgewiesen werden, bis sich die isolierten Datenlücken wieder „ausgewachsen“ haben – im schlechtesten Fall zwei Jahre nach der letzten unplausiblen Fehllieferung. Für die Betrachtung der Veränderungsraten wäre entsprechend ein weiteres Jahr zu warten.

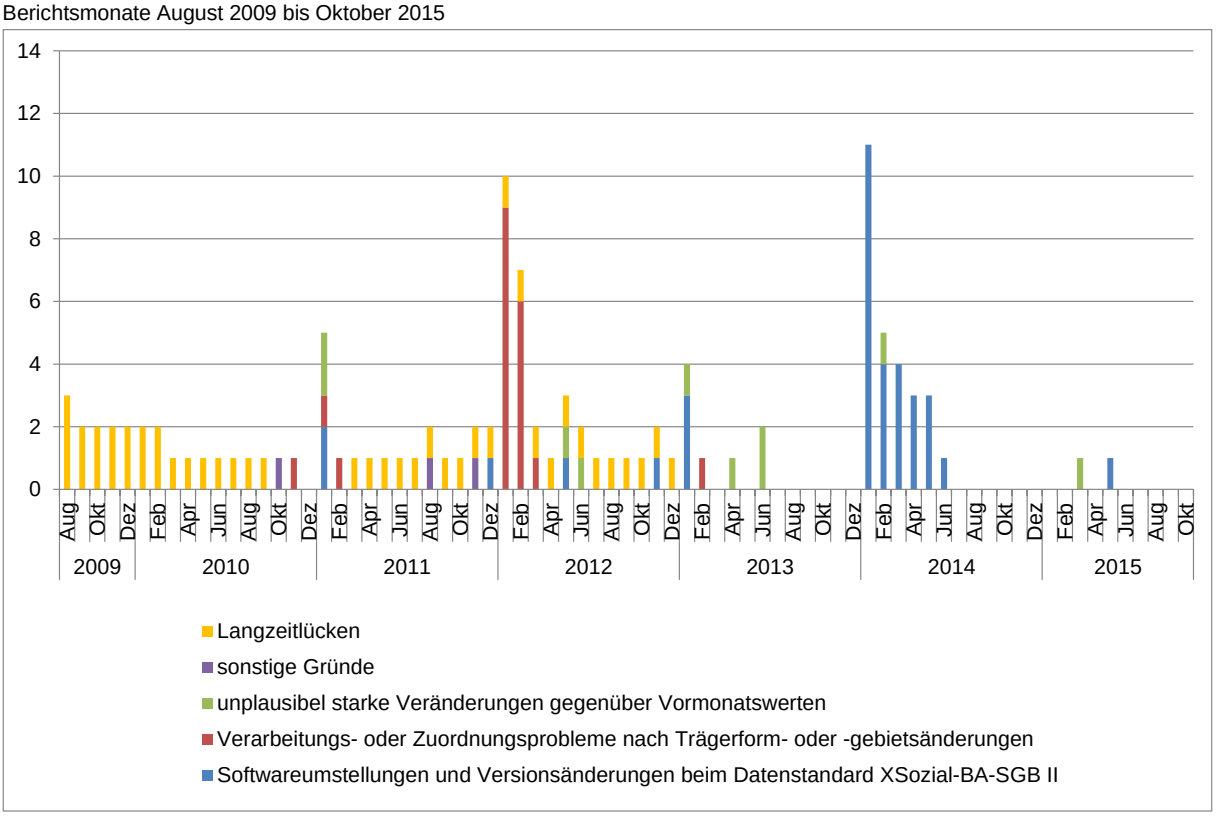
Im Zeitraum von August 2009 bis Oktober 2015 kommt es bei LZB zu insgesamt 112 Datenlücken. Bezogen auf 75 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,37 %. Dabei sind vier JC (knapp 1 %) von den beschriebenen Langzeitlücken betroffen. Diese führen zu 44 fehlenden

[41] Siehe dazu die Ausführungen in den Detailbeschreibungen unter https://www.sgb2.info/ > Kennzahlen > Hilfe und Erläuterungen > Materialien & Downloads -> Detailbeschreibungen.

Berichtsmonatswerten bei LZB. 42 JC (etwa 11 %) sind von isolierten Datenlücken betroffen. Diese häufen sich vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen für isolierte Datenlücken eine untergeordnete Rolle. Dabei fehlen bei 15 JC mehrfach Daten für ELB, wobei die maximale Häufigkeit bei fünf hintereinander gereihten Datenlücken für ein JC liegt. Allerdings führen diese Mehrfachausfälle nie zu den oben genannten Langzeitlücken, da in diesen Konstellationen die Nettodauern für den Leistungsbezug als ELB noch ermittelt werden können.

Abbildung 8: LZB – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei LZB
Alle JC
Berichtsmonate August 2009 bis Oktober 2015



Trotz des geringen Ausmaßes von Datenlücken steht fest: fehlende Daten beim LZB-Bestand insgesamt setzen sich auch bei der Differenzierung nach Geschlecht, bei Bestandsdaten für aktivierte LZB (Abschnitt 4.4) und Integrationen für LZB (Abschnitt 6.3) fort und bringen Datenlücken bei den abgeleiteten Bewegungsdaten (Zu- und Abgänge, Abschnitt 5.2) mit sich.

Zumindest das Ausmaß isolierter Datenlücken bei LZB scheint insgesamt gering. Gleichwohl setzen sich fehlende LZB-Werte auch bei der Differenzierung nach Geschlecht sowie bei allen weiteren Ziel 3-Kenngrößen fort: bei den Integrationen der LZB, bei der Zahl der aktivierten LZB sowie bei den abgeleiteten Bewegungsgrößen (Zu- und Abgänge der LZB). Zudem führen die Datenlücken zu Problemen bei der Berechnung von Jahresfortschrittswerten.

4.3.2 Schätzmodell

Die beiden unterschiedlichen Arten von Datenlücken erfordern verschiedene Herangehensweisen bei der Datenimputation.

4.3.2.1 Schätzmodell bei isolierten Datenlücken

Bei den Analysen erwiesen sich ein Fortschreibungsmodell sowie ein multiples Imputationsmodell im Mittel gleichermaßen geeignet, isolierte fehlende Werte zu schätzen. Das multiple Imputationsmodell produziert allerdings stärkere Ausreißer. Aus diesem Grund wird für die Schätzung bei isolierten Datenlücken das Fortschreibungsmodell gewählt. Dieses entspricht jenem für die Schätzung fehlender ELB-Werte (vgl. Abschnitt 4.1.2), das bisher schon in der amtlichen Grundsicherungsstatistik SGB II verwendet wird und bewährt ist.

Der Fortschreibungswert ergibt sich als Produkt des Vormonatswertes und der geschätzten mittleren Veränderung zum Vormonat in jenem SGB-II-Vergleichstyp (vgl. Abschnitt 2.2), dem das JC angehört. Das heißt die Entwicklung der LZB-Zahl aus dem Vergleichstyp wird auf das betreffende JC übertragen. Die zugrundeliegende Annahme ist, dass sich die LZB-Zahl von JC desselben Vergleichstyps im Zeitverlauf ähnlich entwickeln. Für die Entwicklung innerhalb eines Vergleichstyps werden nur jene JC berücksichtigt, deren Daten sich im Vergleich zum Vormonat nicht „zu stark“ verändern (Ausschluss von Ausreißern).[42]

$$FW_i^{VT_{j(i)}} = \frac{\sum_{k \epsilon VT_j} LZB_{k,t} \cdot D_{k,t}}{\sum_{k \epsilon VT_j} LZB_{k,t-1} \cdot D_{k,t}} \cdot LZB_{i,t-1}$$

$$\text{mit } D_{k,t} = \begin{cases} 1 \; wenn \left(Q_{25,t} - 2 \cdot IQA_t\right) \leq \frac{LZB_{i,t}}{LZB_{i,t-1}} \leq \left(Q_{75,t} + 2 \cdot IQA_t\right) \\ 0 \; sonst \end{cases} \tag{7}$$

mit:

FW	Fortschreibungswert
VT	SGB-II-Vergleichstyp
i	JC
t	Zeitpunkt
j	Nummerierung des SGB-II-Vergleichstyps
$k \epsilon VT_{j(i)}$	Nummerierung der JC aus dem Vergleichstyp j zu i
LZB	Langzeitleistungsbeziehende
$D_{k,t}$	Dummy; 1 wenn Veränderung im plausiblen Bereich, 0 sonst
Q_{25} bzw Q_{75}	25-%- bzw. 75-%-Quantil (Quartile)
IQA	Interquartilsabstand

Sollte auch der Vormonatswert nicht plausibel sein, so wird der dafür ermittelte Fortschreibungswert anstelle des im Statistikverfahren ermittelten unplausiblen Wertes herangezogen. Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

[42] Ausreißer sind in Anlehnung an die Boxplotmethode definiert (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Als Ausreißer gelten hier jene Punkte, die von den Quartilen der Fehlerverteilung weiter als der 2-fache Interquartilsabstand entfernt sind.

4.3.2.2 Schätzmodell bei Langzeitlücken

Das Schätzmodell bei Langzeitlücken ist ein bestandszahlenbasiertes Rekonstruktionsverfahren und im Gegensatz zu allen anderen Modellen in dieser Methodenübersicht kein statistisches Schätzverfahren, das auf den Grundlagen der theoretischen Wahrscheinlichkeitsverteilung beruht. Stattdessen basiert es auf kombinatorischen Abfragen im Auswertungssystem der Statistik der BA. Diese Abfragen werden für das betreffende JC spezifisch auf dessen Ausfallkonstellation hin erstellt und sind auch immer nur für das betreffende JC anwendbar. Ziel ist die näherungsweise Rekonstruktion der LZB-Zahl für das JC, so als hätte es keinen dauermessungsschädlichen Ausfall von Personenbestandsdaten gegeben.

Nach einem solchen Ausfall sind die Bezugsdauern der betroffenen ELB untererfasst. Zwei Auswirkungsformen können unterschieden werden:

- Je länger ELB ausfallen, desto stärker sind ihre Bezugsdauern untererfasst.
- Je mehr ELB ausfallen, bei desto mehr ELB sind die Bezugsdauern untererfasst.

Beide Auswirkungsformen bringen untererfasste LZB-Zahlen mit sich. Das Rekonstruktionsverfahren ergänzt die untererfasste Größe der *LZB* (LZB_u) um eine Korrekturgröße, die Zahl der potenziellen LZB (LZB_{pot}):

$$LZB_{imp} = LZB_u + LZB_{pot} \quad (8)$$

Für ELB aus der Gruppe der LZB_{pot} wird der Schwellwert für die Zugehörigkeit zu den LZB von 21 Monaten um ihre individuell anzunehmende Zahl der Ausfallmonate herabgesetzt – oder umgekehrt ausgedrückt: diese ELB erhalten einen Aufschlag auf ihre gemessene, aber wohl untererfasste Bezugsdauer in Höhe ihrer individuell anzunehmenden Ausfallmonate.

Die entscheidende Frage ist, für welche ELB die Bezugsdauern erhöht werden. Je nach Abgrenzung des Personenkreises kommen Über- und Unterschätzungseffekte in jeweils unterschiedlichem Maße zum Tragen. Würde man die Korrektur der Bezugsdauern auf alle ELB anwenden, die aktuell im Bestand der Leistungsbezieher sind und zugleich in den Ausfallmonaten nicht im Bestand waren, dann würde man die Korrekturgröße der LZB_{pot} gegebenenfalls um jene ELB überschätzen, die während der Ausfallmonate tatsächlich ihren Leistungsbezug unterbrochen haben oder während des Ausfalls oder danach zugegangen sind. Deswegen gilt zusätzlich die Bedingung, die Korrektur der Bezugsdauern nur auf jene ELB anzuwenden, die auch im letzten vollständig gelieferten Berichtsmonat im Bestand der ELB waren. Die zugrundeliegende Annahme ist dabei, dass diese ELB durchgängig im Arbeitslosengeld-II-Bezug waren. Der Überschätzungseffekt verschwindet dadurch zwar nicht vollständig, wird aber verringert. Zugleich wirkt bei der beschriebenen Abgrenzung auf einen ELB-Bestand immer folgender Unterschätzungseffekt: Die Korrekturgröße der LZB_{pot} wird gegebenenfalls um jene ELB unterschätzt, die während der Ausfallmonate tatsächlich in den Leistungsbezug zugegangen sind, aber vom Datenausfall betroffen sind. Insgesamt wirken also zugleich Über- und Unterschätzungseffekte. Analysen zeigen, dass sich durch die beschriebene Abgrenzung des Personenkreises Über- und Unterschätzungseffekte kompensieren können (vgl. nachfolgender Abschnitt zur Bewertung).

Die beschriebene simulative Korrektur – Aufschlag auf die Bezugsdauer für eine spezifische Gruppe an ELB – kann nur durch spezifische Konstruktion der Auswertungen der Statistik der BA erfolgen. Die der Dauermessung zugrundeliegenden statistischen ELB-Bestandsepisoden bleiben auch für ELB der LZB_{pot} unangetastet. Es werden ausschließlich die Auswertungskriterien, wer als LZB zu zählen ist, auf die JC-spezifische Ausfallkonstellation angepasst. Der Personenkreis der LZB_{pot} wird mit Hilfe einer kombinatorischen Abfragelogik ermittelt, deren Eigenschaften direkt von der JC-individuellen Konstellation der Datenlücke abhängen. So steigt die Zahl der kombinatorischen Möglichkeiten mit der Zahl der Ausfallmonate.[43] Prinzipiell ist das Rekonstruktionsverfahren in seiner Grundstruktur allgemeingültig konstruiert. Jedoch hat es die Eigenschaft, dass es zu Beginn einer Langzeitlücke JC-individuell initiiert werden muss – abhängig von der spezifischen Fallkonstellation, also von der Zahl der Ausfallmonate und abhängig davon, ob der Datenausfall zusammenhängend ist oder sich auf mehrere Einzelmonate verteilt. Danach kann das JC-individuelle Vorgehen ohne großen Aufwand monatlich durchgeführt werden. Da es sich um eine statistische Auswertung vorliegender Bestandsdaten mit angepassten Kriterien handelt, sind für die identifizierten Personengruppen (LZB_u und LZB_{pot}) grundsätzlich alle Attribute der Grundsicherungsstatistik (zum Beispiel Geschlecht und andere Merkmale) auswertbar. Entsprechend erfolgt die Differenzierung der Datenimputationswerte nach Geschlecht anhand des gleichen Rekonstruktionsverfahrens wie für die Gesamtgröße der LZB.

4.3.3 Bewertung[44]

Die Bewertung der beiden Schätzmodelle zur Ermittlung von LZB muss für beide Verfahren getrennt erfolgen. Das Schätzmodell bei isoliertem Datenausfall ist ein Fortschreibungsverfahren, also ein statistisches Schätzverfahren, das simulativ mit den üblichen Standardvergleichsverfahren bewertet werden kann. Dagegen kann das Schätzmodell bei Langzeitausfall nur fallindividuellen Analysen unterzogen werden.

4.3.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von August 2009 bis August 2012.[45] Für die quantitative Bewertung wird das Fortschreibungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Das vorgeschlagene Schätzmodell ist dabei identisch mit dem Vergleichsmodell der Fortschreibung. Im Ergebnis ist das Fortschreibungsmodell lediglich der Interpolation geringfügig unterlegen. Damit bewegt sich das vorgeschlagene Fortschreibungsmodell in punkto

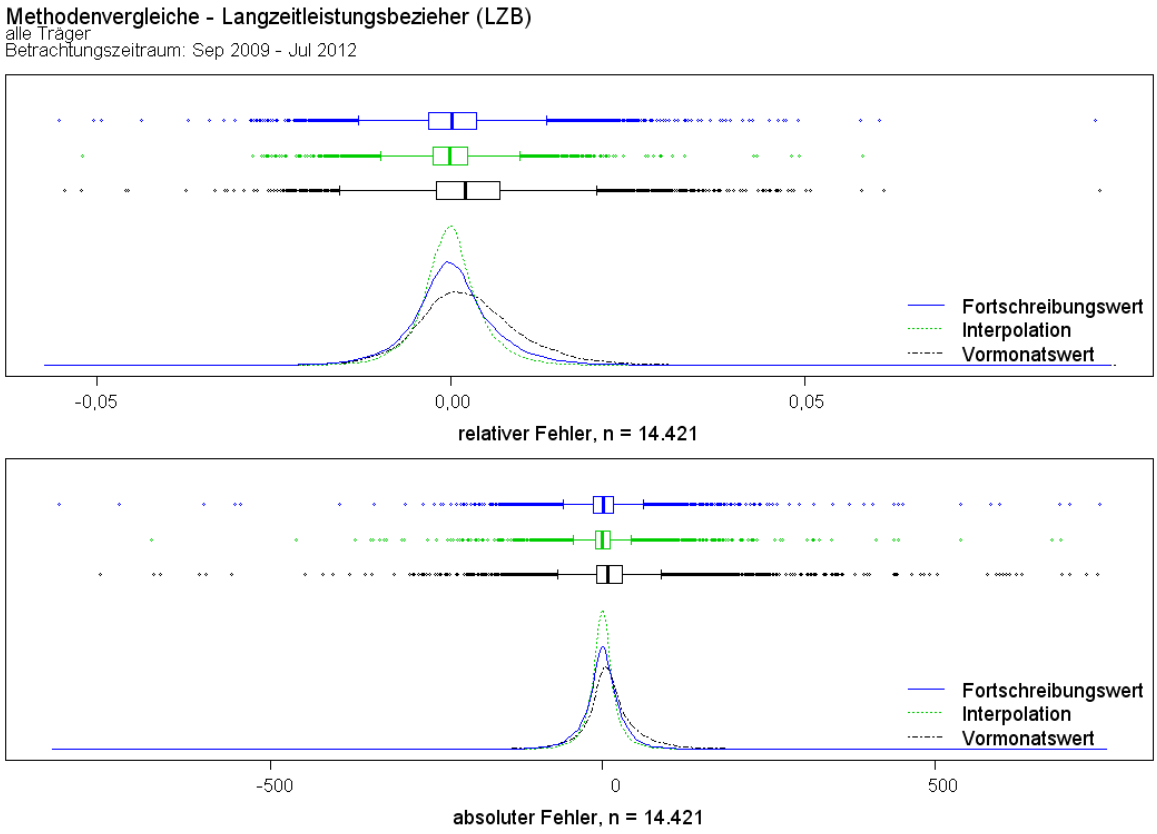
[43] Sei *m* die Zahl der Berichtsmonate mit fehlenden LZB-Werten, dann beträgt die Zahl *k* der kombinatorischen Möglichkeiten: $k = 1/2 \cdot m \cdot 2^m$

[44] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[45] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von September 2009 bis Juli 2012.

Schätzgüte auf einem sehr hohen Niveau gegenüber einem starken Konkurrenten, der als einziges Vergleichsmodell sogar zukünftige, zum Zeitpunkt des Datenausfalls noch unbekannte, Information in die Schätzung einbezieht. Zudem hat das vorgeschlagene Schätzverfahren gegenüber der Interpolation den großen Vorteil, dass im Produktivbetrieb endgültige imputierte Werte sofort zur Verfügung stehen und nicht erst zukünftige Informationen abgewartet werden müssen, um die Interpolation anwenden zu können. Die Verteilung der absoluten und der relativen Fehler weist beim vorgeschlagenen Fortschreibungsmodell eine sehr starke Konzentration bei null auf. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1, Lesehilfe zu Boxplots) der relativen Abweichungen enden bei ca. +/- 1,0 %. Dies spricht für einen sehr geringen Schätzfehler im Allgemeinen.

Abbildung 9: LZB – Methodenvergleich Fehlerverteilungen



Im Median liegt der betragsmäßige Fehler der vorgeschlagenen Fortschreibung für isolierte Datenlücken bei rund 0,009 %. Da die Größe der LZB naturgemäß wenig Variabilität im Zeitablauf aufweist, ist die erzielte hohe Schätzgüte allerdings auch wenig überraschend. Größere prozentuale Abweichungen treten gehäuft bei JC auf, die einen geringen LZB-Anteil haben oder bei denen es strukturelle Änderungen gegeben hat. Mit der vorgeschlagenen Schätzmethode treten jedoch nie Schätzfehler über +/- 6,0 % auf.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei LZB von sehr guter Qualität ist. Denn die geschlechtsspezifischen Anteilswerte variieren von August 2009 bis August 2012 nur geringfügig. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund

0,9 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats sogar nur zwischen 0,2 und 0,8 Prozentpunkten. Die Variabilitätsspannen sind somit als eng anzusehen.

4.3.3.2 Bewertung des Schätzmodells bei Langzeitlücken

Für drei von Langzeitausfällen betroffene JC seit Mitte 2008 wurde die bestandszahlenbasierte Rekonstruktionsmethode historisch initiiert und fallindividuellen Analysen unterzogen. Für die Bewertung des Verfahrens wurde die Rekonstruktion zum Beispiel testweise (a) bei den von Datenausfällen betroffenen JC, jedoch für andere Berichtsmonate als für die tatsächlich ausgefallenen und (b) für andere JC ohne Datenausfall simuliert und anschließend mit den tatsächlichen LZB-Beständen verglichen. Die weiter oben beschriebenen gleichzeitig wirkenden Über- und Unterschätzungseffekte können dabei nicht getrennt voneinander ermittelt werden. Jedoch ist der Gesamteffekt mit Hilfe solcher Einzelfallanalysen messbar. Vor der Einschränkung des Personenkreises auf Personen, die auch im letzten vollständig gelieferten Berichtsmonat im Bestand der Leistungsbezieher waren, wurden bei einem der JC während des 24-monatigen Schätzzeitraums im Extremfall ansteigende Überschätzungen von anfänglich 0,9 % bis 7,0 % am Ende des Schätzzeitraums ermittelt. Infolge der Einschränkung des Personenkreises wurde dieser Fehler im gleichen Zeitraum auf 0,5 % bis 0,9 % reduziert. Die Überschätzung steigt – verfahrensimmanent – am Ende des Schätzzeitraums immer leicht an. Von größeren oder weiteren Schätzfehlern ist nicht auszugehen. Bei Anwendung statistischer Schätzverfahren – welcher Art auch immer – ist insbesondere am Ende des gegebenenfalls extrem langen Schätzzeitraums von 24 Monaten von deutlich größeren Schätzfehlern auszugehen. Insofern wird das vorgeschlagene bestandsbasierte Rekonstruktionsverfahren als sehr gut bewertet, mit gegen Ende des Schätzzeitraums steigender aber akzeptabler Schätzunsicherheit. Ein großer Vorteil des vorgeschlagenen Schätzverfahrens ist, dass das JC-individuelle Verfahren nach seiner initiierenden Konzeption ohne großen Aufwand – mit derselben sehr guten Schätzgüte – auch für Teilgruppen der LZB (nach Geschlecht oder anderen Strukturmerkmalen) angewendet werden kann (vgl. Grunddaten für die Ergänzungsgrößen: Bestand an aktivierten LZB, Abschnitt 4.4.2.2, sowie Integrationen von LZB, Abschnitt 6.3.2.2).[46]

4.3.4 Fazit, Ausblick

Für die Datenimputation bei LZB wird bei isolierten Datenlücken bis zu drei Berichtsmonaten das in der amtlichen Grundsicherungsstatistik bewährte Verfahren der Fortschreibung genutzt. Insgesamt ist von einer sehr guten Schätzgüte auszugehen, die aufgrund der geringen Variabilität der Größe allerdings auch wenig überrascht.

[46] Die Idee, einen Korrekturfaktor zu verwenden, der dieser geringfügigen Überschätzung entgegenwirkt, wurde verworfen. Die Schätzgüte für die Gesamtzahl der LZB hätte sich dadurch geringfügig verbessert, jedoch hätte das Rekonstruktionsverfahren dadurch seine einfache, universelle Verwendbarkeit bei Auswertungen für Teilgruppen der LZB (nach Geschlecht, für Grunddaten der Ergänzungsgrößen) verloren. Weitere alternative Vorgehensweisen, die zugangs- statt bestandszahlenbasiert waren, waren in den Analysen weder wesentlich noch durchgängig überlegen und wurden deshalb ebenfalls verworfen.

Für die Datenimputation bei Langzeitlücken wird ein bestandszahlenbasiertes Rekonstruktionsverfahren genutzt, das JC-individuell angewendet wird. Es wird auf Grundlage der Zeitreihenergebnisse für die betreffenden JC insbesondere angesichts des extrem langen Schätzzeitraums als sehr gut bewertet. Das Verfahren hat zudem den Vorteil, dass es nach seiner initiierenden Konzeption für ein JC ohne großen Aufwand – mit derselben sehr guten Schätzgüte – auch für Teilgruppen der LZB (nach Geschlecht oder anderen Strukturmerkmalen) angewendet werden kann.

4.4 Bestand an aktivierten Langzeitleistungsbeziehenden

4.4.1 Analyse fehlender Werte

Aktivierte Langzeitleistungsbeziehende sind definiert als Langzeitleistungsbeziehende (LZB)[47], die zugleich[48] in einer Maßnahme der aktiven Arbeitsförderung[49] sind. Die Zahl der aktivierten LZB ist also eine Teilgröße der LZB insgesamt. Hier liegen dieselben isolierten Datenlücken und Langzeitlücken vor wie beim Bestand der LZB insgesamt (vgl. Abschnitt 4.3.1). Von September 2010 bis Oktober 2015 liegt ein sehr geringer Ausfallanteil von 0,36 % vor.[50]

Ansonsten liegen aus den Validierungsergebnissen der Förderstatistik der BA für 98 Berichtsmonatswerte Hinweise auf Einschränkungen in der Aussagekraft der Daten zur Arbeitsmarktpolitik (AMP-Daten) vor. In diesen Fällen werden die Werte sowie die daraus abgeleitete Ergänzungsgröße und Jahresfortschrittskennzahl in den Analysewerkzeugen geklammert ausgewiesen.

Insgesamt können bei Datenlücken für aktivierte LZB folgende vier Fallkonstellationen auftreten (detaillierter nachfolgend auch in Abbildung 10):

Fallkonstellation 1: Der LZB-Bestand gilt als nicht plausibel und bei den AMP-Daten gibt es Hinweise auf Einschränkungen in der Aussagekraft.

Fallkonstellation 2: Der LZB-Bestand gilt als nicht plausibel und die AMP-Daten gelten als plausibel.

Fallkonstellation 3: Der LZB-Bestand gilt als plausibel und bei den AMP-Daten gibt es Hinweise auf Einschränkungen in der Aussagekraft.

Fallkonstellation 4: Der LZB-Bestand und die AMP-Daten gelten als plausibel.

Die Fallkonstellationen 1 und 2 führen zu einem Datenausfall bei den aktivierten LZB.

[47] Zum Messkonzept des Langzeitleistungsbezugs vgl. Abschnitt 4.3.1.

[48] Zugleich bedeutet hier: am statistischen Stichtag des Bezugsmonats, vgl. Detailbeschreibungen zu K3E2 unter https://www.sgb2.info > Kennzahlen > Hilfe und Erläuterungen > Materialien & Downloads > Detailbeschreibungen.

[49] Im Sinne der Ergänzungsgröße K3E2 „Aktivierungsquote der Langzeitleistungsbeziehenden nach § 2 Abs. 2 Nr. 1 der geänderten RVO zu Kennzahlen nach § 48a SGB II (15. März 2019).

[50] Die Zeitreihe für aktivierte LZB beginnt aus schnittstellentechnischen Gründen erst im September 2010.

4.4.2 Schätzmodell

Die beiden unterschiedenen Arten von Datenlücken – isolierte Datenlücken einerseits und Langzeitlücken andererseits – erfordern wiederum verschiedene Herangehensweisen bei der Datenimputation.

4.4.2.1 Schätzmodell bei isolierten Datenlücken

Die Daten zur Teilnahme von LZB an arbeitsmarktpolitischen Instrumenten liegen ausschließlich mit dreimonatiger Wartezeit vor. Ein Rückgriff auf Daten mit verkürzter Wartezeit (WZ<3) ist daher nicht möglich. Die Daten unterliegen zudem starken Schwankungen zwischen den Berichtsmonaten. Vor diesem Hintergrund scheidet die Nutzung von Informationen aus dem Vormonat oder Vorjahresmonat zunächst aus.

Analysen haben gezeigt, dass unter bestimmten Bedingungen die ausschnitthaften Informationen der unvollständigen Datenlieferung eines JC für die Schätzung isolierter Datenlücken nutzbar sind. Unvollständige Datenlieferungen sind meistens die Ursache für Eckwerte-Ausfälle und somit für isolierte Datenlücken bei LZB und daraus abgeleiteten (Teil-)Größen. Der komplette Datenausfall – es liegen keine verwertbaren Personendatensätze vor – ist sehr selten. Vertiefte Analysen zeigen nun, dass sich die Personenstrukturen in vielen unvollständigen Datenlieferungen kaum von jenen aus vollständigen Datenlieferungen unterscheiden.

Meist bieten diese unvollständigen Datenlieferungen eine gute Basis, um Imputationswerte mit Hilfe eines Anteilswertverfahrens zu berechnen (Variante a). Dies ist möglich, wenn zwei Bedingungen erfüllt sind:

1. Die isolierte Datenlücke ist auf eine Datenlücke bei LZB insgesamt zurückzuführen.
2. Die unvollständige Datenlieferung stellt statistisch eine Zufallsauswahl aus der unbekannten vollständigen Datenlieferung dar. Man kann also annehmen, dass keine relevante Personengruppe in der unvollständigen Datenlieferung unterrepräsentiert ist.

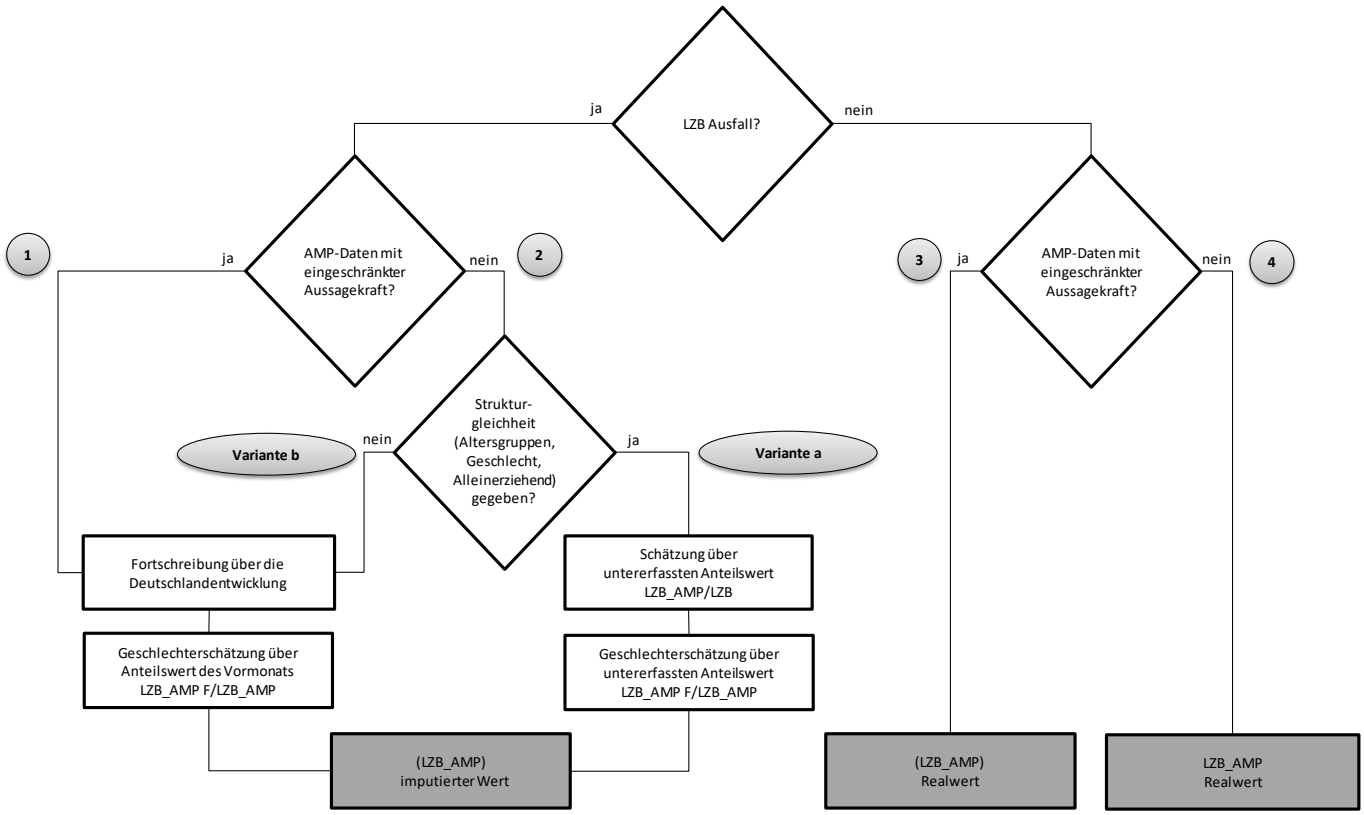
Teststatistische Prüfungen auf Strukturgleichheit bei zentralen Personenmerkmalen zeigen, dass die zweite Bedingung meist zutrifft: bei mehr als drei Viertel der betreffenden isolierten Datenlücken kann von einem zufälligen Datenausfall ausgegangen werden.[51] Liegt keine Strukturgleichheit vor, wird der Bestand an aktivierten LZB in der „Rückfall-Lösung“ mit Hilfe eines Fortschreibungsmodells geschätzt, das auf Vormonatsinformationen aufsetzt (Variante b).

Der Entscheidungsbaum in Abbildung 10 stellt die Prüfschritte dar, anhand derer das Schätzmodell festgelegt wird. Bei Fallkonstellation 1 (LZB-Bestand mit Ausfall, AMP mit eingeschränkter Aussagekraft), werden die Vormonatswerte fortgeschrieben (Variante b). Ist der Bestand an LZB unplausibel, die Förderdaten aber unauffällig (Fallkonstellation 2), wird zunächst getestet, ob der Datenausfall zufällig ist. Trifft dies zu, wird die Annahme getroffen, dass auch der Ausfall der aktivierten LZB zufällig ist. Diese können dann mittels Anteilswertverfahren geschätzt werden (Variante a). Liegt keine Strukturgleichheit vor, wird

[51] Diese Teststatistik beruht auf Daten von August 2009 bis August 2013 (vgl. Version 3.0 dieser Methodenübersicht, Stand Mai 2014) vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Damals lag bei 32 der betreffenden 38 isolierten Datenlücken (84 %) Strukturgleichheit vor. Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

Variante (b) angewendet. Liegen nur bei den Förderdaten Hinweise auf Einschränkungen in der Aussagekraft vor (Fallkonstellation 3), werden die Daten berichtet und als eingeschränkt aussagefähig – mit Klammern – gekennzeichnet. Bei der Fallkonstellation 4 liegen keine Einschränkungen vor und die Realwerte der aktivierten LZB werden ohne Einschränkung berichtet.

Abbildung 10: Aktivierte LZB – Schätzmodell/Entscheidungsbaum



4.4.2.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren

Bei Fallkonstellation 2 wird im ersten Schritt getestet, ob die Anteile wichtiger Personenmerkmale an allen LZB in der unvollständigen Datenlieferung des JC mit den entsprechenden Anteilen vergleichbar sind, die in den Monatsdaten der vorangegangenen 36 Berichtsmonate des jeweiligen JC beobachtet wurden.[52]

[52] Nach einer gebräuchlichen Faustregel sollte für statistische Tests die Mindestfallzahl 30 nicht unterschritten werden. Rundet man noch zum vollen Dreijahreszeitraum auf, kommt man auf 36 Monate.

Mit der im August 2009 aufgenommenen Berichterstattung für LZB existiert für Berichtsmonate bis einschließlich August 2012 kein zurückliegender 36-Monatszeitraum. Um das Verfahren in diesem Zeitraum dennoch nutzen zu können, werden zur Vervollständigung zunächst auch Datenpunkte herangezogen, die zeitlich nach dem Datenausfall liegen. Für Datenlücken ab September 2012 wächst der 36-Monatszeitraum sukzessive aus dem initialen Zeitfenster heraus, um dann rollierend mit den Schätzzeitpunkten mitzulaufen.[53]

Für den Test auf Strukturgleichheit werden die folgenden Strukturmerkmale zur Charakterisierung der Grundgesamtheit berücksichtigt:

- Geschlecht in den Ausprägungen „männlich“ oder „weiblich“
- Alter in den Ausprägungen
  - 15[54] bis unter 25 Jahre
  - 25 bis unter 50 Jahre
  - 50 bis 65 Jahre
- Alleinerziehend in den Ausprägungen „nein“ oder „ja“

Für diese Merkmale werden die jeweiligen Anteile am gesamten LZB-Bestand im Monat des Datenausfalls und dem wie oben definierten 36-Monatszeitraum ermittelt. Auf Basis der plausiblen Anteile im 36-Monatszeitraum werden für jedes einzelne JC die Verteilungsparameter arithmetischer Mittelwert (*MW*) und Standardabweichung (*s*) der Strukturmerkmale berechnet und geprüft, ob der entsprechende Anteil der unvollständigen Datenlieferung in einem Vertrauensintervall zwischen Mittelwert plus bzw. minus drei Standardabweichungen [*MW +/- 3 s*] liegt.[55]

Befinden sich die Anteile von mindestens zwei Strukturmerkmalen außerhalb des Intervalls, wird die Hypothese der Strukturgleichheit abgelehnt und Variante (b) zur Berechnung der imputierten Werte der aktivierten LZB verwendet. Liegt kein oder nur ein Strukturwert außerhalb des Intervalls, wird angenommen, dass Strukturgleichheit vorliegt und der beobachtete Ausfall zufällig ist. Die Anteilsverteilungen aus der

[53] Im initialen Zeitfenster kommt es infolge von Datenlücken zu unvollständigen 36-Monatszeiträumen. Im Anwendungsfall wird die Grenze der Faustregel aber nur ein einziges Mal mit 29 Berichtsmonaten knapp unterschritten.

[54] Ein ELB wird erst als LZB gezählt, wenn er in den letzten 24 Monaten mindestens 21 Monate (1 ¾ Jahre) lang als ELB im Bestand war. Empirisch sind LZB also mindestens 16 ¾ Jahre alt. Für den hier umrissenen Strukturgleichheitstest ist dies jedoch nicht bedeutsam.

[55] Unterstellt man, dass die Anteile der betrachteten Strukturmerkmale normalverteilt sind, dann liegen – nach einer gebräuchlichen Faustformel in der Statistik – 99,7 % der Fälle innerhalb des Intervalls [MW +/- 3 s] und außerhalb nur deutliche Ausreißer (vgl. auch die grafische Darstellung in Anhang 8.1, Lesehilfe zu Boxplots).

unvollständigen Datenlieferung gelten somit als vertrauenswürdige Annäherung an die Verteilungen, die sich aus einer plausiblen, vollständigen Datenlieferung des JC ergeben hätten.

Die Datenimputation erfolgt nun als einfaches Anteilswertverfahren: Der jeweilige Schätzwert für aktivierte LZB ergibt sich als Produkt aus dem per Fortschreibungsmodell geschätzten LZB-Bestand (vgl. Abschnitt 4.3.2.1) und dem Anteil der aktivierten LZB an allen LZB, der aus der unvollständigen Datenlieferung stammt. Bei diesem relativ einfachen Verfahren werden die vorhandenen JC-spezifischen Informationen so weit wie möglich einbezogen.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt für Frauen nach demselben Verfahren wie für die Gesamtschätzung. Die Anzahl der Männer ergibt sich als Differenz zwischen dem Gesamtwert und demjenigen der Frauen.[56]

4.4.2.1.2 Variante b) Fortschreibung mit Hilfe der Deutschlandentwicklung

Wenn die Strukturen nicht gleich sind oder aber der LZB-Bestand unplausibel ist und die AMP-Daten in ihrer Aussagekraft eingeschränkt sind, wird in der „Rückfall-Lösung“ der Anteil der aktivierten LZB mittels Fortschreibung geschätzt. Analysen der Bestandsveränderungen bei aktivierten LZB zeigen teilweise saisonale Veränderungsmuster, die aber keine Ähnlichkeit innerhalb der jeweiligen SGB-II-Vergleichstypen aufweisen. Die Typisierung bietet hier als Prädiktor demnach keine Verbesserung der Schätzgüte. Deswegen wird für die Fortschreibung des Bestandes an aktivierten LZB die Veränderung des Bestands an aktivierten LZB auf Bundesebene herangezogen. Diesem Fortschreibungsmodell liegt somit die Annahme zugrunde, dass sich der Bestand an aktivierten LZB in den von isolierten Datenausfällen betroffenen JC entsprechend dem Bundesdurchschnitt entwickelt hat.

Für die Abbildung der Entwicklung auf Deutschlandebene werden nur jene JC berücksichtigt, deren Daten sich im Vergleich zum Vormonat nicht „zu stark“ verändern (Ausschluss von Ausreißern).[57] Formal lässt sich dieses Verfahren folgendermaßen beschreiben:

$$FW_i^{DE_i} = \frac{\sum_{k \epsilon DE} lzb_amp_{k,t} \cdot D_{k,t}}{\sum_{k \epsilon DE} lzb_amp_{k,t-1} \cdot D_{k,t}} \cdot lzb_amp_{i,t-1}$$

$$\text{mit } D_{k,t} = \begin{cases} 1 \; wenn \left(Q_{25,t} - 2 \cdot IQA_t\right) \leq \frac{lzb_amp_{i,t}}{lzb_amp_{i,t-1}} \leq \left(Q_{75,t} + 2 \cdot IQA_t\right) \\ 0 \; sonst \end{cases} \tag{9}$$

mit:

FW	Fortschreibungswert
DE	Deutschlandebene
i	JC
t	Zeitpunkt
j	Nummerierung des SGB-II-Vergleichstyps
$k \epsilon DE$	Nummerierung der JC in Deutschland

[56] Da die Grunddaten für aktivierte LZB ausschließlich mit dreimonatiger Wartezeit vorliegen, wird hier vom üblichen Muster bei der Geschlechterschätzung (vgl. Abschnitt 2.3) abgewichen. Im Vertrauen auf den Test auf Strukturgleichheit ist dieses Vorgehen aber sachlogisch nur konsequent.

[57] Wie schon beim Fortschreibungsmodell zur Schätzung des LZB-Bestandes werden Ausreißer in Anlehnung an die Boxplotmethode definiert (vgl. Kapitel 8.1, Lesehilfe zu Boxplots). Als Ausreißer gelten hier jene Punkte, die von den Quartilen der Fehlerverteilung weiter als der 2-fache Interquartilsabstand entfernt sind.

lzb_amp	aktivierte Langzeitleistungsbeziehende
$D_{k,t}$	Dummy; 1 wenn Veränderung im plausiblen Bereich, 0 sonst
Q_{25} bzw Q_{75}	25-%- bzw. 75-%-Quantil (Quartile)
IQA	Interquartilsabstand

Sollte bereits der Vormonatswert nicht plausibel sein, so wird der dafür ermittelte Fortschreibungswert anstelle des für diesen Monat gelieferten Wertes herangezogen. Die Grunddaten für aktivierte LZB liegen ausschließlich mit dreimonatiger Wartezeit vor. Deshalb springt das übliche Verfahren für die Geschlechterschätzung (vgl. Abschnitt 2.3) immer direkt auf die Anteilswertinformation des Vormonats.

4.4.2.2 Schätzmodell bei Langzeitlücken

Das Vorgehen bei Langzeitlücken ist dasselbe wie bei Datenimputationen für Langzeitlücken bei LZB insgesamt (vgl. Abschnitt 4.3.2.2). Nach der initiierenden Konzeption der kombinatorischen Abfragelogik kann das JC-individuelle Vorgehen ohne großen Aufwand monatlich durchgeführt und auch für die Teilgruppe der aktivierten LZB angewendet werden. Entsprechend erfolgt die Differenzierung der Datenimputationswerte nach Geschlecht anhand des gleichen Rekonstruktionsverfahrens wie für die Gesamtgröße.

4.4.3 Bewertung[58]

4.4.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken

4.4.3.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren

Zur Optimierung der Auswahlkriterien für die Strukturmerkmale, des Zeitintervalls und des Entscheidungskriteriums wurden mehrere Tests durchgeführt, um das Modell auf Treffsicherheit zu überprüfen.

1. Für JC mit isolierten Lücken wurden die Vertrauensintervalle in unterschiedlichen Zeiträumen bestimmt. Hieraus kann abgeleitet werden, inwiefern der Faktor Zeit die Verteilungen beeinflusst.
2. Für das Alter wurden unterschiedliche Klassengrenzen betrachtet. Mit zu schmalen und somit vielen Klassen (zum Beispiel 5-Jahresintervall) wird die Hypothese auf Strukturgleichheit in vielen Fällen abgelehnt, obwohl sie zutrifft – sogenannter Fehler erster Art. Mit zu breiten und entsprechend wenigen Klassen (zum Beispiel 30-Jahresintervall) wird der Hypothese auf Strukturgleichheit zu oft zugestimmt, obwohl sie nicht zutrifft – sogenannter Fehler zweiter Art.
3. Die Breite des Vertrauensintervalls wurde variiert: Neben dem 99,7-Prozent-Intervall [*MW +/- 3 s*] erfolgte die Testung auf Strukturgleichheit zusätzlich für das 99-Prozent-Intervall [*MW +/- 2,576 s*] und das 95-Prozent-Intervall [*MW +/- 1,96 s*]. Diese zusätzlichen Varianten müssten zu mehr Fehlern erster Art führen.

[58] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

Bei allen Varianten wurde immer bei denselben sechs isolierten Lücken in sechs JC die Hypothese auf Strukturgleichheit abgelehnt, d. h. es lagen immer mindestens zwei Strukturmerkmale außerhalb der festgelegten Intervalle. Die gewählte Variante scheint also zweckmäßig und robust.

Als weiterer Schritt wurden Simulationen durchgeführt, bei denen – nacheinander über alle JC hinweg – in einem spezifischen Monat eine Datenlücke angenommen wurde. Die Simulationen haben gezeigt, dass ein einzelnes Strukturmerkmal als Ablehnungskriterium für die Strukturgleichheit nicht genügt. Ein Einzelkriterium kann gewissen Schwankungen und Trends unterliegen, so dass die Anteilsinformationen am aktuellen Rand fälschlicherweise als „Ausreißer“ erscheinen können („Randeffekt“). Die Simulationen zeigten auch, dass bei der Anwendung von zwei Strukturmerkmalen als Ablehnungskriterien nur 0,5 % der Fälle fälschlicherweise ausschlagen würden (Fehler erster Art). Davon waren maximal zwei JC pro Simulation bzw. Monat betroffen. Insgesamt ergaben die Simulationen mit den Echtdaten, dass die ausgewählten Strukturmerkmale die Grundgesamtheit hinreichend genau beschreiben.

Insgesamt erweist sich der Test auf Strukturgleichheit als sehr treffsicher. Es ist davon auszugehen, dass das vorgeschlagene Verfahren gute Schätzergebnisse bei maximaler Ausnutzung der vorliegenden Informationen des jeweiligen JC bietet. Nach dem vorgeschalteten Test auf Strukturgleichheit wird dieses Verfahren im Zeitraum zwischen September 2010 und August 2013 in den allermeisten Fällen – bei 32 von 39 isolierten Datenlücken – eingesetzt. Der Schätzfehler, der nicht näher quantifizierbar ist, dürfte geringfügig unter oder über dem Schätzfehler der Imputation des LZB-Bestands bei isolierten Datenlücken liegen, auf dem das Anteilswertverfahren aufsetzt.

Da die Differenzierung der Datenimputationswerte nach Geschlecht – abweichend vom üblichen Muster bei der Geschlechterschätzung (vgl. Abschnitt 2.3) – analog zum Verfahren für die Gesamtschätzung erfolgt, ist hier von der gleichen guten Schätzgüte auszugehen – eventuell leicht abgeschwächt infolge geringer Fallzahlen.

Alle potenziellen Stellschrauben scheinen gut gewählt. Das so definierte Testverfahren auf Strukturgleichheit erweist sich damit auch für zukünftige isolierte Lücken als sinnvolles Kriteriensystem für die sichere Anwendung des Anteilswertverfahrens.

4.4.3.1.2 Variante b) Fortschreibung mit Hilfe der Deutschlandentwicklung

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von September 2010 bis August 2013.[59] Für die quantitative Bewertung wird das Anteilswertmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen.

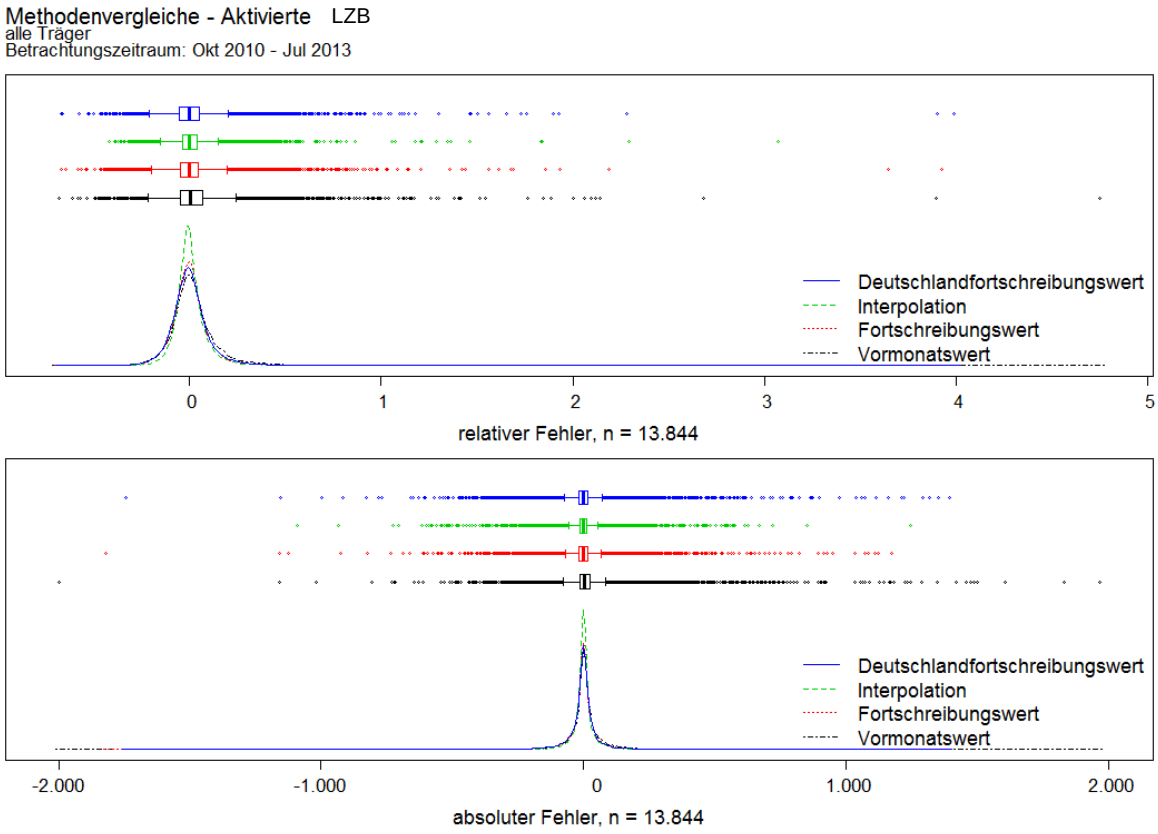
Im Ergebnis ist das Fortschreibungsmodell mit dem Deutschlandwert gleichauf mit dem Fortschreibungsmodell anhand der SGB-II-Typisierung, wie es für ELB-Bestände (vgl. Abschnitt 4.1.2) und LZB-Bestände bei isolierten Datenlücken (vgl. Abschnitt 4.3.2.1) angewendet wird. Die Verteilungen der Schätzfehler

[59] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Oktober 2010 bis Juli 2013.

ähneln sich in hohem Maße. Dies ist ein Hinweis darauf, dass die SGB-II-Typisierung hier keinen wesentlichen Beitrag zur Verbesserung der Schätzgüte leisten kann. Die Aufgliederung nach Vergleichstypen wird deswegen verworfen.

Das vorgeschlagene Anteilswertverfahren ist dem Interpolationsverfahren deutlich unterlegen. Es verliert gegen einen starken Konkurrenten, der als einziges Vergleichsmodell sogar zukünftige, zum Zeitpunkt des Datenausfalls noch unbekannte, Information in die Schätzung einbezieht. Gegenüber der Interpolation hat das vorgeschlagene Schätzverfahren den großen Vorteil, dass im Produktivbetrieb endgültige imputierte Werte sofort zur Verfügung stehen und nicht erst zukünftige Informationen abgewartet werden müssen, um die Interpolation anwenden zu können.

Abbildung 11: Aktivierte LZB – Methodenvergleich Fehlerverteilungen bei der Fortschreibung mit Hilfe der Deutschlandentwicklung



Die Verteilung der absoluten und der relativen Fehler weist beim vorgeschlagenen Fortschreibungsmodell eine mäßige Konzentration bei null auf. Im Median liegt der betragsmäßige Fehler des Modelles bei 18 aktivierten LZB bzw. 5,2 %. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1, Lesehilfe zu Boxplots) der relativen Abweichungen enden bei ca. +/- 21 %. Dies spricht für einen merklichen Schätzfehler im Allgemeinen.

Alle verglichenen Verfahren weisen sehr hohe Schätzfehler auf. Diese sind jedoch insbesondere durch die Datenstruktur bedingt. Es treten JC-individuell Sprünge in den Ursprungsdaten ohne ausgeprägte zeitliche Muster auf, zum Beispiel von 20 auf 140 aktivierte LZB im Nachfolgemonat. Hier stoßen statistische

Imputationsverfahren in der Modellierung zwangsläufig an ihre Grenzen. Insgesamt erscheint die mäßige Schätzgüte angesichts der seltenen Anwendung akzeptabel.

Mangels Grunddaten mit verkürzter Wartezeit springt das übliche Verfahren für die Geschlechterschätzung (vgl. Abschnitt 2.3) immer direkt auf die Anteilswertinformation des Vormonats. Es ist davon auszugehen, dass die Werte zumindest in dieser Hinsicht eine befriedigende Schätzgüte aufweisen, denn die geschlechtsspezifischen Anteilswerte variieren von September 2011 bis August 2013 nur mäßig: die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 12 Prozentpunkte.

4.4.3.2 Bewertung des Schätzmodells bei Langzeitlücken

Das vorgeschlagene bestandsbasierte Rekonstruktionsverfahren wird auch für die Teilgruppe der aktivierten LZB als sehr gut bewertet, mit gegen Ende des Schätzzeitraums steigender aber akzeptabler Schätzunsicherheit. Bei Anwendung statistisch-mathematischer Schätzverfahren – welcher Art auch immer – wäre insbesondere am Ende des gegebenenfalls extrem langen Schätzzeitraums von 24 Monaten von deutlich größeren Schätzfehlern auszugehen.

4.4.4 Fazit, Ausblick[60]

Für die Datenimputation bei isolierten Datenlücken wird in den meisten Fällen (bislang 32-mal) ein Anteilswertverfahren nach Test auf Strukturgleichheit angewendet. Hier ist von einer guten Schätzgüte auszugehen. Die verbleibenden wenigen isolierten Datenlücken (bislang 7) werden mit einer Fortschreibung über die Deutschlandentwicklung geschlossen. Hier ist die Schätzqualität nur mäßig, was aber angesichts der wenigen Schätzpunkte akzeptabel erscheint.

Für die Datenimputation bei bislang 23 Langzeitlücken wird ein bestandszahlenbasiertes Rekonstruktionsverfahren genutzt, das JC-individuell angewendet wird. Es wird angesichts des extrem langen Schätzzeitraums als sehr gut bewertet.

[60] Die in diesem Abschnitt genannten Anwendungshäufigkeiten der verschiedenen Datenimputationsmethoden für Datenlücken bei aktivierten LZB beziehen sich auf Daten bis August 2013 (vgl. Version 3.0 dieser Methodenübersicht, Stand Mai 2014) vor der Revision der Grundsicherungsstatistik SGB II (April 2016), und somit auf 39 isolierte Datenlücken und 23 Langzeitlücken. Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

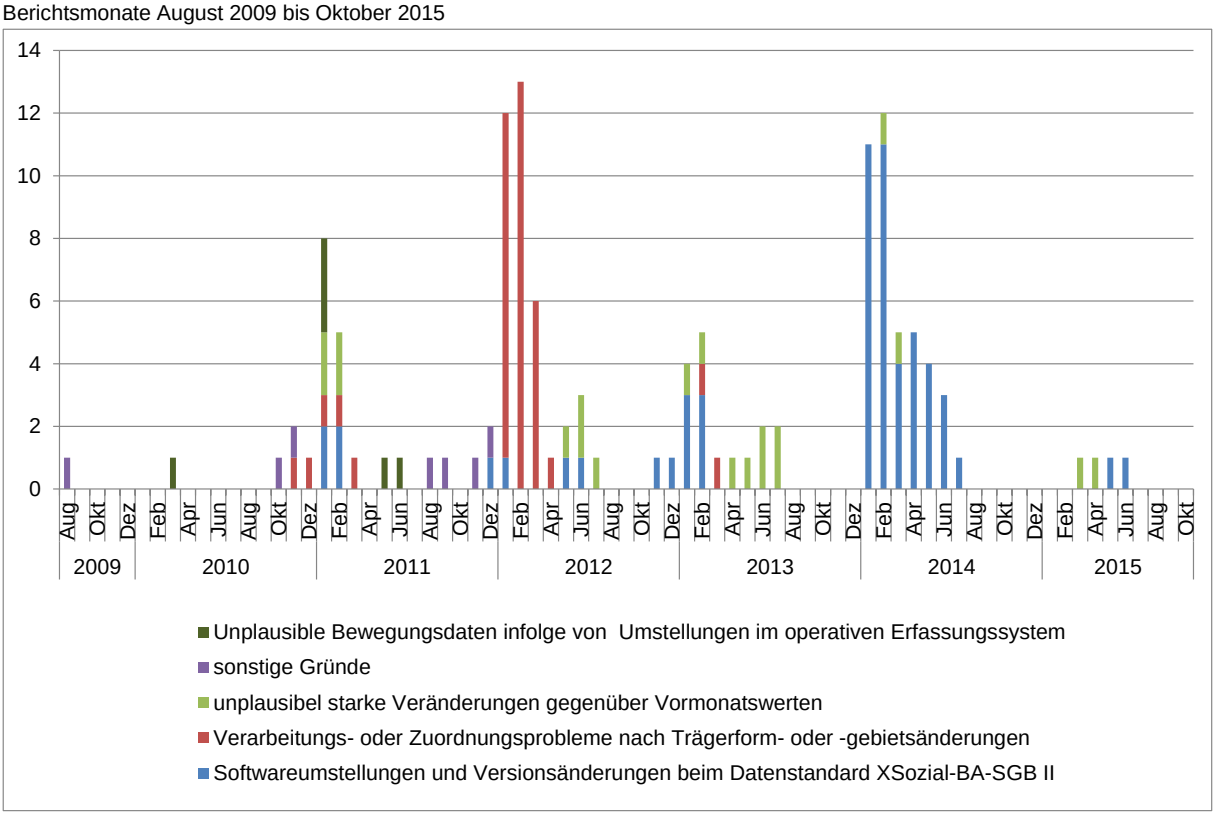
5 Datenimputation zum Themenbereich „Bewegungsdaten“

5.1 Zu- und Abgänge von erwerbsfähigen Leistungsberechtigten

5.1.1 Analyse fehlender Werte

Abbildung 12: ELB-Bewegungen – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei ELB-Bewegungen
Alle JC
Berichtsmonate August 2009 bis Oktober 2015



Ausgehend von der statistischen Bestandszählung von ELB werden bei den Kennzahlen nach § 48a SGB II alle Wechsel in die bzw. aus der Personengruppe der ELB als Bewegungen gezählt – auch Statuswechsel von nicht erwerbsfähig zu erwerbsfähig und umgekehrt.[61]

- Bei einem Zugang wechselt die betreffende Person in die Personengruppe der ELB. Der Status der Person ändert sich von „kein ELB“ auf „ELB“.
- Bei einem Abgang verlässt die betreffende Person die Personengruppe der ELB. Der Status der Person ändert sich von „ELB“ auf „kein „ELB“.

[61] Bis zur Revision der Grundsicherungsstatistik SGB II im April 2016 wurden solche Statuswechsel nicht als Bewegungen gezählt, vgl. vorangehende Versionen dieser Methodenübersicht vor Version 5.0 (Stand Mai 2016).

Kurzzeitige Unterbrechungen des Leistungsbezugs als ELB von bis zu sieben Tagen führen zu keinem Ab- oder Zugang. Hier wird davon ausgegangen, dass es sich um prozessgesteuerte Bewegungen handelt, zum Beispiel aufgrund verspäteter Antragstellung bei der Weiterbewilligung. Somit werden nur Unterbrechungen von mehr als sieben Tagen als Bewegungen gewertet.

Zu- und Abgänge von ELB sind wegen des sogenannten „Stock-Flow-Zusammenhangs“ immer gleichzeitig von Datenlücken betroffen (Erläuterung im nachfolgenden Abschnitt 5.1.2). Im Zeitraum von August 2009 bis Oktober 2015 liegen insgesamt 128 Datenlücken vor. Bezogen auf 75 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,42 %.

Die Datenlücken häufen sich in den vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen eine untergeordnete Rolle. Die Lücken verteilen sich auf 47 JC (etwa 12 % der JC). Bei 43 JC fehlen mehrfach Bewegungsdaten für ELB, wobei die maximale Häufigkeit bei sechs hintereinander gereihten Datenlücken für ein JC liegt.

5.1.2 Schätzmodell

Keines der hier bei anderen Grunddaten verwendeten Modelle ergab für die Datenimputation bei den Bewegungsarten der Zu- und Abgänge befriedigende Ergebnisse. Auch klassische zeitreihenanalytische Modelle schnitten schlecht ab. Zudem verletzten getestete Verfahren, die die beiden Bewegungsarten unabhängig voneinander schätzten, regelmäßig die sogenannte „Stock-Flow-Bedingung“. Diese beschreibt allgemein den Zusammenhang zwischen Stromgrößen und Beständen: der Bestand eines Berichtsmonats ergibt sich rechnerisch aus dem Bestand des Vormonats minus der Abgänge und plus der Zugänge zwischen den Stichtagen. Diese Bedingung erwies sich bei der Bewertung der getesteten Verfahren als zentrales Gütekriterium, sodass das letztlich vorgeschlagene Schätzmodell für beide Größen eigens hierauf zugeschnitten wurde. Somit unterscheidet sich das Modell zur Schätzung für die Bewegungsarten Zu- und Abgänge von allen anderen verwendeten Datenimputationsmodellen in einem Punkt: Die Schätzung erfolgt für beide Größen immer simultan, um die wechselseitige Stock-Flow-Abhängigkeit zu berücksichtigen.

Im vorliegenden Fall wird das Stock-Flow-Modell um den Korrekturterm ρ (Rho) erweitert, da vermeintliche Unterbrechungen des Leistungsbezugs von mehr als sieben Tagen (vgl. vorangehender Abschnitt 5.1.1) das Modell ansonsten verzerren würden: der Korrekturterm ρ fängt Bewegungen auf, die aus technischen oder operativen Gründen gemessen werden, obwohl ein ELB seinen Leistungsbezug eigentlich nicht unterbrochen hat. Somit werden zwar Bewegungen gemessen, aber der ELB-Bestand bleibt unverändert. Ohne die Größe ρ ginge das Stock-Flow-Modell rechnerisch nicht auf. In einer einfachen Formel lautet die Basisgleichung, die der Schätzung fehlender ELB-Zu- und Abgänge zugrunde liegt:

$$Bestand_t = Bestand_{t-1} - Abgang_t + Zugang_t + \rho_t \quad (10)$$

Somit wird, sobald für eine der beiden Bewegungsarten ein Imputationswert ermittelt wurde, der Wert für die andere Bewegungsart simultan mitgeliefert. Über welche Größe man die Schätzgleichung definiert –

über Zugänge oder Abgänge – hat keinen Einfluss auf die Schätzung. Hier werden Zugänge immer in Abhängigkeit von den Abgängen definiert.

Die erweiterte Stock-Flow-Gleichung und einige andere mathematisch formalisierte Bedingungen und Zusammenhänge zu Zugangs- und Abgangsraten finden in eine eigens auf dieses Schätzproblem hin modellierte Optimierungsfunktion Eingang.[62] Die Funktion berücksichtigt mit

- den Bewegungsraten des betreffenden JC in den letzten zwölf Monaten einerseits JC-individuelle Besonderheiten im Niveau und
- der Entwicklung der Bewegungsraten am aktuellen Rand im zugehörigen SGB-II-Vergleichstyp andererseits auch den allgemeinen aktuellen Trend.

Mit Hilfe dieses Optimierungsmodells erfolgt die Schätzung der Abgänge, um aus der Stock-Flow-Gleichung simultan die Zugänge zu erhalten. Um unplausibel niedrige oder sogar negative Schätzwerte zu vermeiden, werden als Mindestfallzahlen statistisch abgeleitete untere Schranken eingezogen. Diese leiten sich wiederum ab aus

- den Bewegungsraten des betreffenden JC in den letzten zwölf Monaten für JC-individuelle Besonderheiten in Niveau, Trend und Saison und
- der Entwicklung der Bewegungsraten am aktuellen Rand über alle JC[63] für den allgemeinen aktuellen Trend.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

5.1.3 Bewertung[64]

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von August 2009 bis August 2013.[65] Für die quantitative Bewertung wird das Optimierungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Im Ergebnis ist das vorgeschlagene Modell allen Vergleichsmodellen deutlich überlegen, sowohl im Mittel als auch bei den Ausreißern. Gleichwohl existieren Ausreißer, die zu sehr hohen relativen Abweichungen führen. Sie sind in erster Linie auf die häufig kleinen Fallzahlen und sehr großen Datensprünge bei Bewegungen zurückzuführen.

[62] Für eine ausführliche Erläuterung des gesamten Optimierungsmodells siehe Anhang 0.

[63] Die erneute Aufteilung nach SGB-II-Vergleichstyp hätte hier zu niedrige Zellbesetzungen zur Folge. Für die zuverlässige, stabile Schätzung werden hier alle JC verwendet.

[64] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[65] Da das Optimierungsmodell aus konzeptuellen Gründen einen 12-monatigen Vorlauf benötigt und die verglichenen Verfahren zusätzlich einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen Werte ab September 2010 bis Juli 2013.

Abbildung 13: Abgänge von ELB – Methodenvergleich Fehlerverteilungen

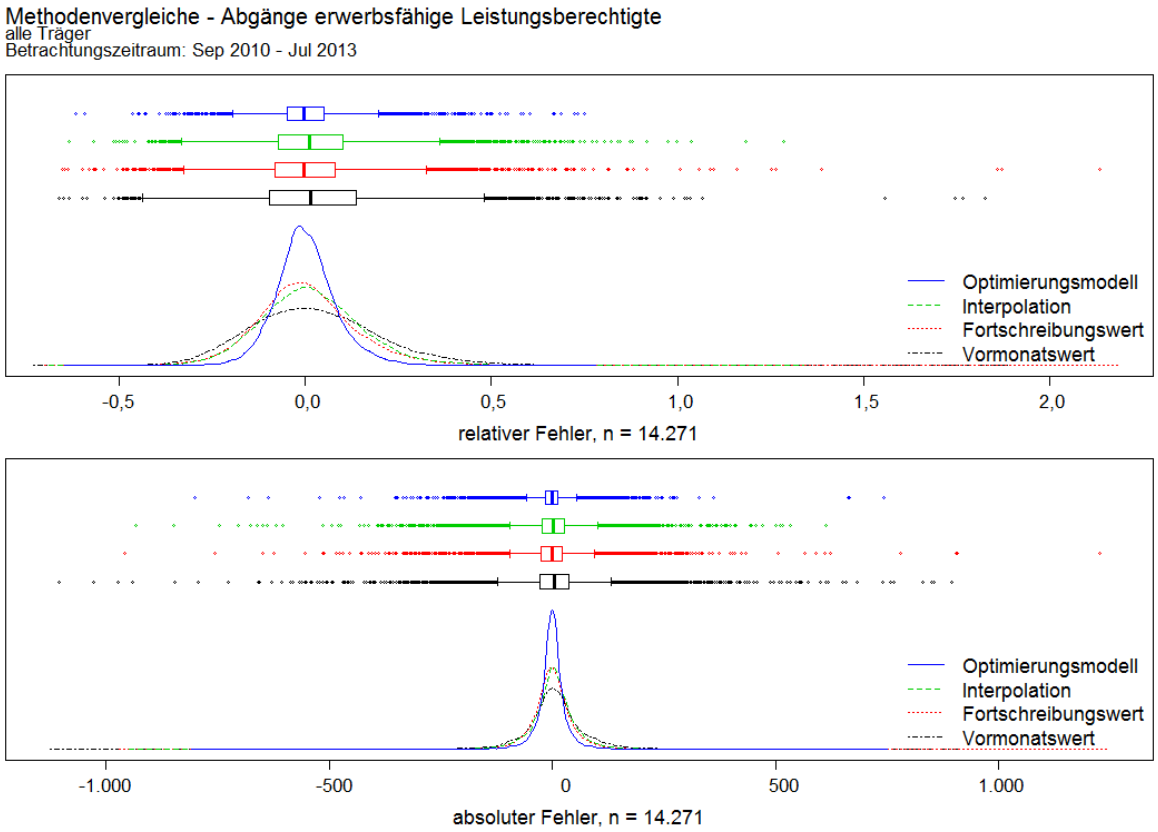
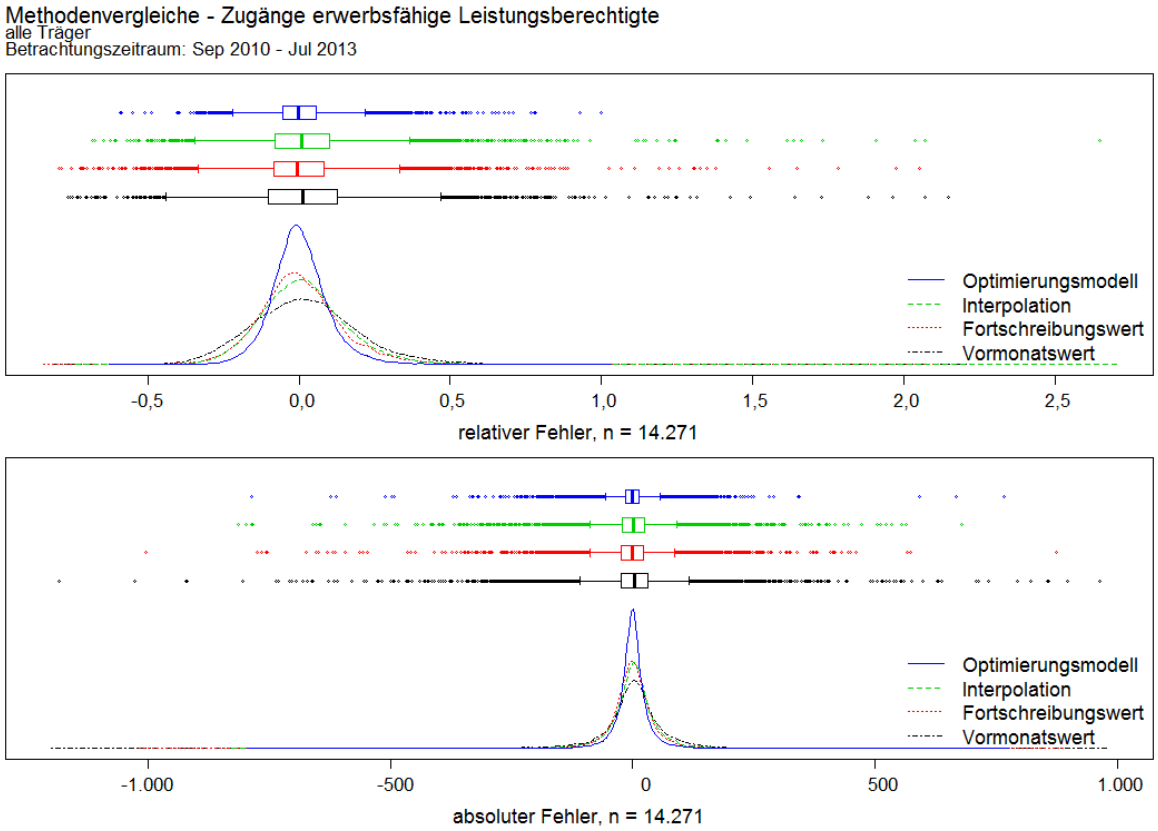


Abbildung 14: Zugänge von ELB – Methodenvergleich Fehlerverteilungen



Die Überlegenheit insbesondere gegenüber dem Fortschreibungsmodell rührt sicher daher, dass es dem Optimierungsverfahren gelingt, neben der Entwicklung am aktuellen Rand auch JC-spezifische längerfristige Trend- und Saisoneffekte einzufangen. Hierdurch wird insbesondere die Streuung beim Schätzfehler verringert.

Bei der Anwendung des Optimierungsmodells wird die tatsächliche Anzahl an Zugängen im Median und Betrag um 14 (5,5 %) und diejenige an Abgängen ebenfalls um 14 (4,9 %) verfehlt, wobei das Bild von JC zu JC variiert. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren für die Zahl der Zugänge bei ca. +/-22 %, für jene der Abgänge bei ca. +/-20 %. Dies spricht für einen mäßigen Schätzfehler im Allgemeinen. Betrachtet man Daten, die sowohl bei absoluten als auch prozentualen Fehlern als Ausreißer klassifiziert werden, sind bei dem vorgeschlagenen Imputationsverfahren lediglich 1,4 % (ELB-Zugänge) bzw. 1,3 % (ELB-Abgänge) der geschätzten Datenpunkte auffällig. In den Monaten, in denen zusätzlich auch der Bestand geschätzt wird, erhöht sich die Wahrscheinlichkeit eines größeren Schätzfehlers für die Zu- bzw. Abgänge.

Die im Modell beschriebenen unteren Schranken für die Fallzahlen mussten während der Simulation selten eingesetzt werden. Die untere Grenze bei den Zugängen wurde 33-mal verwendet und der Minimalwert für die Abgänge 54-mal. Die Verwendung dieses Sicherheitsmechanismus entspricht damit weniger als 1 % aller simulierten Datenausfälle, so dass im Normalfall der zu imputierende Wert aus dem Optimierungsmodell stammt. Die wenigen Sonderfälle sind auf größere Bestandsveränderungen zurückzuführen.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei den Bewegungen der ELB meist von guter Qualität ist. Denn die geschlechtsspezifischen Anteilswerte variieren von September 2010 bis August 2013 nur mäßig. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt bei beiden Bewegungsgrößen in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 17 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats nur zwischen vier und neun Prozentpunkte. Bei über der Hälfte der vorliegenden Datenausfälle (42 von 81) werden WZ2- oder WZ1-Daten mit einer noch guten Variabilitätsspanne unter 10 % herangezogen. Bei den verbleibenden 39 Datenpunkten wird auf die Anteile aus den WZ0-Daten oder den Vormonaten ausgewichen. Deren Anteilsinformation hat mit einer bis zu 20 %-igen Variabilitätsspanne eine befriedigende Güte.

5.1.4 Fazit, Ausblick

Bei der Schätzung der ELB-Bewegungen produziert das vorgeschlagene Datenimputationsverfahren im Median einen betragsmäßigen Schätzfehler von 14 Zugängen (5,5 %) und ebenfalls 14 Abgängen (4,9 %). Das Modell weist eine noch gute Schätzgüte auf. Insgesamt produziert das Modell einige Ausreißer, die angesichts der hohen Variabilität bei beiden Bewegungsgrößen nicht überraschen. Das Modell bietet mit den vorliegenden Mitteln und unter der Berücksichtigung der Stock-Flow-Annahme das bestmögliche Verfahren für die Schätzung fehlender Werte.

5.2 Zu- und Abgänge von Langzeitleistungsbeziehenden

5.2.1 Analyse fehlender Werte

Zugänge von Langzeitleistungsbeziehenden (LZB, zur Definition vgl. Abschnitt 4.3.1) sind alle Leistungsberechtigten, die

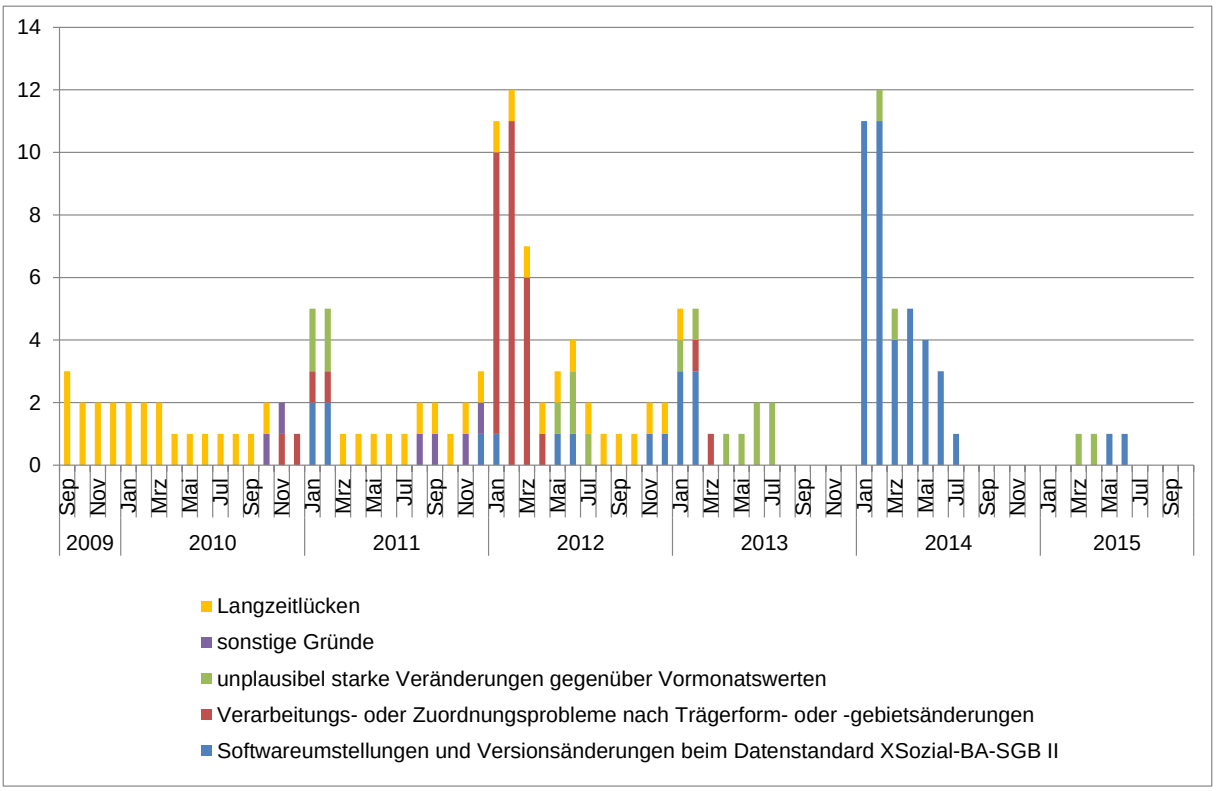
- zum statistischen Stichtag des aktuellen Monats ELB sind,
- in den vergangenen 24 Monaten mindestens 21 Monate ELB waren,
- und zum Stichtag des Vormonats keine LZB waren.[66]

Abgänge von LZB sind alle Leistungsberechtigten, die

- im Bestand des Vormonats LZB waren
- und sich zum Stichtag des aktuellen Monats nicht mehr im Langzeitleistungsbezug befinden.[67]

Abbildung 15: LZB-Bewegungen – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei LZB-Bewegungen
Alle JC
Berichtsmonate September 2009 bis Oktober 2015



[66] Siehe dazu die Ausführungen in den Detailbeschreibungen unter https://www.sgb2.info/ > Kennzahlen > Hilfe und Erläuterungen > Materialien & Downloads > Detailbeschreibungen.

[67] Siehe dazu die Ausführungen in den Detailbeschreibungen unter https://www.sgb2.info/ > Kennzahlen > Hilfe und Erläuterungen > Materialien & Downloads > Detailbeschreibungen.

Wegen des Stock-Flow-Zusammenhangs sind die Zu- und Abgänge von LZB genau wie jene der ELB immer gleichzeitig von Datenlücken betroffen (Begründung anhand der ELB-Bewegungen in Abschnitt 5.1.2). Im Zeitraum von September 2009 bis Oktober 2015 kommt es bei LZB-Bewegungen zu insgesamt 161 Datenlücken. Bezogen auf 74 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,53 %. Dabei sind vier JC (knapp 1 %) von 45 Langzeitlücken, 42 JC (10 %) sind von 116 isolierten Ausfällen betroffen (zur Unterscheidung vgl. Abschnitt 4.3.1). Diese häufen sich vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen für isolierte Datenlücken eine untergeordnete Rolle. Selbst im Fall von isolierten Datenlücken fehlen bei allen betroffenen 42 JC immer mehrfach Bewegungsdaten für LZB, wobei die maximale Häufigkeit bei sechs hintereinander gereihten Datenlücken für ein JC liegt. Allerdings führen diese Mehrfachausfälle nie zu den oben genannten Langzeitlücken, da in diesen Konstellationen die Nettodauern für den Leistungsbezug als ELB noch ermittelt werden können.

Zumindest das Ausmaß isolierter Datenlücken bei den Bewegungen von LZB scheint insgesamt gering. Gleichwohl setzen sich fehlende LZB-Werte auch bei der Differenzierung nach Geschlecht fort.

5.2.2 Schätzmodell

Für Datenlücken bei LZB-Bewegungen soll wie bei ELB-Bewegungen das Optimierungsmodell verwendet werden (vgl. Abschnitt 5.1.2)[68]. Anders als bei den LZB-Beständen wird bei den LZB-Bewegungen somit ein einheitliches Modell über beide Arten von Datenlücken – isolierte und Langzeitlücken – vorgeschlagen. Das Rekonstruktionsverfahren (vgl. Abschnitt 4.3.2.2) wurde für Langzeitlücken bei LZB-Bewegungen zwar im Ansatz mit guten Ergebnissen getestet. Der Test zeigte aber auch, dass dieses Verfahren nicht nur viel aufwändiger wäre als bei der Bestandsschätzung, sondern auch die fristgerechte Datenbereitstellung in der monatlichen Stichtagsverarbeitung gefährden könnte.

Konzeptuell basiert das Optimierungsmodell auch bei LZB-Bewegungen auf der erweiterten Stock-Flow-Gleichung (10) aus Abschnitt 5.1.2. Dabei steht der erweiternde Korrekturterm ρ (Rho) im Bereich der LZB für ein anderes Phänomen wie bei der Schätzung von ELB-Bewegungen. Hier handelt es sich um Personen, die in ein anderes JC-Gebiet gezogen sind und nicht in die Bewegungszählung eingehen, da sie vorher wie nachher LZB sind. Dies hat zwar geringfügige Verschiebungen in den LZB-Beständen der JC zur Folge, aber es werden keine Bewegungen gemessen. Ohne die Größe ρ ginge auch hier das Stock-Flow-Modell rechnerisch nicht vollständig auf. Vom unterschiedlichen Bedeutungsgehalt des Korrekturterms ρ abgesehen, wird hier wie dort dasselbe Datenimputationsverfahren angewendet.

Drei JC sind direkt zu Beginn der Zeitreihe (ab September 2009) von Langzeitlücken betroffenen – im ersten Fall mit nur einem einzigen Schätzpunkt, weshalb ein Imputationswert hier für entbehrlich gehalten wird, im zweiten Fall mit sieben und im dritten Fall mit 14 Schätzpunkten. Für die insgesamt 21 Imputationen bei den beiden letztgenannten JC stehen keine zwölf vorangehenden Berichtsmonatswerte für das Optimierungsmodell zur Verfügung. Um dennoch Imputationswerte zu erhalten, wird das Schätzmodell

[68] Für eine ausführliche Erläuterung des gesamten Optimierungsmodells siehe Anhang 0.

angepasst: das Schätzverfahren wird als Rückwärtsschätzung angewendet. Das heißt, das initiale Zwölfmonatszeitfenster wird zeitlich nach dem Langzeitausfall des jeweiligen JC verlegt und läuft für frühere Schätzpunkte rollierend nach links mit.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens. Im Falle der Rückwärtsschätzung für Langzeitlücken zu Beginn der Zeitreihe wird der gemittelte Frauen- bzw. Männeranteilswert für das jeweilige JC aus dem initialen Zwölfmonatszeitfenster nach dem Langzeitausfall verwendet – eine pragmatische Lösung analog zur Rückwärtsschätzung.

5.2.3 Bewertung[69]

Für Datenlücken bei LZB-Bewegungen wird zwar über beide Arten von Datenlücken – isolierte und Langzeitlücken – einheitlich das Optimierungsmodell verwendet. In der Bewertung des Schätzverfahrens ist es dennoch zweckmäßig, nach beiden Lückenvarianten zu unterscheiden.

5.2.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von September 2009 bis August 2013.[70] Für die quantitative Bewertung wird das Optimierungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Im Ergebnis ist das vorgeschlagene Modell allen Vergleichsmodellen deutlich überlegen, sowohl im Mittel als auch bei den Ausreißern. Gleichwohl existieren Ausreißer, die zu sehr hohen relativen Abweichungen führen. Ausreißer beim relativen Schätzfehler sind in erster Linie auf teils sehr kleine Fallzahlen und große Datensprünge bei LZB-Bewegungen zurückzuführen.

Wiederum rührt die Überlegenheit insbesondere gegenüber dem Fortschreibungsmodell sicher daher, dass es dem Optimierungsverfahren gelingt, neben der Entwicklung am aktuellen Rand auch JC-spezifische längerfristige Trend- und Saisoneffekte einzufangen. Hierdurch wird insbesondere die Streuung beim Schätzfehler verringert.

[69] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[70] Da das Optimierungsmodell aus konzeptuellen Gründen einen 12-monatigen Vorlauf benötigt und die verglichenen Verfahren zusätzlich einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen Werte ab September 2010 bis Juli 2013.

Abbildung 16: Abgänge von LZB – Methodenvergleich Fehlerverteilungen

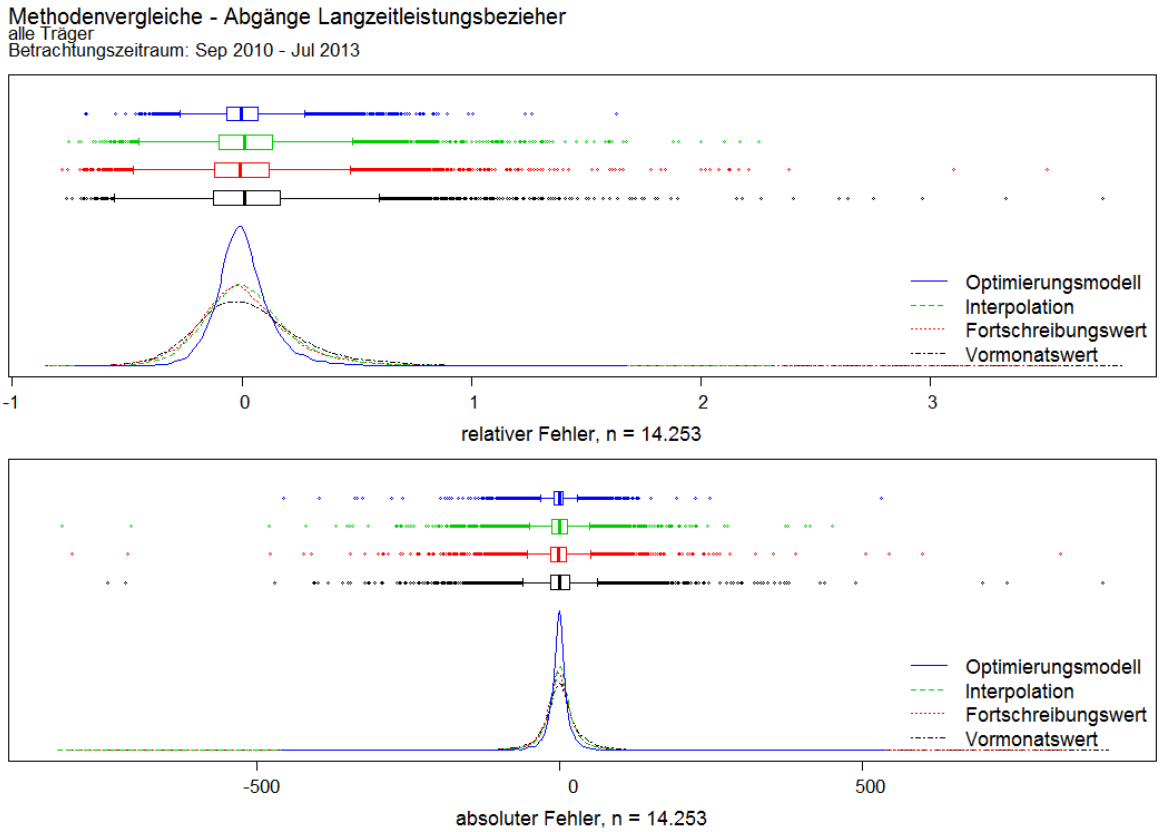
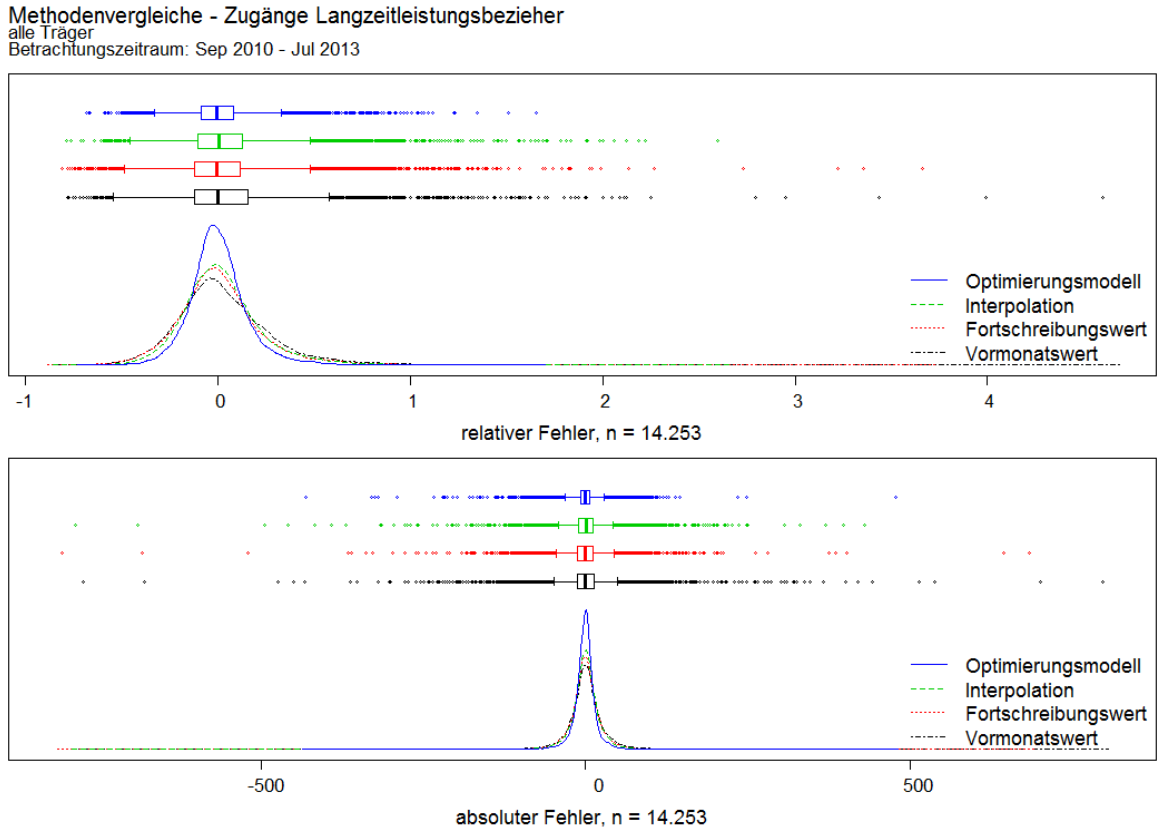


Abbildung 17: Zugänge von LZB – Methodenvergleich Fehlerverteilungen



Bei der Anwendung des Optimierungsmodells wird die tatsächliche Anzahl an Zugängen im Median und Betrag um acht (8,3 %) und diejenige an Abgängen um sieben (6,8 %) verfehlt, wobei das Bild von JC zu JC variiert. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren für die Zahl der Zugänge bei ca. +/- 33 %, für jene der Abgänge bei ca. +/- 27 %. Dies spricht für einen mäßigen Schätzfehler im Allgemeinen. Betrachtet man Daten, die sowohl bei absoluten als auch prozentualen Fehlern als Ausreißer klassifiziert werden, sind bei dem vorgeschlagenen Imputationsverfahren lediglich 1,0 % (LZB-Zugänge) bzw. 1,1 % (LZB-Abgänge) der geschätzten Datenpunkte auffällig. In den Monaten, in denen zusätzlich auch der Bestand geschätzt wird, erhöht sich die Wahrscheinlichkeit eines größeren Schätzfehlers für die Zu- bzw. Abgänge.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei isolierten Lücken der LZB-Bewegungen noch von zufriedenstellender Qualität ist. Die geschlechtsspezifischen Anteilswerte variieren von September 2009 bis August 2013 deutlich. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 bei beiden Bewegungsarten rund 30 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats zwischen neun und 19 Prozentpunkten. Immerhin werden bei etwa einem Drittel der vorliegenden isolierten Datenausfälle (23 von 81) WZ2-Daten mit einer noch guten Variabilitätsspanne unter 10 % herangezogen. Bei den verbleibenden 46 Datenpunkten wird auf die Anteile aus den WZ1- und WZ0-Daten oder den Vormonaten ausgewichen. Deren Anteilsinformation hat mit einer bis zu 30 %igen Variabilitätsspanne eine relativ geringe Güte.

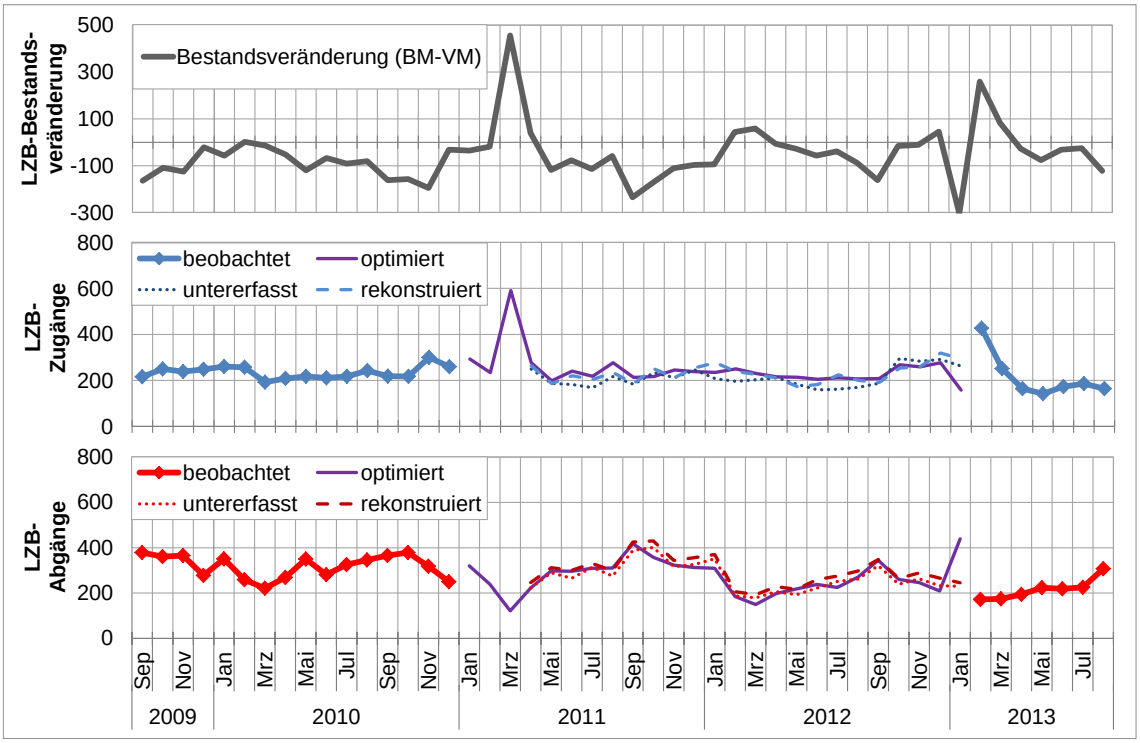
5.2.3.2 Bewertung des Schätzmodells bei Langzeitlücken

Bei Langzeitlücken ist es nicht möglich, für das verwendete Optimierungsmodell den Standardvergleich nach Abschnitt 2.1.3 oder eine andere auf den Ausfall bezogene Simulation vorzunehmen. Dennoch stehen teilweise Datenimputationswerte zur Verfügung, die testweise für eines der betroffenen JC mit dem Rekonstruktionsverfahren für Langzeitlücken geschätzt worden sind. Das Rekonstruktionsverfahren hat sich in Anwendung für LZB-Bestände als sehr gut erwiesen (vgl. Abschnitt 4.3.3.2). Es ist davon auszugehen, dass die testweise rekonstruierten Schätzwerte für LZB-Bewegungen ebenfalls sehr nah an den Realwerten liegen. Auch wenn das Rekonstruktionsverfahren für die Schätzung von Langzeitlücken bei den LZB-Bewegungen aus wirtschaftlichen und zeitkritischen Erwägungen abgelehnt wurde, können dessen Schätzwerte bei Langzeitlücken als Prüfstein für das Optimierungsmodell herangezogen werden.

In Abbildung 18 ist im oberen Abschnitt die Bestandsveränderung des betreffenden JC abgebildet. Im Januar und Februar 2011 (isolierte Datenlücken) liegen der Berechnung die fortgeschriebenen Imputationswerte zugrunde, danach (Langzeitlücken) die Rekonstruktionswerte. Man sieht insbesondere in den Monaten März 2011, Januar und Februar 2013 die Übergänge zwischen Realwerten und Schätzverfahren und umgekehrt.

Im mittleren Abschnitt sind Zugänge abgebildet: Die dicke blaue durchgezogene Linie mit den markierten Datenpunkten zeigt die realen, plausiblen Werte, die blau-gepunktete Linie die untererfassten, unplausiblen Werte. Die hellblau gestrichelte Linie bildet die Imputationswerte aus dem Rekonstruktionsverfahren ab, die dünne lila durchgezogene Linie die Imputationswerte aus dem Optimierungsverfahren.

Abbildung 18: Zu- und Abgänge LZB – Methodenvergleich bei Langzeitlücken: Test-Rekonstruktion versus Optimierungsmodell



Im unteren Abschnitt sind die Abgänge abgebildet: Die dicke rote durchgezogene Linie mit den markierten Datenpunkten zeigt die realen, plausiblen Werte, die rot-gepunktete Linie die untererfassten, unplausiblen Abgänge. Die dunkelrot gestrichelte Linie bildet die Imputationswerte aus dem Rekonstruktionsverfahren ab, die dünne lila durchgezogene Linie die Imputationswerte aus dem Optimierungsverfahren.

In den unteren beiden Abschnitten ist erkennbar, dass sich die jeweiligen Imputationswerte aus dem Optimierungsverfahren in den Übergangsmonaten nicht gut in die übrigen Zeitreihen einfügen. Dem ersten Anschein nach produziert das Optimierungsverfahren in den Übergangsmonaten Ausreißer mit mangelnder Schätzgüte. Die Beobachtung relativiert sich jedoch, wenn man zusätzlich die Zeitreihe der Bestandsveränderung im oberen Abschnitt zu Rate zieht: Die extrem hohen bzw. niedrigen Optimierungswerte in den Übergangsmonaten passen rechnerisch zu den Bestandsveränderungen. Da auch diese Werte in den Übergangsmonaten gewissen Schätzunsicherheiten unterliegen, ist nicht global von einer mangelnden Schätzgüte des Optimierungsmodells an sich auszugehen. Zudem sind die Schätzergebnisse über alle drei Grunddaten hinweg stimmig und die Stock-Flow-Gleichung als zentrales Gütekriterium bzw. als zentrale Modellbedingung ist erfüllt.

Insgesamt produziert das Optimierungsverfahren im Imputationszeitraum der Langzeitlücken sehr ähnliche Ergebnisse wie das Rekonstruktionsverfahren. Offenbar gelingt es dem Optimierungsverfahren, Saison- und Trendmuster gut abzubilden. Nur in vereinzelten Monaten liegen die Optimierungswerte sogar

unterhalb der untererfassten Bewegungsdaten, die eigentlich die tatsächlich messbare Untergrenze für Schätzwerte angeben dürften. Diese Schätzunsicherheit liegt aber im Rahmen der Abweichungen, die bei Schätzungen zwangsläufig auftreten können, insbesondere angesichts des langen Schätzzeitraums.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht bei Langzeitlücken mit deutlichen Schwächen behaftet ist – sowohl bei der Vorwärts- als auch bei der Rückwärtsschätzung. Bei beiden Varianten wird über lange Schätzzeiträume (zwischen sieben und 23 Monate) – quasi statisch – die Anteilswertinformation aus dem Zwölfmonatsfenster vor bzw. nach dem Langzeitausfall verwendet. Mangels Grunddaten mit verkürzter Wartezeit stellt dies eine pragmatische Lösung analog zum zugrundeliegenden Schätzmodell dar. Doch würden die wahren, aber unbekannten, Frauen- bzw. Männeranteile sicherlich variieren. Die bei LZB-Bewegungen übliche merkliche Variabilitätsspanne zwischen den Berichtsmonaten von rund 30 Prozentpunkten bildet die Geschlechterschätzung bei Langzeitausfällen nicht ab.

5.2.4 Fazit, Ausblick

Sowohl bei isolierten als auch bei Langzeitlücken wird bei LZB-Bewegungen das Optimierungsmodell verwendet. Das Modell bietet mit den vorliegenden Mitteln und unter der Berücksichtigung der Stock-Flow-Annahme das bestmögliche Verfahren für die Schätzung fehlender Werte.

Bei isolierten Datenlücken produziert das vorgeschlagene Datenimputationsverfahren im Median einen betragsmäßigen Schätzfehler von acht Zugängen (8,3 %) und sieben Abgängen (6,8 %). Hier ist insgesamt von einer noch guten Schätzgüte auszugehen.

Bei Langzeitlücken gelingt es dem Optimierungsverfahren offenbar, Saison- und Trendmuster gut abzubilden. Nur in vereinzelten Monaten dürften Fehlschätzungen vorliegen, die aber akzeptabel scheinen. Gegenüber der erneuten Anwendung des Rekonstruktionsverfahrens bietet das Optimierungsmodell den Vorteil, dass es bei annähernd gleichen Ergebnissen viel weniger Aufwand und Fehleranfälligkeit mit sich bringt. Insbesondere angesichts des extrem langen Schätzzeitraums wird es als befriedigend bewertet.

6 Datenimputation zum Themenbereich „Beschäftigungsaufnahmen“

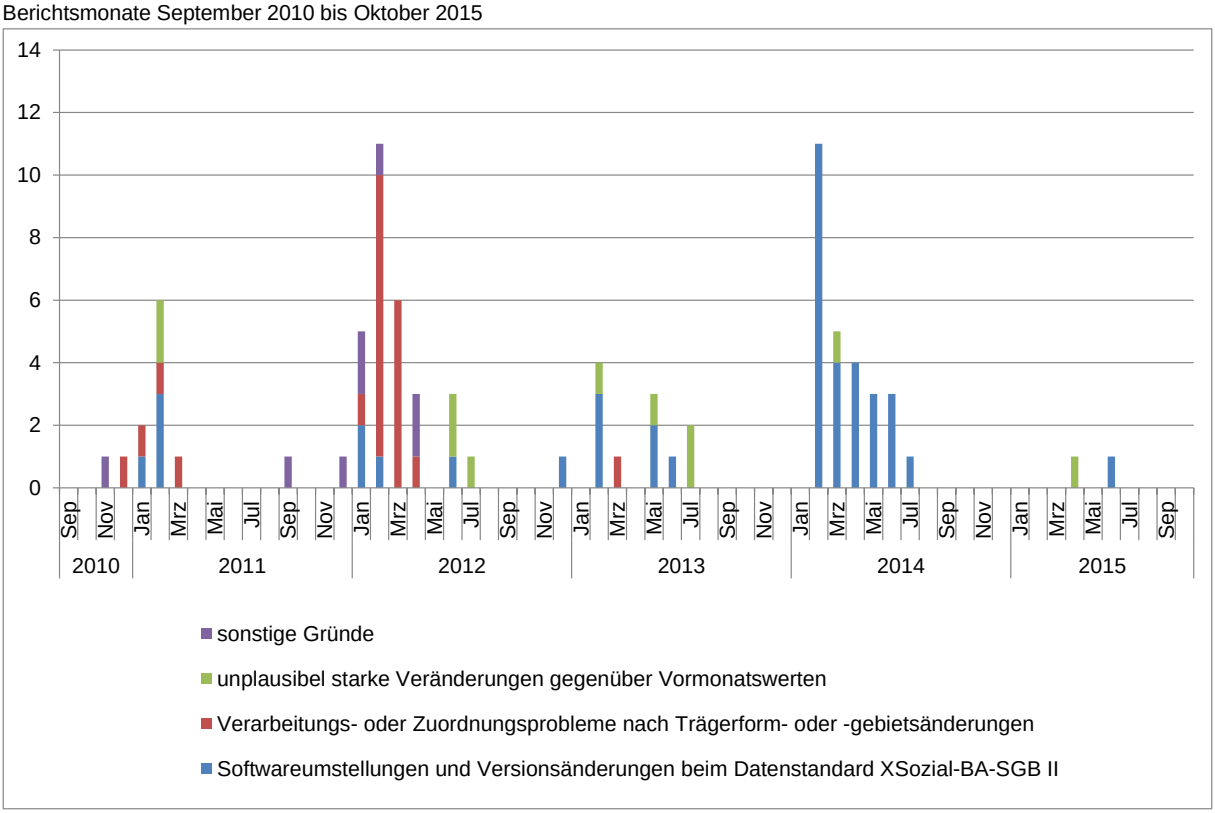
6.1 Integrationen von erwerbsfähigen Leistungsberechtigten

6.1.1 Analyse fehlender Werte

Bei Integrationen (definiert als ELB aus dem Vormonat, die eine Erwerbstätigkeit aufgenommen haben) kommt es bisher nur vereinzelt zu Datenlücken. Von September 2010[71] bis Oktober 2015 liegen 83 fehlende Werte vor. Bezogen auf 62 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,33 %.

Abbildung 19: Integrationen – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei Integrationen
Alle JC
Berichtsmonate September 2010 bis Oktober 2015



Die Datenlücken häufen sich vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen eine untergeordnete Rolle. Die Lücken verteilen sich auf 50 JC (etwa 12 % der JC). Bei 20 JC fehlen

[71] Die Zeitreihe für Integrationen beginnt später als für die meisten hier dargestellten Grunddaten. Die Voraussetzungen für eine einheitliche Ermittlung der Integrationen für die Quellverfahren VerBIS und XSozial waren erst ab Ende des Jahres 2010 vollständig gegeben. Um die Zeitreihe für die hier relevanten Testzwecke zu verlängern, wurden die geringfügig unterschiedlichen Ermittlungsweisen in den vier Monaten zuvor als tolerabel erachtet.

mehrfach Daten für Integrationen, wobei die maximale Häufigkeit bei fünf hintereinander gereihten Datenlücken für ein JC liegt.

In über 80 % der Fälle (in 68 Fällen) kommt es wegen unplausibler Vormonats-Eckwerte (das heißt wegen fehlerhafter Einzeldaten für die ELB) zu fehlenden Grundwerten für Integrationen, weil sich aus der zugrundeliegenden Gesamtheit keine gültige Integrationszahl bestimmen lässt. Nur in 15 Fällen gehen die Datenlücken direkt auf unplausible Informationen zu Beschäftigungsaufnahmen zurück.

Trotz des geringen Ausmaßes von Datenlücken steht fest: fehlende Integrationszahlen setzen sich auch bei der Differenzierung nach Geschlecht sowie bei Integrationen für Alleinerziehende und LZB fort und bringen Datenlücken bei der Teilgröße der sozialversicherungspflichtigen Integrationen und den daraus abgeleiteten kontinuierlichen Beschäftigungen nach Integration mit sich.

6.1.2 Schätzmodell

Die Gesamtgröße der Integrationen steht für ELB mit Beschäftigungsaufnahmen in abhängige und selbständige Erwerbstätigkeit sowie in Ausbildungsverhältnisse. Analysen zeigen, dass diese Größe insbesondere von saisonalen Effekten geprägt ist, die sich regionalspezifisch erheblich unterscheiden können. So schlagen beispielsweise witterungsbedingte Saisoneffekte oder der Beginn des Berufsausbildungsjahres auf dem bundesweiten Arbeitsmarkt durch. Auf regionalen Arbeitsmärkten ergeben sich aber durchaus unterschiedliche Effekte, die sich wiederum in wirtschaftsstrukturell ähnlichen Gebieten auch in ähnlichen Mustern niederschlagen.

Bei den Analysen erwies sich ein zweistufiges Schätzverfahren für die fehlenden Werte am besten, das ein Fortschreibungsmodell mit einer linearen Regression kombiniert.

In der ersten Stufe wird über ein spezielles Fortschreibungsverfahren, ein k-nächste-Nachbarn-Verfahren, die Entwicklung der Integrationszahl ähnlicher Nachbarn auf das betreffende JC übertragen. Der Schätzwert ergibt sich als Produkt des Vormonatswertes und der geschätzten mittleren Veränderung zum Vormonat der k ähnlichsten JC. Die Ähnlichkeit wird anhand der Entwicklung der letzten drei verfügbaren Monate und dem Vorjahresmonat – sofern vorhanden – ermittelt. Die optimale Anzahl k der ähnlichsten JC wird dabei monatsweise so bestimmt, dass der mittlere quadratische Fehler (auf Basis des relativen Fehlers) minimiert wird.[72] Die Zeitpunkte werden systematisch aufsteigend abgearbeitet. Dies bedingt –

[72] Für eine ausführliche Erläuterung des k-nächste Nachbarn-Verfahrens siehe Anhang 0.

vorteilhaft im Rahmen amtlicher Berichterstattung – die Festschreibung der Schätzwerte.[73] Diese erste Verfahrensstufe fängt latente regionalspezifische Entwicklungsmuster ein, die JC-übergreifend in der Zeitreihe zum Tragen kommen.

Die im ersten Schritt gewonnenen Fortschreibungswerte fließen in der zweiten Stufe in eine lineare Regression (Methode der kleinsten Quadrate) ein. Als zweiter Prädiktor fungiert die Größe der „Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II“.[74] Es fließt also eine weitere Saisonkomponente ein, die mit der Zahl der Integrationen in einem engen Zusammenhang steht. Insbesondere, um Ausreißermonate wie August und September mit ihren vielen Ausbildungsbeginner/innen aufzufangen, erfolgt die Regression monatsweise. Zudem wird getrennt nach den SGB-II-Vergleichstypen (vgl. Abschnitt 2.2) geschätzt. Dies gewährleistet, dass auch typische und relativ zeitstabile Aspekte wie zum Beispiel die Aufnahmefähigkeit des jeweiligen lokalen Arbeitsmarktes berücksichtigt werden. Die Regressionsparameter werden also je Monat und Vergleichstyp mit zwei erklärenden Variablen (Simulationsergebnis aus dem k-nächste-Nachbarn-Verfahren und Abgänge aus Arbeitslosigkeit im SGB II) ermittelt.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

6.1.3 Bewertung[75]

Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von September 2010 bis August 2013.[76] Für die quantitative Bewertung wird das Fortschreibungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Im Ergebnis ist das kombinierte zweistufige Verfahren

[73] Alternative Vorgehensweisen, die dies nicht gewährleisten (Berücksichtigung auch von Nachfolgemonaten, globale statt monatsweiser Optimierung), sind im Simulationstest weder wesentlich noch durchgängig überlegen und werden deshalb verworfen.

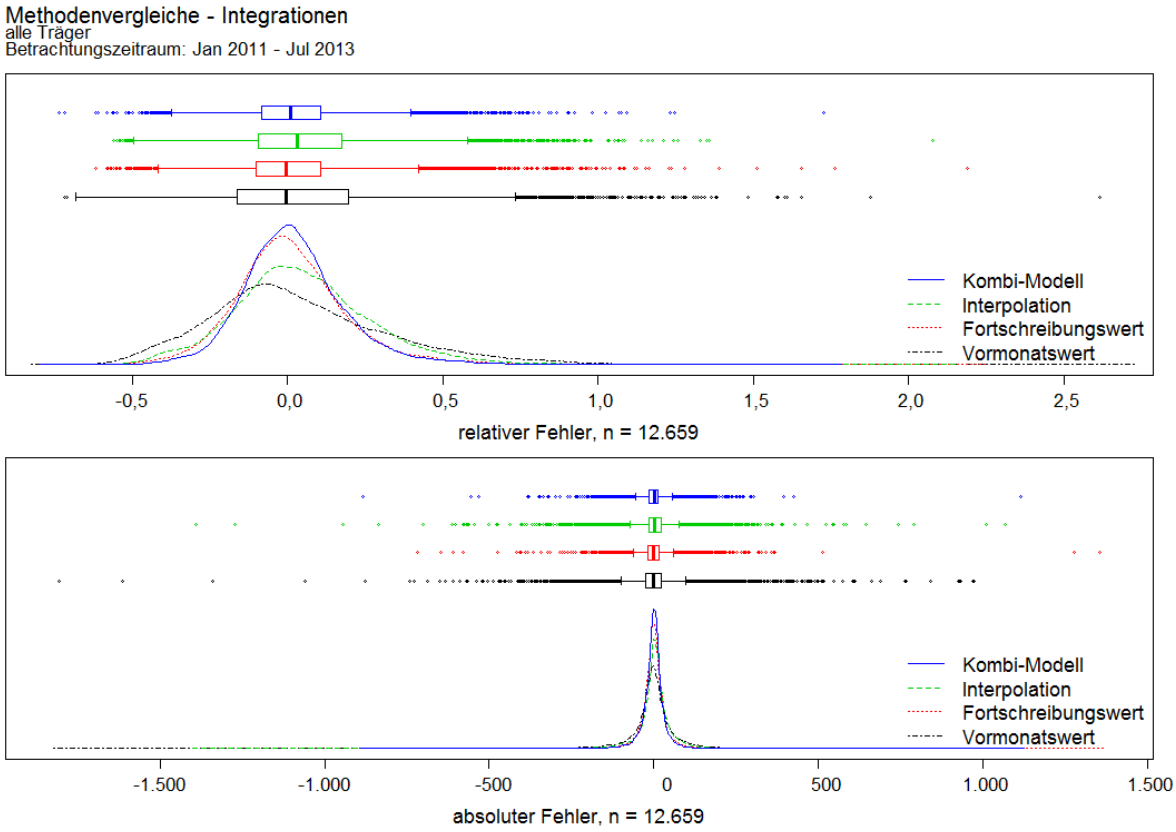
[74] Zu den Arbeitslosen im SGB II gehörten bis Dezember 2016 auch Personen, die gleichzeitig Arbeitslosengeld II und Arbeitslosengeld nach dem SGB III erhielten. Ab dem 1. Januar 2017 zählen diese sog. „Aufstocker/innen“ nun als Arbeitslose im Rechtskreis SGB III und sind nicht mehr in der Prädiktorvariable „Abgänge aus Arbeitslosigkeit im SGB II“ enthalten. In die Zählung von Integrationen nach § 48a SGB II gehen die Beschäftigungsaufnahmen von Aufstocker/innen weiterhin vollumfänglich ein, weshalb sich der Zusammenhang zwischen den beiden Variablen „Abgänge aus Arbeitslosigkeit SGB II“ und „Integrationen“ verändert. Analysen haben jedoch ergeben, dass sich der Wegfall der Aufstocker/innen aus der Prädiktorvariable nicht nennenswert auf die Schätzergebnisse für Integrationen auswirkt.

[75] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[76] Da das knN-Verfahren aus konzeptuellen Gründen einen 4-monatigen Vorlauf benötigt und die verglichenen Verfahren einen 1-monatigen Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Januar 2011 bis Juli 2013.

(Kombi-Modell) den anderen Modellen überlegen: die Verteilung der absoluten und relativen Fehler weist hier die stärkste Konzentration bei null auf. Die Überlegenheit gegenüber den Standardvergleichsmodellen rührt sicher daher, dass das kombinierte zweistufige Modell sowohl kenngrößen- als auch arbeitsmarktspezifische Eigenheiten besonders gut erfassen kann. Die enge Verwandtschaft bzw. hohe Korrelation der beiden Regressoren befördert zwar Multikollinearität.[77] Etwaige Effekte wie resultierende Schätzungenauigkeiten für die Koeffizienten können jedoch in Kauf genommen werden. Denn im vorliegenden Anwendungsfall interessiert weniger, wie gut einzelne Koeffizienten geschätzt werden. Das Verfahren wird hier als prädiktives Modell eingesetzt, um im Einzelfall Schätzwerte möglichst nah an den Ursprungswerten zu erhalten. Folglich werden die gewählten Regressoren aufgrund ihrer hohen inhaltlichen Relevanz im Modell belassen.

Abbildung 20: Integrationen – Methodenvergleich Fehlerverteilungen



Bei der Anwendung des kombinierten zweistufigen Modells wird die tatsächliche Anzahl an Integrationen im Median und Betrag um 14 Integrationen (9,6 %) verfehlt, wobei das Bild von JC zu JC variiert. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren

[77] Eine gute Erläuterung des Problems der Multikollinearität im Zusammenhang mit multivariaten Regressionsanalysen findet sich zum Beispiel in Backhaus, K; Erichson, B; Plinke, W.; Weiber, R. (2018): Multivariate Analysemethoden, 15. Auflage, Springer, Seite 98ff.

bei ca. +/- 40 %. Dies spricht für einen merklichen Schätzfehler im Allgemeinen. Ausreißer der absoluten Fehler treten häufiger bei großen JC, Ausreißer der relativen Fehler eher bei kleinen JC auf. Betrachtet man Daten, die sowohl bei absoluten als auch prozentualen Fehlern als Ausreißer klassifiziert werden, sind bei dem vorgeschlagenen Imputationsverfahren lediglich 0,7 % der geschätzten Datenpunkte auffällig. Simuliert man testhalber längere Datenlücken, so tendiert die Methode zu höheren Schätzfehlern, sodass Integrationen stärker unter- bzw. überschätzt würden.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei Integrationen von zufriedenstellender Qualität ist. Denn die geschlechtsspezifischen Anteilswerte variieren von August 2009 bis August 2013 mäßig. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 28,1 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats zwischen 4,7 und 15,7 Prozentpunkten. Bis auf 17 Datenpunkte werden bei den vorliegenden Datenausfällen WZ2- oder WZ1-Daten mit einer noch guten Variabilitätsspanne unter 10 % herangezogen. Lediglich bei elf Datenpunkten wird auf die Anteile aus den Vormonaten ausgewichen. Deren Anteilsinformation hat mit einer fast 30 %igen Variabilitätsspanne eine relativ geringe Güte.

6.1.4 Fazit, Ausblick

Bei der Schätzung der Integrationen produziert das vorgeschlagene Datenimputationsverfahren im Median einen betragsmäßigen Schätzfehler von 14 Fällen bzw. 9,6 %. Das Modell bietet mit den vorliegenden Mitteln das bestmögliche Verfahren für die Schätzung fehlender Integrationswerte. Auf Basis dieses Schätzkonzepts und der bisherigen Datenlückenmuster ist von einer befriedigenden Schätzgüte auszugehen. Mit anwachsender Zeitreihe ist zu prüfen, ob mit anderen Modellen gegebenenfalls eine verbesserte Schätzung erreicht werden kann.

6.2 Integrationen von alleinerziehenden erwerbsfähigen Leistungsberechtigten

6.2.1 Analyse fehlender Werte

Die Zahl der Integrationen von alleinerziehenden ELB ist definiert als alleinerziehende ELB aus dem Vormonat, die eine Erwerbstätigkeit aufgenommen haben. Es handelt sich somit um eine Teilgröße der Integrationen insgesamt. Hier liegen dieselben vereinzelten Datenlücken und derselbe sehr geringe Ausfallanteil von 0,33 % vor (vgl. Abschnitt 6.1.1).

6.2.2 Schätzmodell[78]

Da es sich bei den Integrationen der Alleinerziehenden um eine – sehr kleine – Teilgröße der Integrationen insgesamt handelt, sind die Fallzahlen sehr klein. So fallen im Zeitraum von September 2010 bis August 2013[79] 19,6 % der beobachteten Fälle kleiner als zehn aus. Der Median über alle JC liegt für den Beobachtungszeitraum bei 19 Integrationen Alleinerziehender. Deshalb ist die naheliegende Schätzung fehlender Werte per Anteilswertverfahren[80] hier nicht sinnvoll: Durch die sehr kleinen Fallzahlen variiert der Anteil von Integrationen Alleinerziehender an allen Integrationen in den JC über die Zeit extrem. Selbst geringfügige Schwankungen im Absolutwert zwischen einzelnen Berichtsmonaten oder zwischen Wartezeitständen innerhalb eines Berichtsmonats schlagen sich in sehr hohen prozentualen Schwankungen nieder.

Dennoch bieten Informationen aus der näheren Vergangenheit der Datenlücke eine Brücke zu Schätzwerten für fehlende Werte. Mit der zweiten Entwicklungsstufe der Imputationsverfahren wurde für die Zahl der Integrationen Alleinerziehender zunächst ein Ersatzwertverfahren eingeführt, bei dem – teils mit Anpassungen – auf Berichtsmonatswerte aus früheren Wartezeitständen (WZ<3) und auf Vormonatsdaten zurückgegriffen wurde. Bei Verwendung der Daten ohne Wartezeit und der Daten mit Wartezeit von einem Monat wurde die dort übliche zeitliche Untererfassung[81] mithilfe eines relativ groben Anpassungsfaktors ausgeglichen: Die untererfassten Werte wurden mit berichtsmonats- und JC-spezifischen Füllgradkehrwerten der verwandten Gesamtgröße der Integrationen multipliziert. Der Rückgriff auf die Untererfassungs- bzw. Füllgradinformationen bei der Gesamtgröße der Integrationen lag nahe: Die Untererfassung verhält sich bei Gesamt- und Teilgröße sehr ähnlich.[82] Über die Integrationsschätzung (vgl. Abschnitt 6.1) sind die entsprechenden Informationen immer verfügbar. Zudem bietet der Rückgriff auf die Untererfassungsinformationen bei der Gesamtgröße mehr Robustheit wegen der größeren Fallzahlen.

In der dritten Entwicklungsstufe hat dieses Ersatzwertverfahren eine Verfeinerung erfahren Vertiefte Analysen haben gezeigt, dass ein regressionsbasiertes Schätzverfahren (Methode der kleinsten Quadrate)

---

[78] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[79] Analog zur Zeitreihe für Integrationen insgesamt beginnt die Zeitreihe für Integrationen von Alleinerziehenden später als für die meisten hier dargestellten Grunddaten. Die Voraussetzungen für eine einheitliche Ermittlung für die Quellverfahren VerBIS und XSozial waren erst ab Ende des Jahres 2010 vollständig gegeben. Um die Zeitreihe für die Lückenanalyse zu verlängern, wurden die geringfügig unterschiedlichen Ermittlungsweisen in den vier Monaten zuvor als tolerabel erachtet.

[80] Vgl. Vorgehen nach der Anteilswertmethode bei der Schätzung fehlender Werte nach Geschlecht in Abschnitt 2.3 bzw. bei alleinerziehenden ELB in Abschnitt 4.2.2.

[81] Die Berichtsmonatswerte mit verkürzter Wartezeit weisen bei Integrationen Alleinerziehender in der Regel eine merkliche Untererfassung auf, die jedoch mit wachsender Wartezeit schwindet. Sie liegt im Zeitraum von September 2010 bis August 2013 über alle JC bei Daten ohne Wartezeit im Median bei 27,9 %, fällt auf 7,7 % in WZ1 und weicht in WZ2 der annähernden Vollerfassung. Eine Übererfassung kommt nur in den Randbereichen der Verteilung bei WZ1 und WZ2 vor.

[82] Die Ergebnisse der Untererfassungsanalyse bei der Gesamtgröße der Integrationen unterscheiden sich nur unwesentlich von den eben genannten: Die Untererfassung liegt im Zeitraum von September 2010 bis August 2013 über alle JC bei Daten ohne Wartezeit im Median bei 27,5 %, fällt auf 7,5 % in WZ1 und weicht in WZ2 der annähernden Vollerfassung. Eine Übererfassung kommt nur in den Randbereichen der Verteilung bei WZ1 und WZ2 vor.

mit einigen ergänzenden Prädiktoren zu verbesserten Schätzergebnissen führt. Die Auswahl der Prädiktoren erfolgt aufgrund sachlogischer Überlegungen sowie messbarer Zusammenhänge mit der Zielgröße. Als abhängige Variable fungiert die Zahl der Integrationen Alleinerziehender. Die Schätzung folgt dem Schema:

Sind die Daten im Referenzmonat unplausibel, so wird der regressionsbasierte Schätzwert verwendet auf Basis der

1. WZ2-Daten, falls diese auch fehlen
2. WZ1-Daten, falls diese auch fehlen
3. WZ0-Daten, falls diese auch fehlen,
4. dann werden die Vormonatsdaten mit WZ3 genutzt.

Als zentrale Prädiktoren wirken somit erneut – entsprechend ihrer Verfügbarkeit – die Berichtsmonatswerte mit verkürzter Wartezeit bzw. die Vormonatswerte. Ausgehend von den Überlegungen des Ersatzwertverfahrens in der zweiten Entwicklungsstufe geht in das Basismodell außerdem die jeweilige „Nacherfassungsquote“ als Indiz für die jeweilige Untererfassung ein, wiederum im Rückgriff auf die Gesamtgröße der Integrationen mit den o. g. Vorteilen. Sie ist wie folgt definiert:

$$NEQ_i^{WZ0} = \frac{int_i^{WZ0} - int_i^{WZ3}}{int_i^{WZ3}} \qquad NEQ_i^{WZ1} = \frac{int_i^{WZ1} - int_i^{WZ3}}{int_i^{WZ3}} \qquad NEQ_i^{WZ2} = \frac{int_i^{WZ2} - int_i^{WZ3}}{int_i^{WZ3}} \tag{11}$$

Als Niveauvariable wirkt zudem die Zahl der alleinerziehenden ELB des Vormonats, aus deren Grundgesamtheit Integrationen Alleinerziehender überhaupt nur möglich sind. Als weitere Regressoren dienen der SGB-II-Vergleichstyp (vgl. Abschnitt 2.2) und das Bundesland, mit denen Informationen zu den wirtschaftlichen und gebietspolitischen Rahmenbedingungen des jeweiligen JC einfließen. Auch die integrationsverwandte Anzahl der Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II wurde teilweise als Verbesserungsfaktor für die Schätzung identifiziert.[83] Den starken Saisonschwankungen der Größe Rechnung tragend, erfolgt die Regression erfolgt monatsweise. Je nach Verfügbarkeit der Wartezeitinformationen variieren die Prädiktoren in den Schätzgleichungen:[84]

a) Wenn Daten aus verkürzter Wartezeit genutzt werden können

$$\begin{aligned} int_allein_{t,i} = {} & \beta_{0,i} + \beta_1 \cdot (int_allein^{WZ0..2})_{t,i} + \beta_2 \cdot (NEQ^{WZ0..2})_{t,i} + \beta_3 \cdot (elb_allein^{WZ3})_{t-1,i} \\ & + \beta_4 \cdot (VT)_{t,j(i)} + \beta_5 \cdot (BL)_{t,k(i)} + \varepsilon_{t,i} \end{aligned} \tag{12}$$

[83] Zum Wegfall der Aufstocker/innen aus der Zahl der „Abgänge aus Arbeitslosigkeit SGB II“ ab dem 1.1.2017 vgl. Fußnote 74 in Abschnitt 6.1.2.

[84] Die Auswahl des Schätzmodells zur Schätzung der Integrationen Alleinerziehender erfolgte schrittweise. Für eine ausführliche Erläuterung der schrittweisen Erweiterung vgl. Anhang 8.6.

b) Wenn keine Informationen aus kürzerer Wartezeit vorliegen

$$int_allein_{t,i} = \beta_{0,i} + \beta_1 \cdot (alo_abg^{WZ0})_{t,i} + \beta_2 \cdot (int_allein^{WZ3})_{t-1,i} + \beta_3 \cdot (elb_allein^{WZ3})_{t-1,i} + \beta_4 \cdot (VT)_{t,j(i)} + \beta_5 \cdot (BL)_{t,k(i)} + \varepsilon_{t,i} \quad (13)$$

mit:

int_allein	Integrationen Alleinerziehender
β_0 bis β_5	Regressionskoeffizienten
$NEQ^{WZ0..2}$	Nacherfassungsquote für den jeweiligen Wartezeitstand
elb_allein	Bestand an alleinerziehenden ELB
alo_abg^{WZ0}	Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II mit WZ0 (gemäß der amtlichen Arbeitsmarktstatistik ohne Wartezeit)
VT	SGB-II-Vergleichstyp
BL	Bundesland
i	JC
t	Zeitpunkt
$j(i)$	Nummerierung des SGB-II-Vergleichstyps j zu i
$k(i)$	Nummerierung des Bundeslandes k zu i
ε	Fehlerterm

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

6.2.3 Bewertung[85]

Für die quantitative Bewertung des regressionsbasierten Schätzverfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von Januar 2011 bis August 2013.[86] Für 30 der insgesamt 51 Datenausfälle kann mit dem Regressionsverfahren auf den WZ2-Wert zurückgegriffen werden. In vier Fällen wird der WZ1-Wert und in sechs weiteren Fällen der WZ0-Wert verwendet. In insgesamt elf Fällen liegen keine WZ<3-Daten vor, so dass das Verfahren hier mit den Vormonatswerten angewendet wird.

Für die quantitative Bewertung wird das vorgeschlagene Imputationsverfahren mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Es ist eine deutliche Überlegenheit des regressionsbasierten Schätzverfahrens gegenüber den Standardvergleichsmodellen erkennbar.[87] Da das Regressionsverfahren vornehmlich auf WZ2-Daten zurückgreifen kann, die in der Regel bereits sehr dicht am endgültigen WZ3-Wert liegen, ist die erzielte hohe Schätzgüte allerdings auch wenig überraschend. Im Median liegt der betragsmäßige Fehler des regressionsbasierten Schätzverfahrens bei einer Integration bzw. 2,0 %. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/- 8 %. Dies spricht für einen geringen Schätzfehler im Allgemeinen. Gleichwohl existieren Ausreißer, die zu sehr hohen relativen Abweichungen (bis 350 %) führen. Hierbei ist zu berücksichtigen, dass der simulierte Schätzwert in 5.599 von 12.251

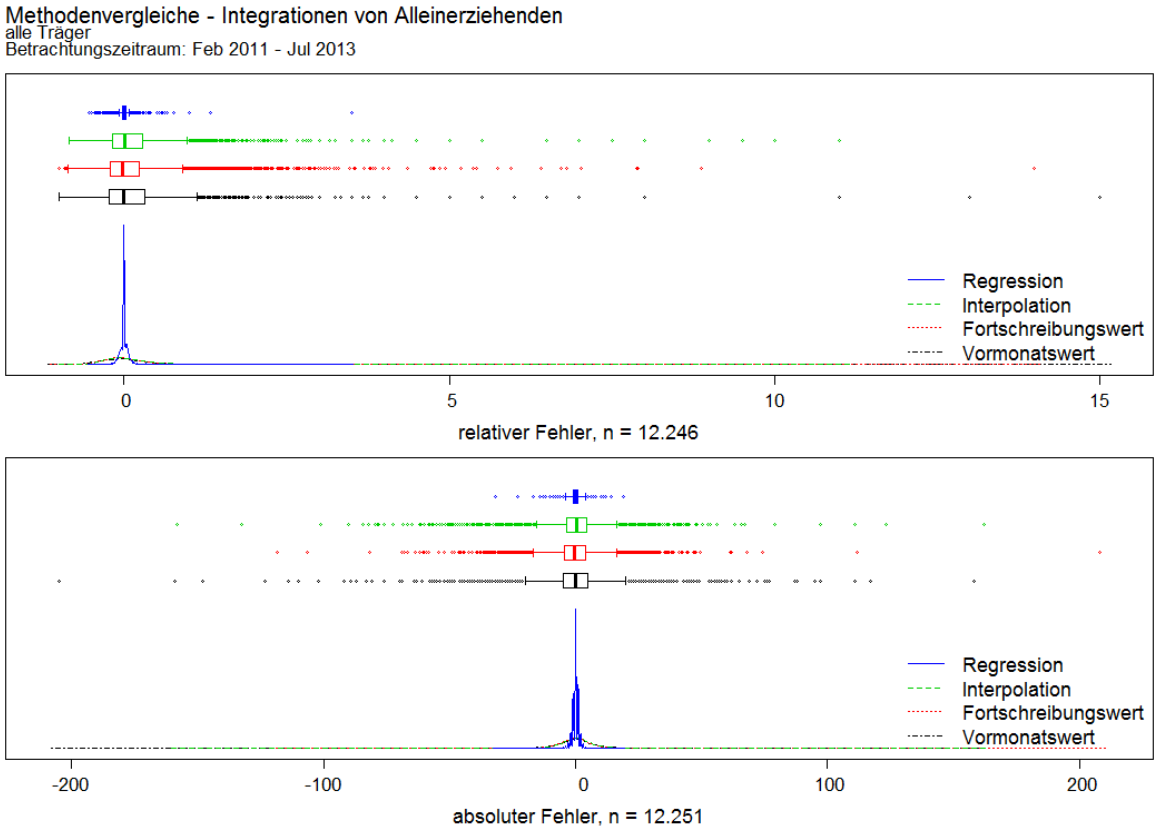
[85] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[86] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Februar 2011 bis Juli 2013.

[87] Auch gegenüber dem in der zweiten Entwicklungsstufe der Imputationsverfahren vorgeschlagenen gröberen Ersatzwertverfahren ergibt sich eine deutliche Verbesserung der Schätzgüte (vgl. Anhang 8.6).

Fällen (45,7 %) exakt mit dem tatsächlichen Wert übereinstimmt – die Abweichung beträgt hier also genau Null. Somit liegt auch der Interquartilsabstand sehr dicht am Nullwert. Bei dieser engen Fehlerverteilung fallen auch Werte mit sehr geringen Abweichungen unter die Ausreißer. Insgesamt sind die hohen relativen Abweichungen zudem natürlich auch auf die sehr kleinen Fallzahlen zurückzuführen. Betrachtet man Daten, die sowohl beim prozentualen als auch beim absoluten Fehler als Ausreißer klassifiziert werden, sind bei dem vorgeschlagenen Imputationsverfahren lediglich 0,6 % der geschätzten Datenpunkte auffällig. Somit bleibt es – auch unter der Berücksichtigung von Ausreißern – bei der Einschätzung, dass der Schätzfehler gering ist.

Abbildung 21: Integrationen von Alleinerziehenden – Methodenvergleich Fehlerverteilungen



Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei der Ergänzungsgröße Integrationen von Alleinerziehenden von befriedigender Qualität ist. Die geschlechtsspezifischen Anteilswerte variieren von September 2010 bis August 2013 stark, wiederum ein Effekt der sehr kleinen Fallzahlen. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den verschiedenen Wartzeitständen eines Berichtsmonats zwischen ca. sechs (WZ2) und 24 Prozentpunkten (WZ0). Die mittlere Variabilitätsspanne zwischen den Berichtsmonaten mit WZ3 liegt sogar bei fast 38 Prozentpunkten. Bei 30 Datenpunkten wird bei den vorliegenden 51 Datenausfällen auf WZ2-Daten mit einer noch guten Variabilitätsspanne von unter 10 % zurückgegriffen. In allen anderen Fällen müssen Anteilsinformationen mit geringer Güte verwendet werden.

6.2.4 Fazit, Ausblick

Bei der Schätzung der Integrationen von Alleinerziehenden produziert das regressionsbasierte Schätzverfahren im Median einen betragsmäßigen Schätzfehler von einer Integration bzw. 2,0 %. Es ist von einer befriedigenden Schätzgüte auszugehen. Angesichts der sehr kleinen Fallzahlen bietet das Modell mit den vorliegenden Mitteln das bestmögliche Verfahren für die Schätzung fehlender Werte.

6.3 Integrationen von Langzeitleistungsbeziehenden

6.3.1 Analyse fehlender Werte

Die Zahl der Integrationen von Langzeitleistungsbeziehenden (LZB)[88] ist definiert als LZB aus dem Vormonat, die eine Erwerbstätigkeit aufgenommen haben. Es handelt sich somit um eine Teilgröße der LZB insgesamt.

- Somit liegen von September 2010[89] bis Oktober 2015 zum einen dieselben Datenlücken vor wie bei LZB insgesamt in diesem Zeitraum, allerdings um einen Monat verschoben (92 Fälle, vgl. Abschnitt 4.3.1).
- Hinzu kommen 15 Datenlücken infolge fehlender Integrationswerte (vgl. Abschnitt 6.1.1).

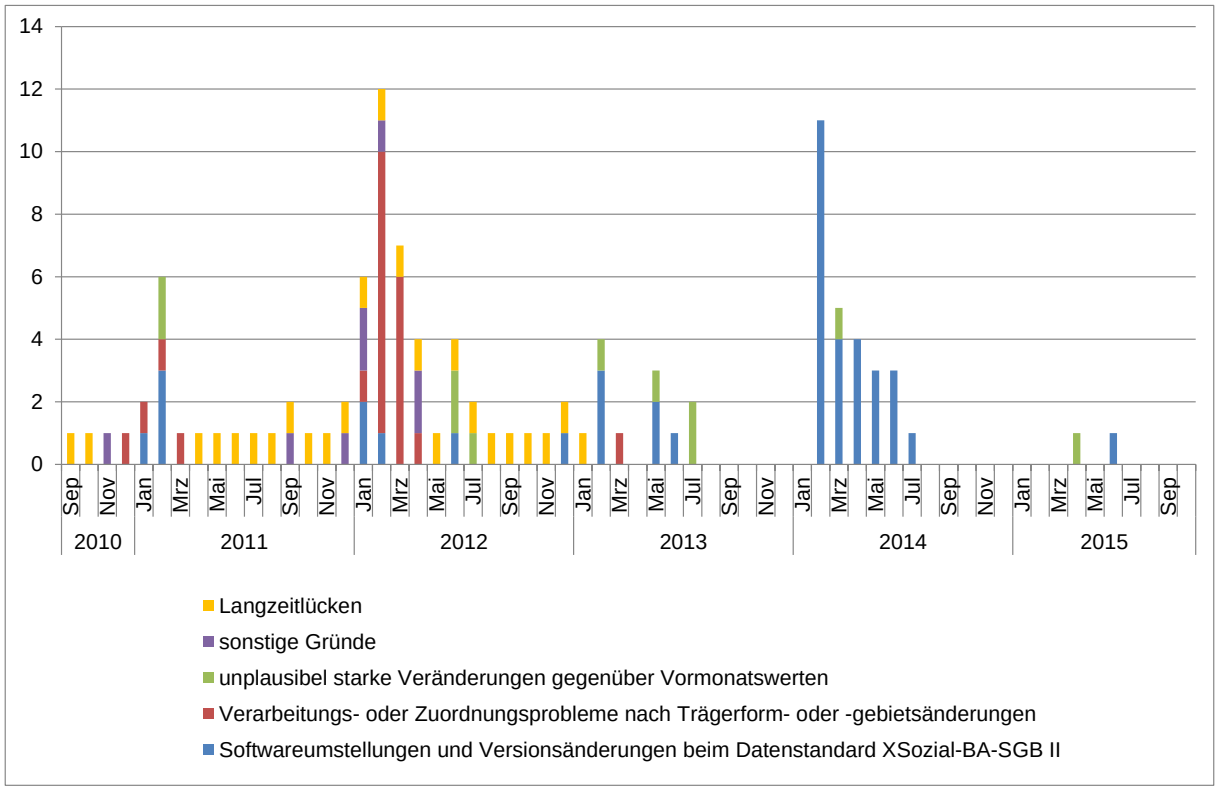
Von September 2010 bis Oktober 2015 liegen somit insgesamt 107 fehlende Werte vor. Bezogen auf 62 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,42 %. Dabei sind zwei JC von Langzeitlücken betroffen. Diese führen zu 24 fehlenden Berichtsmonatswerten bei Integrationen von LZB. 51 JC sind von insgesamt 83 isolierten Datenlücken betroffen. Diese häufen sich vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen für isolierte Datenlücken eine untergeordnete Rolle. Dabei fehlen bei 15 JC mehrfach Daten für Integrationen von LZB, wobei die maximale Häufigkeit bei fünf hintereinander gereihten Datenlücken für ein JC liegt. Allerdings führen diese Mehrfachausfälle nie zu den oben genannten Langzeitlücken, da in diesen Konstellationen die Nettodauern für den Leistungsbezug als ELB noch ermittelt werden können.

[88] Zum Messkonzept des Langzeitleistungsbezugs vgl. Abschnitt 4.3.1.

[89] Die Zeitreihe für Integrationen beginnt später als für die meisten hier dargestellten Grunddaten. Die Voraussetzungen für eine einheitliche Ermittlung der Integrationen für die Quellverfahren VerBIS und XSozial waren erst ab Ende des Jahres 2010 vollständig gegeben. Um die Zeitreihe für die Lückenanalyse zu verlängern, wurden die geringfügig unterschiedlichen Ermittlungsweisen in den vier Monaten zuvor als tolerabel erachtet.

Abbildung 22: Integrationen von LZB – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei Integrationen von LZB
Alle JC
Berichtsmonate September 2010 bis Oktober 2015



Ganz ähnlich wie beim Bestand von aktivierten LZB (vgl. Abschnitt 4.4.1) können bei Datenlücken für Integrationen von LZB folgende Fallkonstellationen auftreten (detaillierter nachfolgend auch in Abbildung 23):

Fallkonstellation 1: Der LZB-Bestand des Vormonats gilt als nicht plausibel. Hieraus lassen sich im Berichtsmonat keine plausiblen Integrationen bei LZB ableiten.

Fallkonstellation 2: Der LZB-Bestand des Vormonats gilt als plausibel und Integrationen des Berichtsmonats gelten als nicht plausibel.

Fallkonstellation 3: Der LZB-Bestand des Vormonats und Integrationen des Berichtsmonats gelten als plausibel.

Die Fallkonstellationen 1 und 2 führen zu einem Datenausfall bei den Integrationen von LZB.

6.3.2 Schätzmodell

Die beiden unterschiedenen Arten von Datenlücken – isolierte Datenlücken einerseits und Langzeitlücken andererseits – erfordern wiederum verschiedene Herangehensweisen bei der Datenimputation.

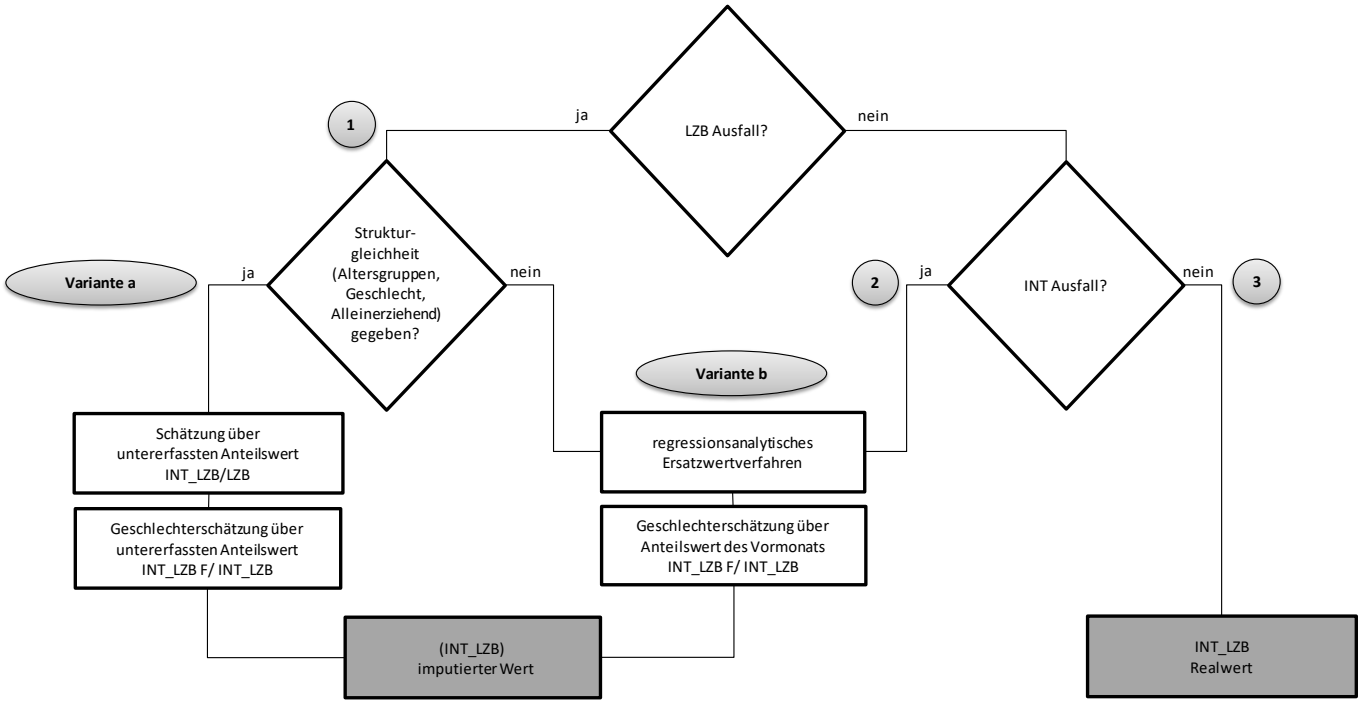
6.3.2.1 Schätzmodell bei isolierten Datenlücken

Das Vorgehen bei isolierten Datenlücken ist ähnlich wie bei Datenimputationen für aktivierte LZB (vgl. Abschnitt 4.4.2.1). Wenn der jeweilige Test auf Strukturgleichheit in den Vormonatsdaten der LZB es erlaubt, dann werden die Imputationen auch hier auf Basis der unvollständigen Datenlieferung mit dem Anteilswertverfahren berechnet (Variante a). Lehnt der Test die Hypothese der Strukturgleichheit ab, werden die Integrationen von LZB analog zu Integrationen von Alleinerziehenden (vgl. Abschnitt 6.2.2) mit Hilfe eines regressionsbasierten Schätzmodells ermittelt (Variante b).

6.3.2.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren

Der Entscheidungsbaum in Abbildung 23 stellt die Prüfschritte dar, anhand derer das Schätzmodell festgelegt wird.

Abbildung 23: Integrationen von LZB – Schätzmodell/Entscheidungsbaum



Bei Fallkonstellation 1 wird im ersten Schritt die Hypothese auf Strukturgleichheit überprüft. Wenn sie nicht abgelehnt wird, wird angenommen, dass Strukturgleichheit vorliegt und der beobachtete Ausfall zufällig ist. Die Anteilsverteilungen aus der unvollständigen Datenlieferung gelten somit als vertrauenswürdige Annäherung an die Verteilungen, die sich aus einer plausiblen Datenlieferung des JC ergeben hätten, wenn alle Personendatensätze vollständig geliefert worden wären. In diesem Fall greift wieder das Anteilswertverfahren: Der jeweilige Schätzwert für LZB-Integrationen ergibt sich als Produkt aus dem per Fortschrittsverfahrenen geschätzten LZB-Bestand (vgl. Abschnitt 4.3.2.1) und dem Anteil von LZB mit Integrationen an allen LZB, der aus der unvollständigen Datenlieferung stammt.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt auch hier für Frauen nach demselben Verfahren wie für die Gesamtschätzung. Die Anzahl der Männer ergibt sich als Differenz zwischen dem Gesamtwert und demjenigen der Frauen.[90]

6.3.2.1.2 Variante b) Regressionsbasiertes Schätzverfahren[91]

Wenn die Strukturen nicht gleich sind, werden die Integrationen von LZB mit Hilfe eines regressionsbasierten Schätzverfahrens (Methode der kleinsten Quadrate) ermittelt. Grundsätzlich ist das Vorgehen das gleiche wie bei Datenimputationen für Integrationen Alleinerziehender (vgl. Abschnitt 6.2.2). Als abhängige Variable fungiert hier die Zahl der Integrationen von LZB. Als zentrale Prädiktoren wirken erneut – entsprechend ihrer Verfügbarkeit – die Berichtsmonatswerte mit verkürzter Wartezeit sowie die jeweiligen „Nacherfassungsquoten“ im Rückgriff auf die Gesamtgröße der Integrationen.[92] Andere Prädiktoren werden aufgrund sachlogischer Überlegungen sowie messbarer Zusammenhänge mit der Zielgröße ausgewählt: Als Niveauvariable fungieren die Integrationen von LZB des Vormonats. Erneut fließen der SGB-II-Vergleichstyp (vgl. Abschnitt 2.2) und das Bundesland in das Modell ein. Auch die integrationsverwandte Anzahl der Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II wurde teilweise wieder als Verbesserungsfaktor für die Schätzung identifiziert.[93] Den starken Saisonschwankungen der Größe Rechnung tragend, erfolgt die Regression monatsweise. Je nach Verfügbarkeit der Wartezeitinformationen variieren die Prädiktoren in den Schätzgleichungen.[94]

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

---

[90] Somit wird hier vom üblichen Muster bei der Geschlechterschätzung (vgl. Abschnitt 2.3) abgewichen. Im Vertrauen auf den Test auf Strukturgleichheit ist dieses Vorgehen aber sachlogisch nur konsequent.

[91] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[92] Die Ergebnisse der Untererfassungsanalyse unterscheiden sich nur unwesentlich von jenen bei der Gesamtgröße der Integrationen oder jenen bei der Teilgröße der Integrationen Alleinerziehender: Die Berichtsmonatswerte mit verkürzter Wartezeit weisen bei Integrationen von LZB in der Regel eine merkliche Untererfassung auf, die jedoch mit wachsender Wartezeit schwindet. Sie liegt im Zeitraum von September 2010 bis August 2013 über alle JC bei Daten ohne Wartezeit im Median bei 26,6 %, fällt auf 7,7 % in WZ1 und weicht mit 2,9 % in WZ2 der annähernden Vollerfassung. Übererfassungen kommen vereinzelt in den Randbereichen der Verteilung bei WZ1 und WZ2 vor.

[93] Zum Wegfall der Aufstocker/innen aus der Zahl der „Abgänge aus Arbeitslosigkeit SGB II“ ab dem 1.1.2017 vgl. Fußnote 74 in Abschnitt 6.1.2.

[94] Die Auswahl des Schätzmodells zur Schätzung der Integrationen von LZB erfolgte schrittweise. Für eine ausführliche Erläuterung der schrittweisen Erweiterung vgl. Anhang 8.7.

6.3.2.2 Schätzmodell bei Langzeitlücken

Das Vorgehen bei Langzeitlücken ist dasselbe wie bei Datenimputationen für Langzeitlücken bei LZB insgesamt (vgl. Abschnitt 4.3.2.2). Nach der initiierenden Konzeption der kombinatorischen Abfragelogik kann das JC-individuelle Vorgehen ohne großen Aufwand monatlich durchgeführt und auch für die Teilgruppe der LZB mit Integrationen angewendet werden. Entsprechend erfolgt die Differenzierung der Datenimputationswerte nach Geschlecht anhand des gleichen Rekonstruktionsverfahrens wie für die Gesamtgröße.

6.3.3 Bewertung

6.3.3.1 Bewertung des Schätzmodells bei isolierten Datenlücken[95]

6.3.3.1.1 Variante a) Test auf Strukturgleichheit und Anteilswertverfahren

Analog zur Bewertung des Schätzmodells bei aktivierten LZB (vgl. Abschnitt 4.4.3.1) ist auch hier davon auszugehen, dass das vorgeschlagene Verfahren gute Schätzergebnisse bei maximaler Ausnutzung der Informationen des jeweiligen JC bietet. Nach dem vorgeschalteten Test auf Strukturgleichheit wird dieses Verfahren meistens – bei 30 von 51 (über 60 %) isolierten Datenlücken – zwischen Januar 2011 und August 2013 eingesetzt. Da die Differenzierung der Datenimputationswerte nach Geschlecht – abweichend vom üblichen Muster bei der Geschlechterschätzung (vgl. Abschnitt 2.3) – analog zum Verfahren für die Gesamtschätzung erfolgt, ist hier von der gleichen guten Schätzgüte auszugehen – eventuell abgeschwächt infolge sehr geringer Fallzahlen.

6.3.3.1.2 Variante b) Regressionsbasiertes Schätzverfahren

Für die quantitative Bewertung des regressionsbasierten Schätzverfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von Januar 2011 bis August 2013.[96] Für 30 der hier insgesamt betrachteten 49 isolierten Datenausfälle könnte im Anwendungsfall auf den WZ2-Wert zurückgegriffen werden. In vier Fällen würde der WZ1-

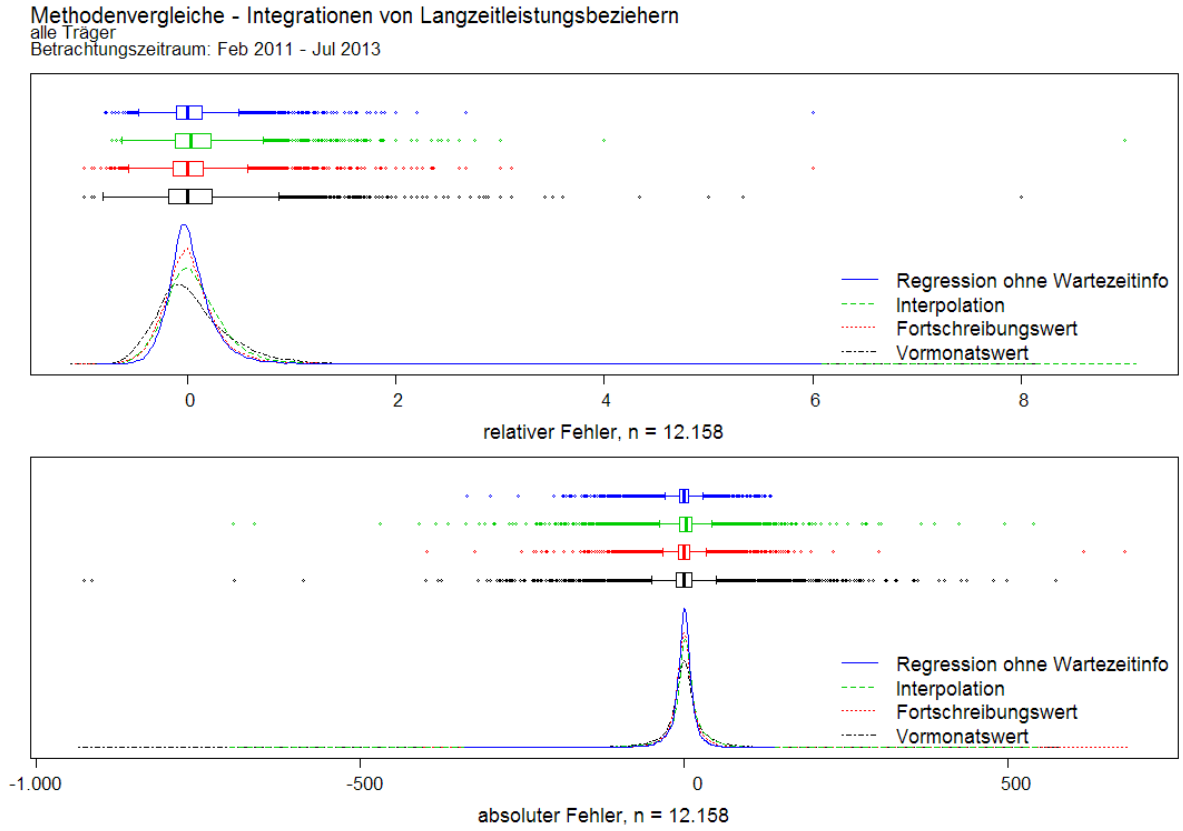
[95] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[96] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Februar 2011 bis Juli 2013.

Wert und in fünf weiteren Fällen der WZ0-Wert verwendet. In insgesamt zehn Fällen liegen keine WZ<3-Daten vor, so dass das Verfahren hier mit den Vormonatswerten angewendet werden würde.

Das vorgeschlagene Imputationsverfahren wird mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Die Grafik enthält für das vorgeschlagene regressionsbasierte Schätzverfahren nur Simulationen für den Schlechtfall, bei dem auf die Vormonatswerte zurückgegriffen wird. Selbst in diesem Fall ist das regressionsbasierte Schätzverfahren allen Standardvergleichsmodellen überlegen. Im Schlechtfall liegt der betragsmäßige Fehler des regressionsbasierten Schätzverfahrens bei sieben Integrationen bzw. 11,9 %. Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/- 48 %. Dies spricht für einen merklichen Schätzfehler im Allgemeinen. Da jedoch in drei Viertel der Fälle, also meistens, auf die besseren WZ<3-Informationen zurückgegriffen würde, ist insgesamt in den Fällen, in denen das regressionsbasierte Schätzverfahren angewendet werden würde, von einer guten Schätzgüte auszugehen.[97]

Abbildung 24: Integrationen von LZB – Methodenvergleich Fehlerverteilungen beim regressionsbasierten Schätzverfahren



[97] Aus technischen Gründen ist es nicht möglich, den gesamten Schätzfehler über alle Kaskadenschritte hinweg (WZ2-Modell, WZ1-Modell, WZ0-Modell und Vormonatswert-Modell) zu bestimmen. Unterstellt man, die jeweiligen Schätzfehler gehen in den Gesamtfehler mit dem Gewicht der jeweiligen Anwendungshäufigkeit (30, vier, fünf und zehn) ein, liegt der gesamte betragsmäßige absolute Schätzfehler bei drei Integrationen bzw. der gesamte relative Schätzfehler bei etwa 4,2 %.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei der Ergänzungsgröße Integrationen von LZB von mäßiger Qualität ist. Die geschlechtsspezifischen Anteilswerte variieren von September 2010 bis August 2013 stark, wiederum ein Effekt der sehr kleinen Fallzahlen. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den verschiedenen Wartzeitständen eines Berichtsmonats zwischen ca. sieben (WZ2) und 23 Prozentpunkten (WZ0). Die mittlere Variabilitätsspanne zwischen den Berichtsmonaten mit WZ3 liegt sogar bei fast 45 Prozentpunkten. Bei mehr als der Hälfte der Datenpunkte würde bei den vorliegenden isolierten Datenausfällen auf WZ2-Daten mit einer noch guten Variabilitätsspanne von unter 10 % zurückgegriffen. In allen anderen Fällen müssten Anteilsinformationen mit geringer Güte verwendet werden.

6.3.3.2 Bewertung des Schätzmodells bei Langzeitlücken

Das vorgeschlagene bestandsbasierte Rekonstruktionsverfahren wird auch für Auswertungen von Integrationen von LZB und deren Geschlechterschätzung als sehr gut bewertet, mit gegen Ende des Schätzzeitraums steigender aber akzeptabler Schätzunsicherheit. Bei Anwendung statistisch-mathematischer Schätzverfahren – welcher Art auch immer – wäre insbesondere am Ende des gegebenenfalls extrem langen Schätzzeitraums von 24 Monaten von deutlich größeren Schätzfehlern auszugehen.

6.3.4 Fazit, Ausblick[98]

Für die Datenimputation bei isolierten Datenlücken von LZB-Integrationen wird in den meisten Fällen (bislang 30) ein Anteilswertverfahren nach Test auf Strukturgleichheit angewendet. Die verbleibenden isolierten Datenlücken (bislang 21) werden mit einem regressionsbasierten Schätzverfahren geschlossen. Bei beiden Verfahren ist von einer guten Schätzgüte auszugehen.

Für die Datenimputation bei 22 Langzeitlücken von LZB-Integrationen wird ein bestandszahlenbasiertes Rekonstruktionsverfahren genutzt, das JC-individuell angewendet wird. Es wird angesichts des extrem langen Schätzzeitraums als sehr gut bewertet.

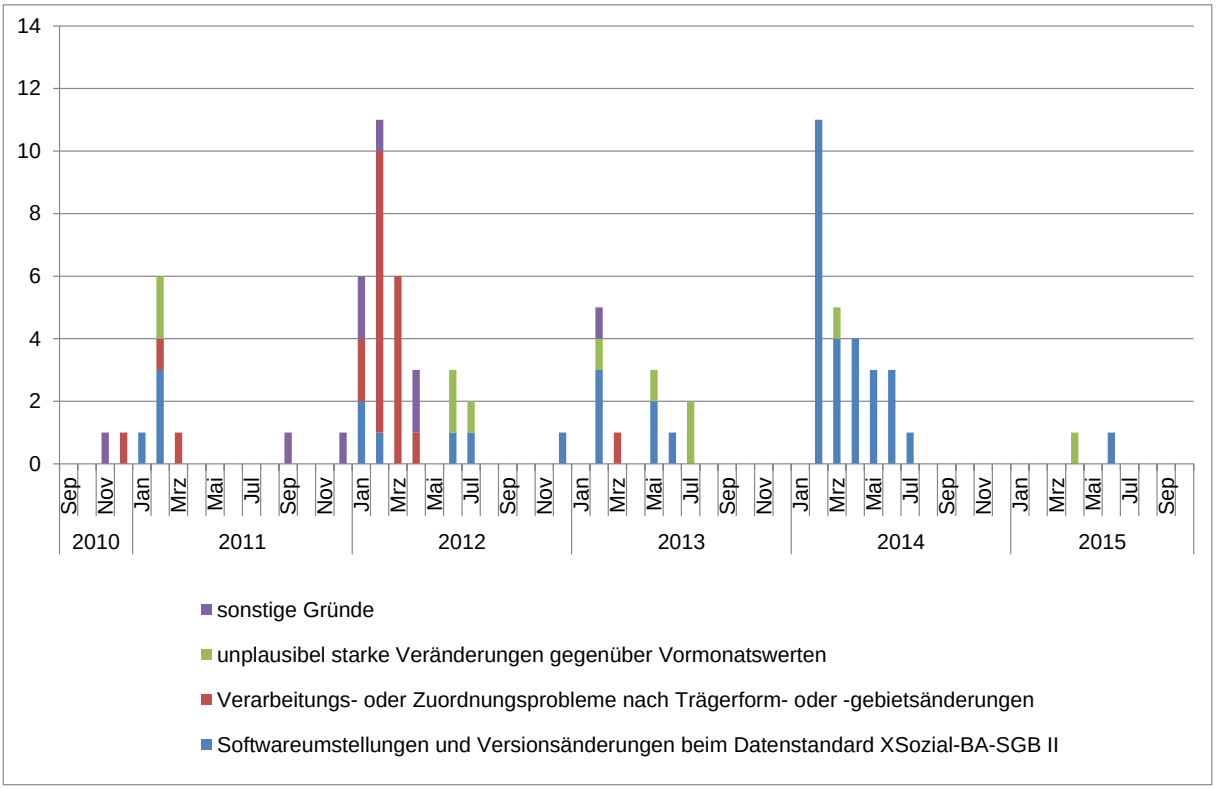
[98] Die in diesem Abschnitt genannten Anwendungshäufigkeiten der verschiedenen Datenimputationsmethoden für Datenlücken bei aktivierten LZB beziehen sich auf Daten bis August 2013 (vgl. Version 3.0 dieser Methodenübersicht, Stand Mai 2014) vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

6.4 Eintritte in geringfügige Beschäftigung

6.4.1 Analyse fehlender Werte

Abbildung 25: Eintritte in geB – fehlende Werte

Häufigkeiten und Zeitpunkte von Datenlücken bei geringfügigen Beschäftigungsaufnahmen
Alle JC
Berichtsmonate September 2010 bis Oktober 2015



Bei den geringfügigen Beschäftigungsaufnahmen (geB, definiert als ELB aus dem Vormonat, die eine geringfügige Beschäftigung aufgenommen haben) kommt es bisher nur vereinzelt zu Datenlücken. Von September 2010[99] bis Oktober 2015 liegen 85 fehlende Werte vor. Bezogen auf 62 Berichtsmonate und 408 JC bedeutet dies einen sehr geringen Ausfallanteil von 0,34 %. Die Datenlücken häufen sich vor allem Anfang 2012, was auf Verarbeitungs- oder Zuordnungsprobleme nach den zahlreichen Trägerformänderungen zurückzuführen ist, sowie Anfang 2014, was einer Häufung von Softwareproblemen beim Datenstandard XSozial-BA-SGB II geschuldet ist. Weitere Ursachen spielen eine untergeordnete Rolle. In etwa 80 % der Fälle (in 68 Fällen) kommt es wegen unplausibler Vormonats-Eckwerte (das heißt wegen fehlerhafter Einzeldaten für die ELB) zu fehlenden Grundwerten für geB, weil sich aus der zugrundeliegenden Gesamtheit keine gültige Zahl aufgenommener Beschäftigungsaufnahmen bestimmen lässt. Nur in 17 Fällen gehen die Datenlücken direkt auf unplausible Informationen zu geB zurück.

[99] Die Zeitreihe für geB beginnt später als für die meisten hier dargestellten Grunddaten. Die Voraussetzungen für eine einheitliche Ermittlung für die Quellverfahren VerBIS und XSozial waren erst ab Ende des Jahres 2010 vollständig gegeben. Um die Zeitreihe für die Lückenanalyse verlängern, wurden die geringfügig unterschiedlichen Ermittlungsweisen in den vier Monaten zuvor als tolerabel erachtet.

Die Lücken verteilen sich auf 50 JC (etwa 12 % der JC). Bei 22 JC fehlen mehrfach Daten für Integrationen, wobei die maximale Häufigkeit bei fünf hintereinander gereihten Datenlücken für ein JC liegt.

6.4.2 Schätzmodell

Analysen zeigen, dass die Größe der geB von saisonalen Effekten mit Spitzen insbesondere vor dem Weihnachtsgeschäft und im Frühjahr geprägt ist. Es liegt nahe, analog zum Schätzverfahren bei Integrationen (vgl. Abschnitt 6.1.2) die Teilverfahren oder das Gesamtverfahren für die Größe der geB zu testen.

Bei den Analysen erwies sich das Teilverfahren der k-nächste-Nachbarn-Fortschreibung (knN-Fortschreibung) für die Größe der geB als am besten geeignet. Dabei wird die Entwicklung der geB ähnlicher Nachbarn auf das betreffende JC übertragen. Der Schätzwert ergibt sich als Produkt des Vormonatswertes und der geschätzten mittleren Veränderung zum Vormonat der k ähnlichsten JC. Die Ähnlichkeit wird anhand der Entwicklung der letzten drei verfügbaren Monate und dem Vorjahresmonat – sofern vorhanden – ermittelt. Die optimale Anzahl k der ähnlichsten JC wird dabei monatsweise so bestimmt, dass der mittlere quadratische Fehler (auf Basis des relativen Fehlers) minimiert wird.[100] Die Zeitpunkte werden systematisch aufsteigend abgearbeitet. Dies bedingt – vorteilhaft im Rahmen amtlicher Berichterstattung – die Festschreibung der Schätzwerte.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

6.4.3 Bewertung[101]

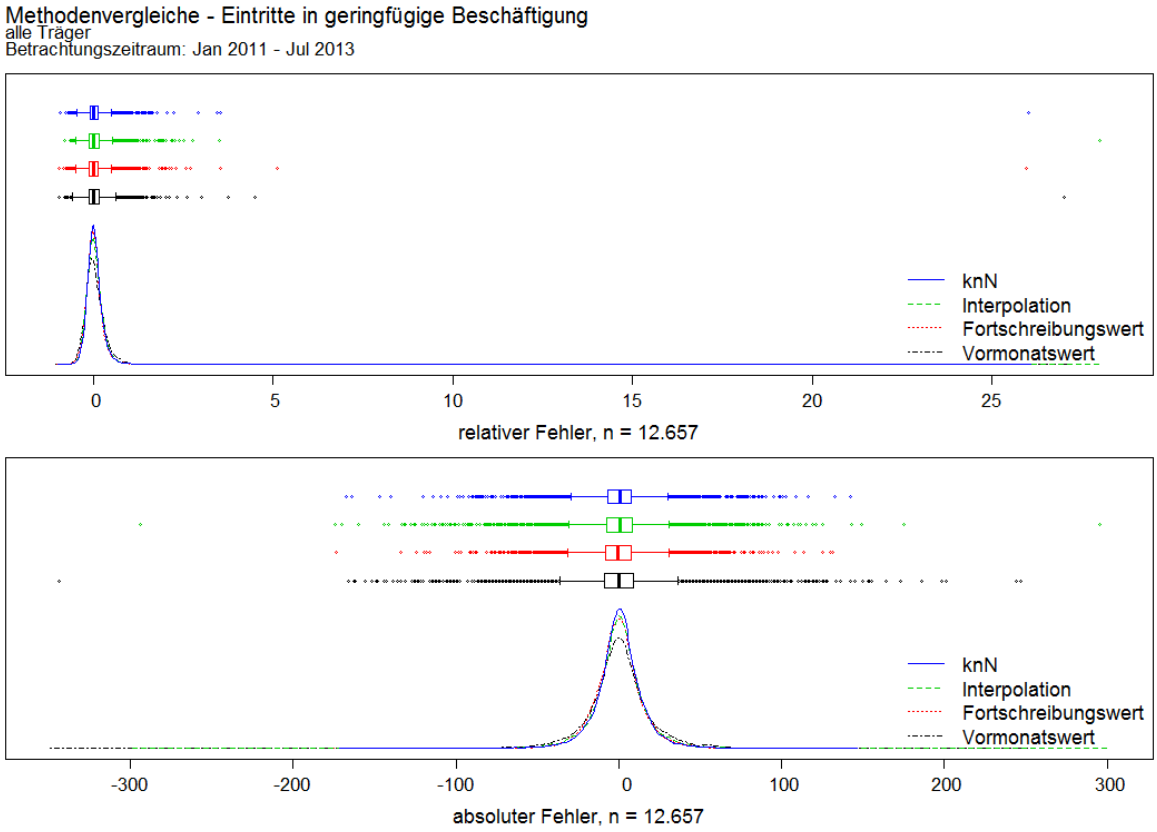
Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von September 2010 bis August 2013.[102] Für die quantitative Bewertung wird das Fortschreibungsmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Im Ergebnis ist das vorgeschlagene knN-Fortschreibungsverfahren den anderen Modellen überlegen: die Verteilung der absoluten und relativen Fehler weist hier die stärkste Konzentration bei null auf. Die Überlegenheit gegenüber den Standardvergleichsmodellen rührt sicher daher, dass das vorgeschlagene Imputationsverfahren die kenngrößenspezifischen Abhängigkeiten der JC untereinander besser herauszuarbeiten vermag.

[100] Für eine ausführliche Erläuterung des k-nächste Nachbarn-Verfahrens am Beispielthema Integrationen siehe Anhang 0.

[101] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[102] Da das knN-Verfahren aus konzeptuellen Gründen einen 4-monatigen Vorlauf benötigt und die verglichenen Verfahren einen 1-monatigen Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Januar 2011 bis Juli 2013. Das Anwachsen der bislang relativ kurzen Zeitreihen bei geB wurde genutzt: Gegenüber der Version 2.0 dieser Methodenübersicht (Stand Mai 2013) fließen zwölf weitere Berichtsmonate in die Analysen ein.

Abbildung 26: Eintritte in geB – Methodenvergleich Fehlerverteilungen



Bei der Anwendung des vorgeschlagenen Verfahrens wird die tatsächliche Anzahl an Eintritten in geB im Median und Betrag um sieben Eintritte (11,5 %) verfehlt, wobei das Bild von JC zu JC variiert. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/- 45 %. Dies spricht für einen merklichen Schätzfehler im Allgemeinen. Ausreißer der absoluten Fehler treten häufiger bei großen JC, Ausreißer der relativen Fehler eher bei kleinen JC auf. Betrachtet man Daten, die sowohl bei absoluten als auch prozentualen Fehlern als Ausreißer klassifiziert werden, sind bei dem kombinierten zweistufigen Modell lediglich 0,3 % der geschätzten Datenpunkte auffällig. Simuliert man testhalber längere Datenlücken, so tendiert die Methode zu höheren Schätzfehlern, sodass Eintritte in geB stärker unter- bzw. überschätzt würden.

Wie bei allen Fortschreibungsmethoden übertragen sich auch beim knN-Verfahren ungewöhnlich hohe oder niedrige Vormonatswerte in großem Maße auf den Schätzwert und darum können besonders sprunghafte Zeitreihen schwer geschätzt werden. Dieses Problem tritt vor allem bei JC auf, die besonders kleine Fallzahlen aufweisen.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei der Größe der geB von mäßiger Qualität ist. Denn die geschlechtsspezifischen Anteilswerte variieren von August 2009 bis August 2013 deutlich. Die mittlere Variabilitätsspanne über alle

JC in der Definition nach Abschnitt 2.3 beträgt in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 41 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats zwischen neun (WZ2) und 32 Prozentpunkten (WZ0). Bis auf 24 Datenpunkte werden bei den vorliegenden Datenausfällen WZ2-Daten mit einer noch guten Variabilitätsspanne unter 10 % herangezogen. Lediglich bei zwölf Datenpunkten muss auf die Anteile aus den Vormonaten ausgewichen, deren Anteilsinformation eine noch sehr geringe Güte hat.

6.4.4 Fazit, Ausblick

Bei der Schätzung der Eintritte in geB produziert das vorgeschlagene Datenimputationsverfahren im Median einen betragsmäßigen Schätzfehler von sieben Fällen bzw. 11,5 %. Das Modell bietet mit den vorliegenden Mitteln das bestmögliche Verfahren für die Schätzung fehlender Daten. Auf Basis dieses Schätzkonzepts und der bisherigen Datenlückenmuster ist von einer befriedigenden Schätzgüte auszugehen. Mit anwachsender Zeitreihe ist zu prüfen, ob mit anderen Modellen gegebenenfalls eine verbesserte Schätzung erreicht werden kann.

6.5 Eintritte in öffentlich geförderte Beschäftigung

6.5.1 Analyse fehlender Werte

Bei den öffentlich geförderten Beschäftigungsaufnahmen (ögB, definiert als ELB aus dem Vormonat, die eine öffentlich geförderte Beschäftigung aufgenommen haben) liegen von August 2009 bis Oktober 2015 dieselben Datenlücken vor wie bei ELB, allerdings um einen Monat verschoben (vgl. Abschnitt 4.1.1). Es sind insgesamt 69 Datenlücken. Bezogen auf 75 Berichtsmonate und 408 JC liegt somit ein sehr geringer Ausfallanteil von 0,23 % vor.

Ansonsten liegen aus den Validierungsergebnissen der Förderstatistik (FST) der Bundesagentur für Arbeit für 136 Berichtsmonatswerte Hinweise auf Einschränkungen in der Aussagekraft der Monatswerte vor. In diesen Fällen werden die Werte sowie die daraus abgeleitete Ergänzungsgröße und Jahresfortschrittskennzahl in den Analysewerkzeugen geklammert ausgewiesen.

6.5.2 Schätzmodell[103]

Die Daten mit verkürzter Wartezeit zeigen insbesondere für den Wartezeitstand mit zweimonatiger Wartezeit (WZ2) eine geringe Differenz zum endgültigen Wert. Der durchschnittliche Wert der Über- bzw. Unterfassung liegt über alle JC und Zeitpunkte (September 2010 bis August 2012) bei etwa 1,5 Prozent. Die Daten mit noch kürzeren Wartezeiten fallen diesbezüglich etwas ab (WZ1: 4,3 %, WZ0: 17,3 %) und werden nicht für ein Ersetzungswertverfahren herangezogen. Als eine weitere Quelle für Ersatzwerte können die Eintritte in Maßnahmen der Arbeitsmarktpolitik, die für die Bestimmung der Eintritte in ögB relevant sind, aus der FST herangezogen werden.[104] Analysen zeigen, dass diese Daten ebenfalls eine hohe Übereinstimmung zu den Daten der Eintritte in öffentlich geförderter Beschäftigung erzielen.

Die Eintritte in ögB unterliegen starken Schwankungen zwischen den Berichtsmonaten. Vor diesem Hintergrund scheidet die Nutzung eines Wertes aus dem Vormonat oder Vorjahresmonat aus. Aus der skizzierten Datenlage ergibt sich somit folgendes Schema zur Ersetzung nicht plausibler Werte:

Sind die Daten im Referenzmonat unplausibel, so wird der Wert geschätzt anhand von:

1. WZ2-Daten, falls diese auch fehlen
2. Daten aus der FST, falls diese auch fehlen
3. dem Median[105] aus den verfügbaren letzten zwölf Berichtsmonatswerten des JC.

Nach dem in Abschnitt 2.3 beschriebenen Anteilswertverfahren, um die Informationen der FST ergänzt, werden die relevanten Anteilswerte für die Differenzierung der Datenimputationswerte nach Geschlecht gewonnen. Dies stellt eine kleine – sachlogisch aber nur konsequente – Abwandlung des grundsätzlichen Verfahrens dar.

Die Schätzung der Anteilswerte für das Geschlecht erfolgt in folgender Reihenfolge:

1. WZ2-Daten, falls diese auch fehlen
2. WZ1-Daten, falls diese auch fehlen
3. Daten der FST, falls diese auch fehlen
4. WZ0-Daten, falls diese auch fehlen
5. Vormonatsdaten mit WZ3.

---

[103] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[104] „Arbeitsgelegenheiten“ in der Mehraufwandsvariante nach § 16d SGB II, Leistungen zur „Teilhabe am Arbeitsmarkt“ nach § 16i SGB II, Altfälle für die Bereiche „Arbeitsgelegenheiten in der Entgeltvariante“ (§ 16d SGB II a. F. bis 31.03.2012), „Beschäftigungszuschuss“ (§ 16a SGB II i. d. F. ab 01.10.2007 bis 31.12.2008, § 16e SGB II i. d. F. ab 01.01.2009 bis 31.03.2012), „Förderung von Arbeitsverhältnissen“ (§ 16e SGB II i. d. F. ab 01.04.2012 bis 31.12.2018) und nach dem Modellprojekt „Bürgerarbeit“ (Beschäftigungsphase, bis 31.12.2014) sowie nach dem Bundesprogramm „Soziale Teilhabe am Arbeitsmarkt“ (bis 31.12.2018).

[105] Die Wahl fällt hier auf den Median, da er deutlich robuster gegen Ausreißer ist als das arithmetische Mittel.

6.5.3 Bewertung[106]

Abbildung 27: Eintritte in ögB – Methodenvergleich Fehlerverteilungen

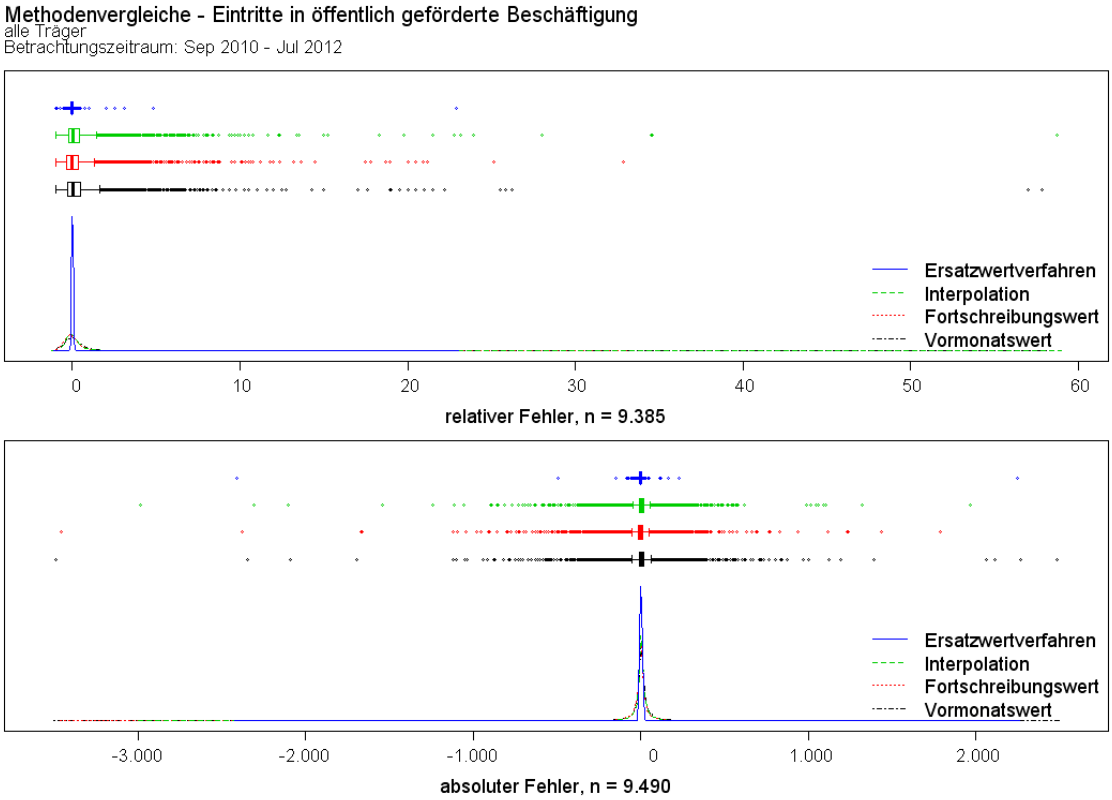
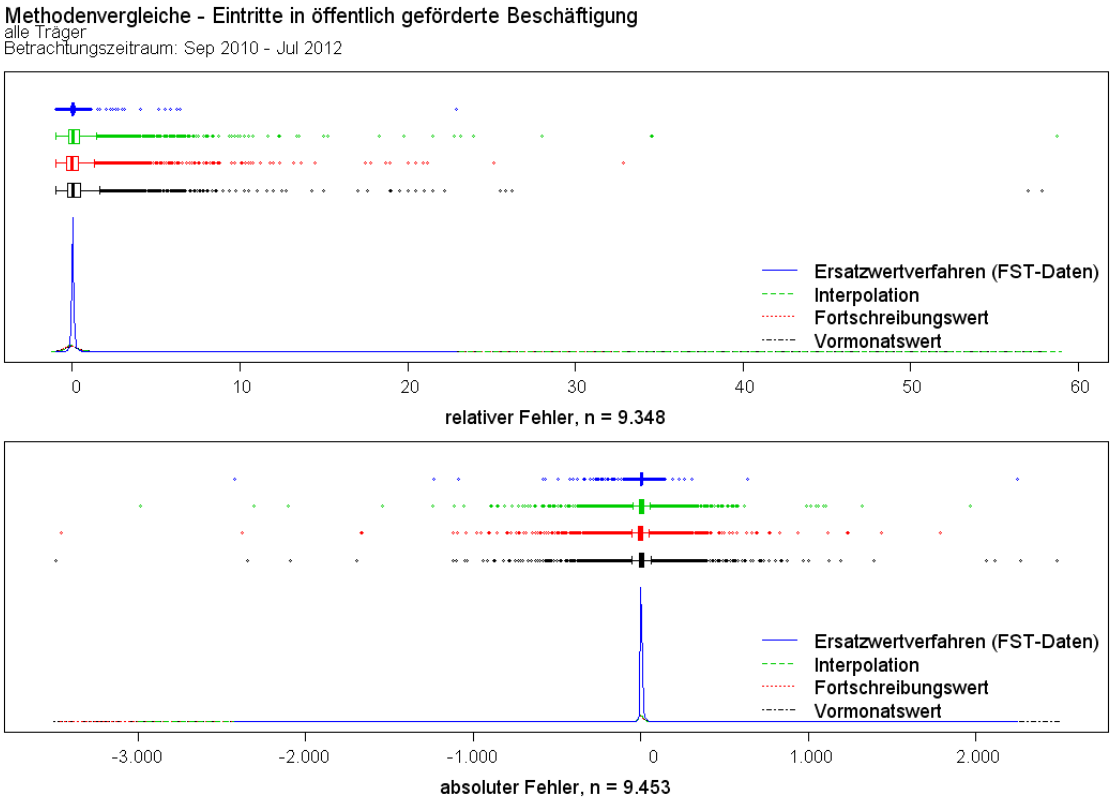


Abbildung 28: Eintritte in ögB – Schätzung auf Basis der Daten der FST – Methodenvergleich Fehlerverteilungen



Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von August 2010 bis August 2012.[107] Für die quantitative Bewertung wird das vorgeschlagene Imputationsverfahren mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Abbildung 27 zeigt, dass das vorgeschlagene Imputationsverfahren den anderen Modellen deutlich überlegen ist: die Verteilung der absoluten und relativen Fehler weist hier die stärkste Konzentration bei null auf. Die deutliche Überlegenheit gegenüber den anderen Verfahren rührt wohl daher, dass die Simulation überwiegend auf die Daten mit einer Wartezeit von zwei Monaten zurückgreift. Deswegen wurde auch ein Testlauf nur auf Basis von Ersatzwerten aus der FST durchgeführt – quasi als Szenario des Schlechtfalls. Abbildung 28 zeigt, dass das Modell, selbst wenn nur Ersatzwerte aus der FST herangezogen würden, mit einem Median der relativen Abweichung von 1,6 % immer noch deutlich bessere Ergebnisse erzielt würden als bei den Standardvergleichsmodellen.

Bei der Anwendung des Schätzverfahrens wird die tatsächliche Anzahl an Eintritten in ögB im Median exakt geschätzt. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler, definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots) ist hier nicht angebracht, da der Interquartilsabstand null beträgt. Somit fallen Werte, die ebenfalls keine Abweichung zum tatsächlichen Wert aufzeigen, unter die Ausreißer. Insgesamt beträgt für 5.992 der 9.904 Fälle (60,5 %) die absolute Abweichung null. Im Vergleich zu den Standardvergleichsmodellen zeigt sich die Überlegenheit des hier gewählten Verfahrens sehr deutlich in Abbildung 27.

Der Rückgriff auf Daten aus der FST und die damit verbundene kleine Abwandlung des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens für die Schätzung der Datenimputationswerte nach Geschlecht bietet bessere Ergebnisse als das sonst übliche Anteilswertverfahren. Dennoch: Die Schätzqualität nach Geschlecht ist bei der Anzahl ögB von mäßiger Qualität. Die geschlechtsspezifischen Anteilswerte variieren von September 2010 bis August 2012 relativ stark, was insbesondere auf die kleinen Fallzahlen zurückzuführen ist. Die mittlere Variabilitätsspanne der Anteilswerte für verkürzte Wartezeitstände und die Daten der FST innerhalb eines Berichtsmonats erstreckt sich von fünf (WZ2) bis 30 Prozentpunkte (WZ0). Die mittlere Variabilitätsspanne über die Differenzen der Anteilswerte zum Vormonat liegt sogar bei fast 70 Prozentpunkten. In etwa der Hälfte der vorliegenden Datenausfälle wird auf WZ2-Daten mit einer noch guten Variabilitätsspanne unter 10 % herangezogen. In allen anderen Fällen muss auf Anteilsinformation mit relativ geringer Güte zurückgegriffen werden.

6.5.4 Fazit, Ausblick

Das Schätzmodell liefert sehr gute Ergebnisse, insbesondere, wenn auf die Daten mit zweimonatiger Wartezeit zurückgegriffen werden kann, dies erfolgt in etwa der Hälfte der aktuellen Datenausfälle. Insgesamt ist die Schätzgüte in der mittleren Tendenz auf einem sehr guten Niveau, allerdings mit einigen deutlichen

[107] Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von September 2010 bis Juli 2012.

Ausreißern in Einzelfällen. Angesichts der häufig kleinen Fallzahlen und der starken Datensprünge ist dies erwartbar. Dementsprechend wird die Schätzgüte als gut bewertet.

6.6 Integrationen in sozialversicherungspflichtige Beschäftigung

6.6.1 Analyse fehlender Werte

Die Zahl der Integrationen in sozialversicherungspflichtige Beschäftigung (svB) fließt in den Nenner der Ergänzungsgröße K2E3 und ist eine Teilgröße der Integrationen insgesamt (vgl. Abschnitt 6.1), bei denen neben den sozialversicherungspflichtigen auch selbstständige Tätigkeiten und Ausbildungen gezählt werden. Bei Integrationen in svB liegen dieselben vereinzelten Datenlücken und derselbe sehr geringe Ausfallanteil von 0,33 % vor wie bei Integrationen insgesamt (vgl. Abschnitt 6.1.1).

6.6.2 Schätzmodell

Analysen zeigen, dass die Teilgröße der Integrationen in svB von saisonalen Effekten geprägt ist, die sich regionalspezifisch erheblich unterscheiden können. So schlagen beispielsweise Saisoneffekte wie die Belebung im Frühjahr oder in den Wintermonaten durch. Auf regionalen Arbeitsmärkten ergeben sich aber durchaus unterschiedliche Effekte, die sich wiederum in wirtschaftsstrukturell ähnlichen Gebieten auch in ähnlichen Mustern niederschlagen.

Bei den Analysen erwies sich ein Anteilswertverfahren, das auf den Schätzwerten der Gesamtzahl der Integrationen beruht, als überlegen.

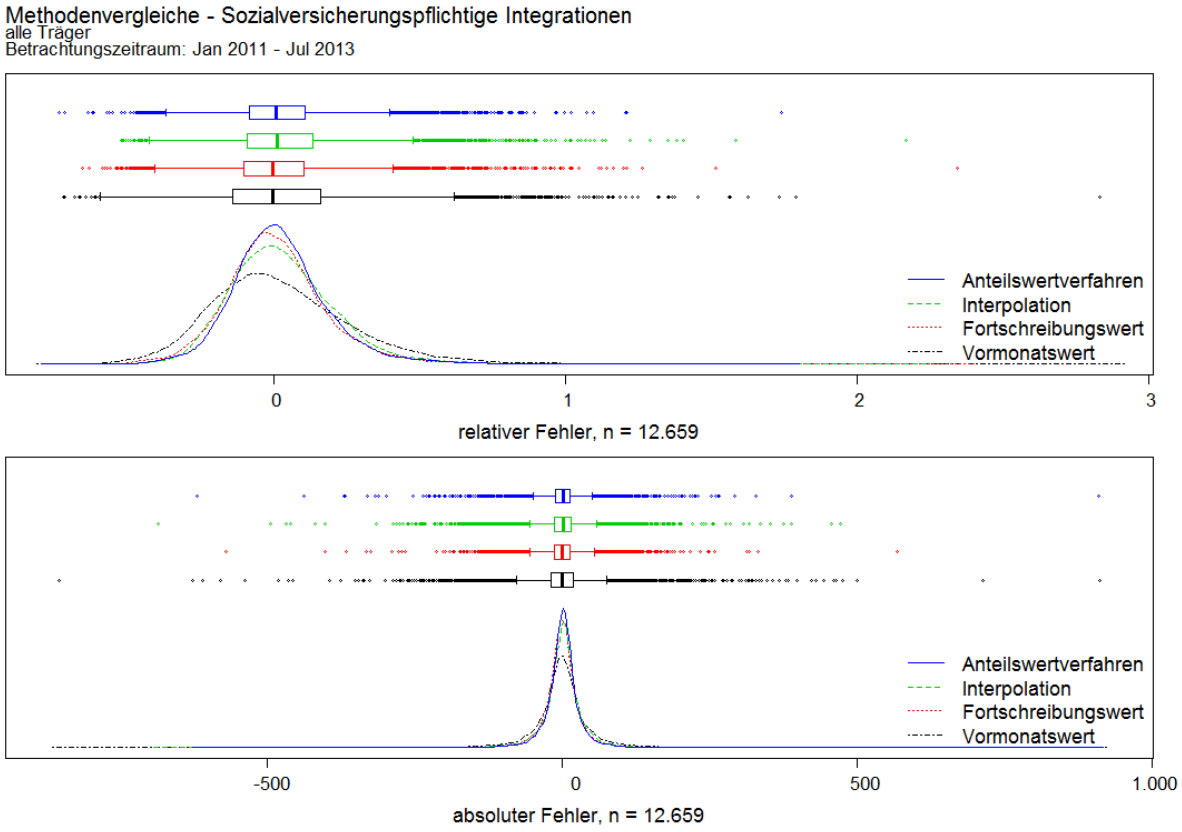
Der Schätzwert bei diesem Anteilswertverfahren ergibt sich als Produkt des Schätzwertes der Gesamtzahl der Integrationen (vgl. zweistufiges Modell in Abschnitt 6.1.2) und eines geschätzten Anteils von sozialversicherungspflichtigen an der Gesamtzahl der Integrationen. Analog zu dem Anteilswertverfahren bei der Geschlechterschätzung (vgl. Abschnitt 2.3) werden zur Anteilsschätzung auch hier die kürzeren Wartezeitstände in der gleichen Reihenfolge angewandt. Da sich das Verhältnis von sozialversicherungspflichtigen zur Gesamtzahl der Integrationen in den Ausbildungsmonaten deutlich ändert, ist es hier nicht sinnvoll im vierten Schritt auf den Vormonat auszuweichen, sondern den Vorjahresmonat hinzuzuziehen. Damit ergibt sich bei der Anteilsschätzung folgende Reihenfolge:

1. WZ2-Daten, falls diese auch fehlen
2. WZ1-Daten, falls diese auch fehlen
3. WZ0-Daten, falls diese auch fehlen
4. Vorjahresmonatsdaten mit WZ3, falls diese auch fehlen
5. Vormonatsdaten mit WZ3.

Die Differenzierung der Datenimputationswerte nach Geschlecht erfolgt anhand des in Abschnitt 2.3 beschriebenen Anteilswertverfahrens.

6.6.3 Bewertung[108]

Abbildung 29: Integrationen in svB – Methodenvergleich Fehlerverteilungen



Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von September 2010 bis August 2013. Für die quantitative Bewertung wird das vorgeschlagene Imputationsverfahren mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Neben den Standardvergleichsmodellen wurde – analog zum Schätzverfahren bei Integrationen (vgl. Abschnitt 6.1.2) – auch das k-nächste-Nachbarn-Fortschreibungsverfahren (knN-Fortschreibung) beim Vergleich berücksichtigt.[109] Im Ergebnis ist das vorgeschlagene Imputationsverfahren den anderen Modellen überlegen: die Verteilung der absoluten und relativen Fehler weist hier die stärkste Konzentration bei null auf. Die Überlegenheit gegenüber den anderen Verfahren rührt sicher daher, dass zum einen die Schätzung der Gesamtzahl der Integrationen eine hohe Güte besitzt und zum anderen die Anteilswerte hauptsächlich auf Daten mit einer Wartezeit von zwei Monaten beruhen.

[108] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[109] Da das knN-Verfahren aus konzeptuellen Gründen einen 4-monatigen Vorlauf benötigt und die verglichenen Verfahren einen 1-monatigen Nachlauf benötigen, enthält die Abbildung zu den Fehlerverteilungen nur Werte von Januar 2011 bis Juli 2013. Für eine ausführliche Erläuterung des k-nächste Nachbarn-Verfahrens am Beispielthema Integrationen siehe Anhang 0.

Bei der Anwendung des Anteilswertverfahrens wird die tatsächliche Anzahl an sozialversicherungspflichtigen Integrationen im Median und Betrag um 13 sozialversicherungspflichtige Integrationen (9,6 %) verfehlt, wobei das Bild von JC zu JC variiert. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/-40 %. Dies spricht für einen merklichen Schätzfehler im Allgemeinen. Ausreißer der absoluten Fehler treten häufiger bei großen JC, Ausreißer der relativen Fehler eher bei kleinen JC auf. Betrachtet man Daten, die sowohl bei absoluten als auch prozentualen Fehlern als Ausreißer klassifiziert werden, sind bei dem Anteilswertverfahren 0,7 % der geschätzten Datenpunkte auffällig. Simuliert man testhalber längere Datenlücken, so tendiert auch diese Methode (analog zu dem zweistufigen Modell in Abschnitt 6.1.2) zu höheren Schätzfehlern, sodass die sozialversicherungspflichtigen Integrationen stärker unter- bzw. überschätzt würden.

Es ist davon auszugehen, dass die Schätzung der Datenimputationswerte nach Geschlecht nach der Anteilswertmethode bei der Größe der sozialversicherungspflichtigen Integrationen von mindestens befriedigender Qualität ist. Die mittlere Variabilitätsspanne über alle JC in der Definition nach Abschnitt 2.3 beträgt zwischen August 2009 und August 2013 in diesem Zeitraum zwischen den Berichtsmonaten mit WZ3 rund 29 Prozentpunkte, zwischen den verschiedenen Wartezeitständen eines Berichtsmonats zwischen fünf und 17 Prozentpunkten. Bis auf 17 Datenpunkte werden bei den vorliegenden Datenausfällen WZ2-Daten mit einer noch guten Variabilitätsspanne unter 10 % herangezogen. Lediglich bei elf Datenpunkten wird auf die Anteile aus den Vor(jahres)monaten ausgewichen. Deren Anteilsinformation hat mit einer rund 30 %igen Variabilitätsspanne eine geringe Güte.

6.6.4 Fazit, Ausblick

Bei der Schätzung der sozialversicherungspflichtigen Integrationen produziert das vorgeschlagene Datenimputationsverfahren im Median einen betragsmäßigen Schätzfehler von 13 Fällen bzw. 9,6 %. Das Modell bietet mit den vorliegenden Mitteln das bestmögliche Verfahren für die Schätzung fehlender Daten. Auf Basis dieses Schätzkonzepts und der bisherigen Datenlückenmuster ist von einer befriedigenden Schätzgüte auszugehen.

6.7 Kontinuierliche Beschäftigung nach Integration in sozialversicherungspflichtige Beschäftigung

6.7.1 Analyse fehlender Werte

Ausgangspunkt für die Messung kontinuierlicher Beschäftigung nach Integration (KBNI) bilden die Integrationen in sozialversicherungspflichtige Beschäftigung (svB, vgl. Abschnitt 6.6). Eine Beschäftigung nach Integration gilt als kontinuierlich, wenn die betreffende Person am Stichtag von jedem der sechs auf die Integration folgenden Monate sozialversicherungspflichtig beschäftigt ist. Der Abgleich erfolgt über die Daten der Beschäftigungsstatistik (BST) der Bundesagentur für Arbeit, die auf den Meldungen der Arbeitgeber zur Sozialversicherung basiert. Datenlücken treten auf, wenn die Daten zu den Integrationen in svB

im Vorjahresmonat ausgefallen sind oder keine Informationen aus der BST vorliegen. Bislang rühren alle aufgetreten Datenlücken aus Datenausfällen bei den Integrationen in svB. Diese bilden wiederum eine Teilgröße der Integrationen insgesamt (vgl. Abschnitt 6.1). Somit liegen über alle genannten Größen dieselben Datenlücken, derselbe sehr geringe Ausfallanteil und dieselben Hintergründe für Datenausfälle vor (vgl. Abschnitt 6.1.1 bzw. Abbildung 19). Bei den Ergebnissen zur KBNI werden die Datenlücken erst ein Jahr später relevant, zum einen wegen des sechsmonatigen Verbleibszeitraums in dem die Kontinuität der Beschäftigung geprüft wird, außerdem kommt die sechsmonatige Wartezeit der BST hinzu. Im Zeitraum von Januar 2011 bis einschließlich Dezember 2018 gibt es 113 Datenlücken. Dies bedeutet einen sehr geringen Ausfallanteil von 0,29 %.[110]

6.7.2 Schätzmodell

Die Teilgröße der KBNI ist von ähnlichen saisonalen Effekten geprägt wie Integrationen in svB (Belebung im Frühjahr, rückläufig in den Wintermonaten). Beim Anteil der KBNI an Integrationen in svB treten andere Muster auf: Er ist in der ersten Jahreshälfte hoch und sinkt danach ab. Solche saisonalen Muster können sich jedoch regional erheblich unterscheiden (z. B. wirtschaftsstrukturell oder witterungsbedingt).

Unter allen geprüften Verfahren[111] erwies sich ein regressionsbasiertes Schätzverfahren (Methode der kleinsten Quadrate) mit – je nach Verfügbarkeit – zwei oder drei Prädiktoren als überlegen. In bestimmten Fällen erfolgt außerdem eine Saisonkorrektur.

Die abhängige Variable der Regressionen ist der Anteil der KBNI an allen Integrationen in svB. Die Anzahl der KBNI ergibt sich aus dem geschätzten Anteilswert multipliziert mit der Anzahl der Integrationen in svB. Im Schätzfall ist die Anzahl der Integrationen in svB bereits selbst ein Schätzwert (vgl. Abschnitt 6.6.2), der wiederum auf den geschätzten Integrationen beruht (vgl. Abschnitt 6.1.2). Zweifelsohne können sich hier Schritt für Schritt Schätzunsicherheiten kumulieren. Analysen haben jedoch gezeigt, dass eine Schätzung der Anteilswerte zu besseren Ergebnissen führt als eine direkte Schätzung der absoluten Anzahl der KBNI. Außerdem hat das Verfahren, die Anteilswerte zu schätzen, den Vorteil, dass die KBNI nie höher sein können als die Zahl der Integrationen in svB.

Das Grundmodell enthält zwei Prädiktoren. Der erste ist der durchschnittliche Anteil der KBNI an allen Integrationen in svB in den zwölf Monaten vor dem Datenausfall im jeweiligen Jobcenter. Diese Anteilswertinformation aus der Vergangenheit dient als Niveauvariable. Der zweite Prädiktor ist der SGB-II-Vergleichstyp. Dieser trägt als Trend- und Saisoninformation für die Vergleichstypen zur Modellverbesserung bei.

Das Grundmodell wird um einen dritten Prädiktor ergänzt, wenn die entsprechenden Daten dazu verfügbar sind. Hierbei handelt es sich um den Anteil der kontinuierlichen Beschäftigungen nach Übergängen von ELB in svB, nachfolgend abgekürzt mit KBNÜ. Die KBNÜ sind stark mit KBNI korreliert. Somit ergänzt der

[110] Die in diesem Kapitel berichteten Ergebnisse umfassen einen längeren Zeitraum als in anderen Kapiteln der Methodenübersicht, da das Imputationsverfahren für KBNI erst mit Version 5.2 dieser Methodenübersicht (Stand Mai 2020) entwickelt wurde, also deutlich nach den drei ersten Entwicklungsstufen (bis Mai 2014).

[111] Unter anderem wurden einfache Anteilswertverfahren, Fortschreibungen, Zeitreihen-Modelle sowie Regressionsmodelle mit anderen Prädiktoren getestet.

dritte Prädiktor das Modell um JC-spezifische Informationen, insbesondere zu Saisonmustern, aber auch zu atypischen Verläufen.

Übergänge von ELB in svB werden durch einen Abgleich der ELB-Bestände mit Daten der BST ermittelt. Ein Übergang von „nicht svB“ in „svB“ wird gemessen, wenn eine Person (a) im Vormonat *t-1* im Bestand der ELB und zugleich nicht im Bestand der svB war und (b) im Berichtsmonat *t* im Bestand der svB ist (vgl. blaue Einträge in u. s. Abbildung 30, zweite Zeile).[112] Im Sinne einer möglichst guten Annäherung an die Messung von Integrationen in svB werden Übergänge in Ausbildung ausgeschlossen. Die wesentlichen Unterschiede zwischen Integrationen in svB und Übergängen in svB sind:

1. Übergänge basieren auf den Arbeitgebermeldungen zur Sozialversicherung, Integrationen auf operativen Daten bzw. Meldungen der Jobcenter.
2. Für die Messung eines Übergangs darf ein ELB am Stichtag des Vormonats (*t-1*) nicht svB gewesen sein, am Stichtag des Berichtsmonats *t* muss eine svB bestanden haben. Für die Integrationsmessung genügt es, wenn irgendwann im Zeitraum zwischen *t-1* und *t* eine neue Beschäftigung aufgenommen wurde, es ist hier irrelevant, ob am Stichtag *t-1* bereits eine (andere) svB bestand und ob die neue svB am Stichtag t immer noch besteht.
3. Bei den Übergängen sind auch sozialversicherungspflichtige öffentlich geförderte Beschäftigungen sowie Freiwilligendienste enthalten, als Integrationen zählen diese Beschäftigungsformen nicht.

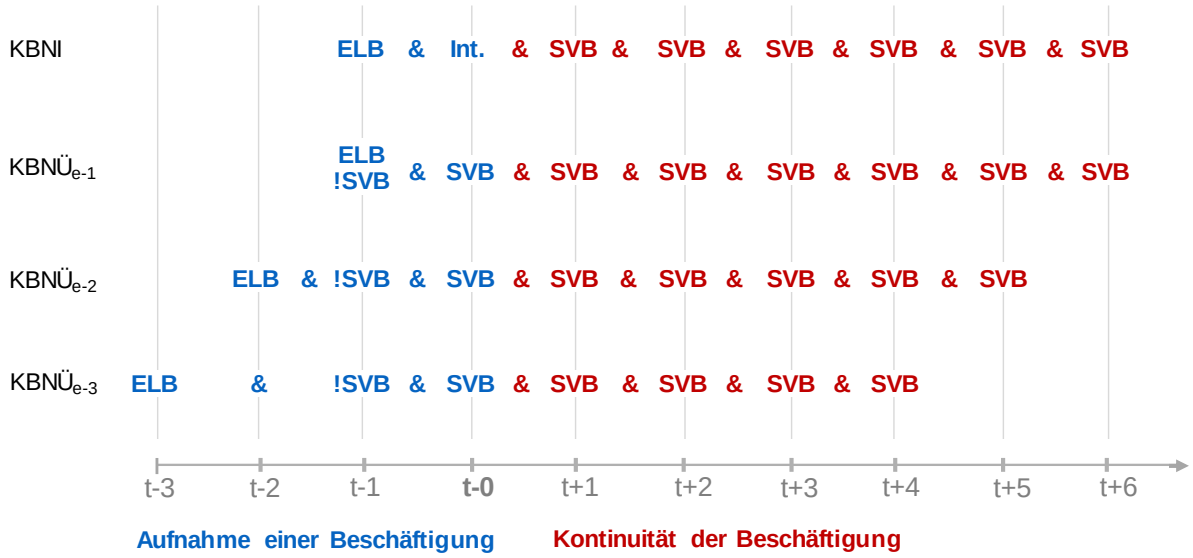
Die Kontinuität der Beschäftigung nach einem Übergang wird genauso ermittelt wie bei den Integrationen: Es wird geprüft, ob an allen Stichtagen der sechs Monate nach dem Übergang eine svB besteht (vgl. rote Einträge in u. s. Abbildung 30, zweite Zeile).

Da es bislang keine Datenausfälle bei Beschäftigungsdaten gibt, können Übergänge immer ermittelt werden, wenn die Ausgangsmenge der ELB verfügbar ist. Um bei möglichst vielen Datenausfällen die Schätzergebnisse durch Übergangsinformationen verbessern zu können, wird bei einem Ausfall der ELB-Bestände im Vormonat auf Bestandsdaten aus weiter zurückliegenden Monaten zurückgegriffen. Die Zeitpunkte für die Ermittlung von Übergängen von „nicht svB“ in „svB“ bleiben bestehen, die Überprüfung der Kontinuität der Beschäftigung verkürzt sich allerdings aufgrund von Beschränkungen im Auswertesystem um jeweils einen Monat zur Vorstufe. Dieses Vorgehen ist möglich, weil mit den Anteilswerten in der Regression gerechnet wird. Es hat sich gezeigt, dass die Schätzergebnisse im Vergleich zum Grundmodell auch dann noch deutlich besser sind, wenn als Ausgangsmenge für die Ermittlung der Übergänge die ELB-Bestände drei Monate vor dem Datenausfall herangezogen werden und sich der Zeitraum für die Messung des Verbleibs auf vier Monate verkürzt. Die drei Varianten der KBNÜ mit unterschiedlichen Zeitpunkten der Ermittlung des ELB-Ausgangsbestandes werden als $KBN\ddot{U}_{e-1}$, $KBN\ddot{U}_{e-2}$ und $KBN\ddot{U}_{e-3}$ bezeichnet (vgl. u. s. Abbildung 30, dritte und vierte Zeile).

Abbildung 30 zeigt die Messung von KBNI und den drei Stufen der KBNÜ-Messung im Vergleich.

[112] Eine ausführliche Beschreibung findet sich in: Wyrwoll et al. (2014): Messung von Beschäftigungsaufnahmen – Vergleichende Darstellung der Messkonzepte. Methodenbericht. Nürnberg: Statistik der BA, zum Download verfügbar unter: https://statistik.arbeitsagentur.de/ > Grundlagen > Methodik und Qualität > Methodenberichte und Hintergrundinfos > Grundsicherung für Arbeitsuchende (SGB II).

Abbildung 30: KBNI bzw. KBNÜ (Stufen 1 bis 3) – Überblick über die Messmodelle



KBNI	Kontinuität der Beschäftigung nach Integration
$KBNÜ_{e-x}$	Kontinuität der Beschäftigung nach Übergang für den ELB-Bestand *x* Monate vor dem Übergangsmonat
Int	Integration zum Zeitpunkt t_i
ELB	zum Zeitpunkt t_i im Bestand ELB
SVB	zum Zeitpunkt t_i im Bestand svB
!SVB	zum Zeitpunkt t_i nicht im Bestand svB

Bei 9 von 10 Datenausfällen können die Schätzungen durch KBNÜ-Informationen verbessert werden.

Liegen für mehr als drei Monate hintereinander keine Daten zum ELB-Bestand vor, werden keine KBNÜ-Informationen in das Regressionsmodell aufgenommen. In diesen Fällen wird das Grundmodell mit den zwei Prädiktoren gleitender Jahresdurchschnitt des Anteils der KBNI und SGB-II-Vergleichstyp verwendet. Da sich in Simulationen gezeigt hat, dass das Grundmodell für saisonal geprägte Zeitreihen zu vereinfachend sein kann, erfolgt bei Vorliegen von jobcenterspezifischer Saisonalität eine Nachkorrektur. Die Nachkorrektur soll saisonale Muster abbilden, wenn diese nicht durch den SGB-II-Vergleichstyp ausreichend gut berücksichtigt wurden. Um zu testen, ob eine Nachkorrektur notwendig ist, werden die Schätzwerte anhand des Grundmodells für das Jobcenter *i* über den Zeitraum der vergangenen 60 Monate simuliert. Danach werden die Residuen, also die Differenzen von Schätzwerten und Ist-Werten, berechnet und innerhalb eines linearen Regressionsmodells auf Saisonalität überprüft. Liegt Saisonalität der Residuen vor, wird der Schätzwert aus dem Grundmodell um den Mittelwert der Residuen der fünf Vorjahresmonate vor dem Datenausfall korrigiert.

Das Schätzverfahren für den Anteil der KBNI lässt sich so zusammenfassen: Liegt ein Datenausfall vor, wird die folgende Kaskade durchlaufen und der erste mögliche Schätzwert zur Imputation genutzt:

1. Grundmodell erweitert um den Prädiktor $KBNÜ_{e-1}$
2. Grundmodell erweitert um den Prädiktor $KBNÜ_{e-2}$
3. Grundmodell erweitert um den Prädiktor $KBNÜ_{e-3}$
4. Grundmodell, ggf. mit Saisonkorrektur

Formalisiert lautet das Schätzmodell:

$$y_{t,i} = \underbrace{\alpha_t \cdot \bar{y}_{t,i} + \sum_{j=1}^{15} \beta_{t,j} \cdot I(i,j)}_{Grundmodell} + \underbrace{\gamma_t \cdot z_{t,i}}_{KBN\ddot{U}-Term} + \varepsilon_{t,i}, \quad i = 1, \ldots, n \tag{14}$$

mit:

KBNI	kontinuierliche Beschäftigungen nach Integration in svB
KBNÜ	kontinuierliche Beschäftigungen nach Übergängen in svB
$y_{t,i}$	Anteil der *KBNI* an allen Integrationen in svB des Jobcenters *i* zum Zeitpunkt *t*
t	Zeitpunkt *t*, zu dem imputiert wird
i	JC *i*, für das imputiert wird
α_t	Regressionskoeffizient zum gleitenden Jahreswert $\bar{y}_{t,i}$ im Monat *t*
$\bar{y}_{t,i}$	gleitender Jahreswert des Anteils der *KBNI* an Integrationen in svB des Jobcenters *i* zum Zeitpunkt *t-1* $\left(\bar{y}_{t,i} = \frac{1}{12} \cdot \sum_{k=1}^{12} y_{t-k,i}\right)$
$\beta_{t,1}$ bis $\beta_{t,15}$	Regressionskoeffizienten für die Indikatorfunktionen *I* zum SGB-II-Vergleichstyp im Monat *t*
$I(i,j)$	Indikatorfunktion die den Wert 1 annimmt, wenn Jobcenter *i* im SGB-II-Vergleichstyp *j* liegt und ansonsten 0 ist
j	Index für den SGB-II-Vergleichstyp
γ_t	Regressionskoeffizient zum Anteil der $KBN\ddot{U}_{z_{t,i}}$ für Monat *t*
$z_{t,i}$	Anteil der *KBNÜ* an allen Übergängen in svB des Jobcenters *i* zum Zeitpunkt *t* (für eine der Stufen *e-1*, *e*-2 oder *e-3*)
$\varepsilon_{t,i}$	Fehler für Jobcenter i zum Zeitpunkt *t*
n	Anzahl an Jobcentern

Der KBNÜ-Term entfällt, wenn in den letzten drei Monaten vor dem Datenausfall keine plausiblen Daten zu ELB-Beständen vorliegen.

Die Imputation erfolgt für jeden Monat separat und in zeitlich aufsteigender Reihenfolge. Die Berichterstattung zu Integrationen beginnt im Januar 2011. Bis einschließlich Dezember 2011 liegt bei Datenausfällen noch kein vollständiger Zwölfmonatszeitraum von Vormonatswerten vor. Um Datenlücken im Jahr 2011 dennoch füllen zu können, werden die gleitenden Jahresdurchschnitte mit in der Zukunft liegenden Datenpunkten ermittelt.

Für die Differenzierung der Datenimputationswerte nach Geschlecht wird vom üblichen Verfahren bei der Geschlechterschätzung (vgl. Abschnitt 2.3) abgewichen, weil für die KBNI keine Daten mit verkürzter Wartezeit vorliegen. Stattdessen werden für die Ermittlung der Anteilswerte der Frauen Informationen zu den KBNÜ herangezogen. Es wird folgende Kaskade durchlaufen:

1. $KBN\ddot{U}_{e-1}$, falls diese auch fehlen
2. $KBN\ddot{U}_{e-2}$, falls diese auch fehlen
3. $KBN\ddot{U}_{e-3}$, falls diese auch fehlen
4. $KBNI$ des Vormonats.

Die Anzahl der KBNI der Männer ergibt sich als Differenz zwischen dem Gesamtwert der KBNI und der Anzahl der KBNI der Frauen.

6.7.3 Bewertung

Abbildung 31: KBNI – Methodenvergleich Fehlerverteilungen

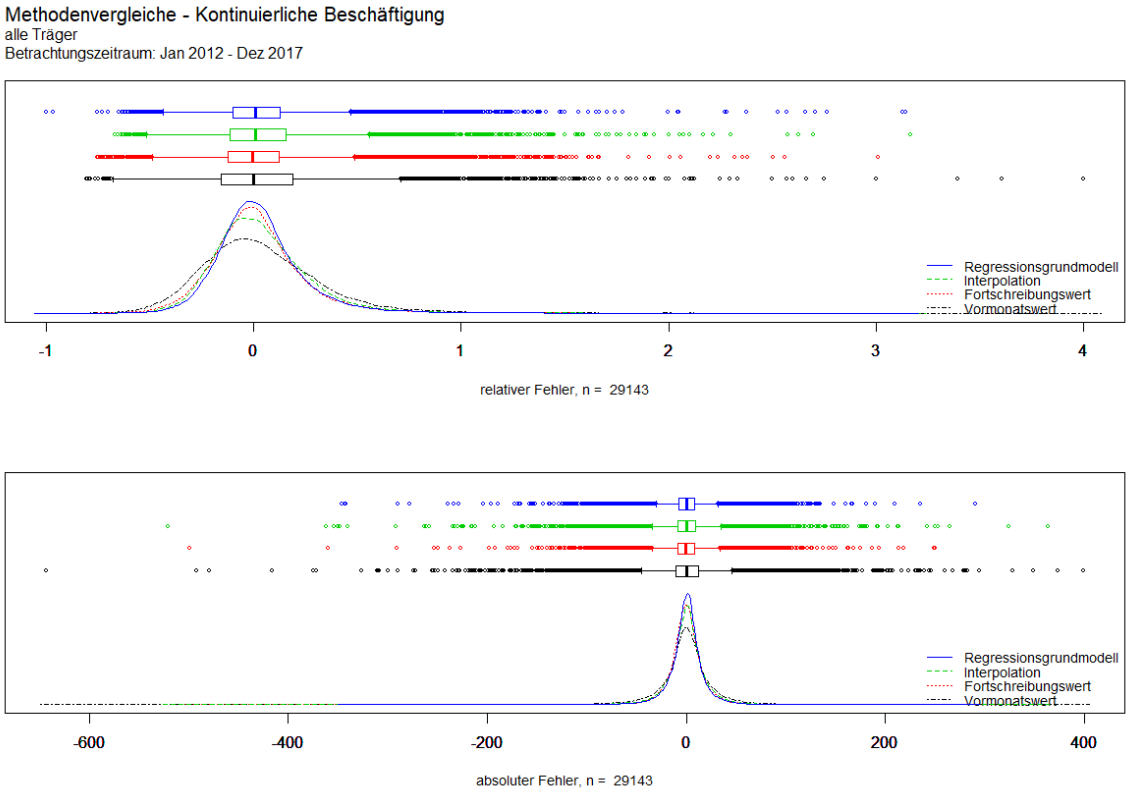
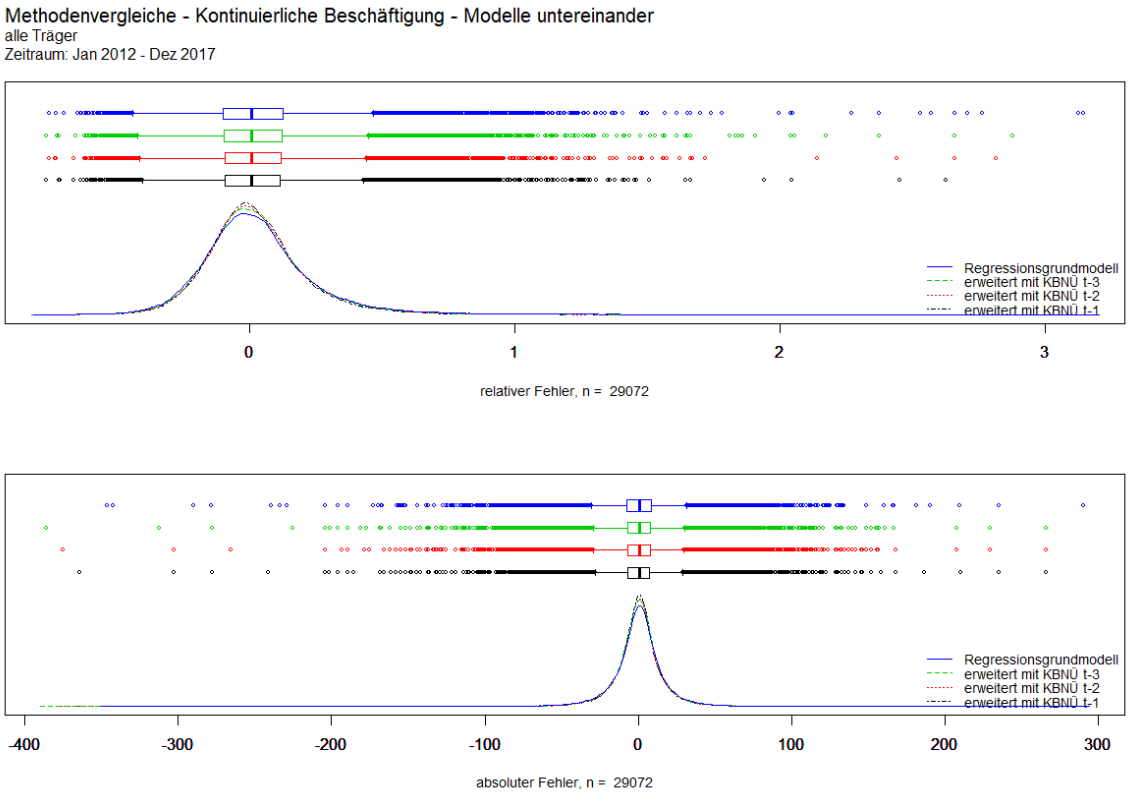


Abbildung 32: KBNI – Methodenvergleich Fehlerverteilungen der Kaskaden



Für die quantitative Bewertung des Verfahrens werden absolute und relative Abweichungen zwischen Schätz- und Ursprungswerten herangezogen. Als Datengrundlage dienen die Daten von Januar 2011 bis Januar 2018.[113] Das Grundmodell ist von den vier Modellen der oben beschriebenen Kaskade dasjenige mit der schlechtesten Bewertung. Für die quantitative Bewertung wird daher das Grundmodell mit den Standardvergleichsmodellen aus Abschnitt 2.1.3 verglichen. Im Ergebnis ist das Grundmodell, insbesondere beim absoluten Fehler, den anderen Standardvergleichsmodellen überlegen.

Bei der Anwendung des vorgeschlagenen Datenimputationsverfahrens wird die tatsächliche Anzahl an kontinuierlicher Beschäftigung nach Integration im Median und Betrag um acht Integrationen (11,2 %) verfehlt, wobei das Bild von JC zu JC variiert. Eine ergänzende Ausreißeranalyse der absoluten und prozentualen Fehler definiert Ausreißer nach der Boxplotmethode als Punkte, die von den Quartilen der Fehlerverteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind (vgl. Anhang 8.1, Lesehilfe zu Boxplots). Die Whisker der Boxplots (jeweiliges Quartil +/- 1,5-facher Interquartilsabstand) der relativen Abweichungen enden beim vorgeschlagenen Verfahren bei ca. +/-45 %. Dies spricht für einen merklichen Schätzfehler im Allgemeinen. Betrachtet man Daten, die sowohl bei absoluten als auch prozentualen Fehlern als Ausreißer klassifiziert werden, sind bei dem vorgeschlagenen Imputationsverfahren lediglich 0,5 % der geschätzten Datenpunkte auffällig. Da von den bislang 113 Datenausfällen nur neun Fälle (8 %) mit dem Grundmodell ermittelt wurden und für alle anderen 104 Fälle (92 %) Übergangsinformationen für die Schätzungen herangezogen wurden, sind die Fehler größtenteils geringer (vgl. u. s. Abbildung 32 gegenüber Abbildung 31). In keinem dieser neun Fälle, die mit dem Grundmodell geschätzt wurden, gab es eine signifikante Saisonalität der Residuen, so dass die Nachkorrektur bislang nicht zum Einsatz kam.

Da die Differenzierung der Datenimputationswerte nach Geschlecht – abweichend vom üblichen Muster bei der Geschlechterschätzung (vgl. Abschnitt 2.3) – anhand der Geschlechterinformationen bei den KBNÜ erfolgt, ist hier zumindest vom gleichen merklichen Schätzfehler auszugehen – eventuell verstärkt durch geringe Fallzahlen.

6.7.4 Fazit, Ausblick

Das Schätzmodell ist grundsätzlich gut geeignet, fehlende Daten zur KBNI zu imputieren, insbesondere dann, wenn Informationen zur KBNÜ vorliegen. Da die Schätzwerte der KBNI jedoch unmittelbar von den Schätzwerten für die Integrationen in svB abhängen und diese wiederum von den Schätzwerten für die Integrationen insgesamt, können sich Schätzunsicherheiten kumulieren. Das vorgeschlagene Datenimputationsverfahren produziert im Median einen betragsmäßigen Schätzfehler von acht Fällen bzw. 11,2 %. Das ist ein wenig mehr als der Schätzfehler von 9,6 %, der für die Integrationen in svB ermittelt wurde (vgl. Abschnitt 6.6.3). Wie bei den Integrationen in svB ist bei den Schätzergebnissen für die KBNI von einer befriedigenden Schätzgüte auszugehen.

[113] Die Zeitreihe für die Ausgangsgröße Kontinuierliche Beschäftigung nach Integration beginnt aus messtechnischen und wartezeittechnischen Gründen erst im Dezember 2011 und endet bereits im Januar 2018. Da die verglichenen Verfahren aus konzeptuellen Gründen einen 1-monatigen Vor- und Nachlauf benötigen, enthalten die Abbildungen zu den Fehlerverteilungen nur Werte von Januar 2012 bis Dezember 2017.

7 Analysen zu Mehrfachausfällen

7.1 Einleitung

Während der Entwicklungsarbeiten an den Verfahren zur Datenimputation für Kennzahlen nach § 48a SGB II (Februar 2012 bis Mai 2014) wurden alle Imputationsverfahren standardmäßig einer quantitativen Bewertung unterzogen.[114] Diese erfolgte anhand statistischer Simulationsanalysen: die tatsächlich vorhandenen Ursprungswerte wurden einzeln als fehlend angenommen, Schätzwerte simuliert und die Abweichungen zwischen tatsächlichen Werten und simulierten Schätzwerten bestimmt. Je kleiner die Abweichungen sind, umso kleiner ist der Schätzfehler und umso besser das Imputationsverfahren. Hierbei wurden zunächst nur Einzelwerte als fehlend behandelt („leave-one-out", vgl. Abschnitt 2.1.2). Tatsächlich liegen aber manchmal, wenn auch sehr selten, Mehrfachausfälle vor, sodass sich die Datenlücken bei einzelnen JC häufen. Nachfolgend werden die Auswirkungen von Mehrfachausfällen auf den Schätzfehler dargestellt.

Es ist zwischen gereihten und nicht gereihten Mehrfachausfällen zu unterscheiden. Im ersten Fall treten die Ausfälle bei dem betreffenden JC in direkt aufeinander folgenden Berichtsmonaten auf. Dagegen liegen im zweiten Fall für das betreffende JC zwischen Datenausfällen Echtwerte vor. Im Bereich der Datenimputation für Kennzahlen nach § 48a SGB II werden überwiegend Fortschreibungsverfahren angewendet. Dabei werden fehlende Werte durch das Produkt des Vormonatswertes (Niveaufaktor) mit einem Fortschreibungsfaktor (Entwicklungsfaktor) imputiert. Hierbei ist anzunehmen, dass gereihte Mehrfachausfälle mit größeren Schätzunsicherheiten behaftet sind als nicht gereihte Mehrfachausfälle. Je länger ein gereihter Ausfall andauert, umso größer ist die Unsicherheit, dass das unterstellte Schätzniveau und/oder die unterstellte zeitliche Entwicklung noch zutreffend sind. Dagegen können bei nicht gereihten Mehrfachausfällen die Zeitreihen mit Hilfe der Informationen aus Echtwerten und zeitlichen Veränderungen zwischen den Datenausfällen immer wieder abgesichert werden. In Bezug auf Schätzfehler stellen gereihte Mehrfachausfälle also den schlechteren Fall dar. Im Sinne einer Maximalabschätzung konzentrieren sich die folgenden Analysen auf gereihte Mehrfachausfälle. Darüber hinaus fokussieren sich die Analysen auf die drei Kennzahlen K1 (Veränderung der Summe der Leistungen zum Lebensunterhalt), K2 (Integrationsquote) und K3 (Veränderung des Bestands an LZB) sowie die Ergänzungsgröße K1E2 (Veränderung der Zahl der ELB). Zum einen stehen diese vier Kenngrößen in der Berichterstattung am deutlichsten im Fokus. Zum anderen werden bei allen betreffenden Grunddaten Fortschreibungsverfahren zur Datenimputation verwendet. Es gibt bei den betreffenden Grunddaten bislang (April 2016) höchstens eine Häufung von bis zu fünf direkt aufeinander folgenden Datenlücken je JC Nachfolgend werden maximal sechs gereihte Mehrfachausfälle simuliert, um die Auswirkungen auf die Schätzunsicherheit zu analysieren.[115]

[114] Verfahrensbedingt wurde bei einigen Imputationsverfahren auf dieses Standardvorgehen verzichtet und geeignetere Maßnahmen zur Verfahrensbewertung ergriffen, vgl. Fußnote 10 in Abschnitt 2.1.2.

[115] Die Analyseergebnisse dieses Kapitels 7 (Analysen zu Mehrfachausfällen) beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.2).

7.2 Kennzahl K1 „Veränderung der Summe der Leistungen zum Lebensunterhalt“

7.2.1 Kennzahl K1, ihre Grunddaten und dazugehörigen Imputationsverfahren

Die Kennzahl K1 „Veränderung der Summe der Leistungen zum Lebensunterhalt“ misst die Veränderung der Hilfebedürftigkeit von BG, indem sie die Summe der Leistungen zum Lebensunterhalt (ohne Leistungen für Unterkunft und Heizung) der BG eines JC im Berichtsmonat (Zähler) ins Verhältnis zum entsprechenden Vorjahresmonatswert (Nenner) setzt.

Die Imputation der Datenlücken erfolgt durch ein regressionsbasiertes Fortschreibungsverfahren: fehlende Werte bei der Summe der Leistungen zum Lebensunterhalt (LLU) in einem Berichtsmonat werden durch das Produkt des Vormonatswertes der LLU mit einem Fortschreibungsfaktor imputiert. Der Fortschreibungsfaktor wird mit Hilfe einer linearen Regression ermittelt und ist als geschätzte prozentuale Veränderung der LLU zum Vormonat zu interpretieren (vgl. Abschnitt 3.1.2). Die Regressoren wurden nach sachlogischen Erwägungen ausgewählt und haben eine sehr gute Vorhersagekraft.

7.2.2 Simulation von Mehrfachausfällen

Die Simulation von Mehrfachausfällen beruht auf den Daten der Berichtsmonate Oktober 2010 bis März 2014. Für jedes JC wurden bis zu sechs aufeinanderfolgende Datenausfälle der LLU simuliert. Durch Zufallsauswahl wird der erste Ausfallmonat bestimmt und ein Imputationswert für LLU ermittelt. Im Falle von zwei aufeinanderfolgenden Datenausfällen setzt die Berechnung auf dem ersten Imputationswert auf. Dieser Vorgang wiederholt sich gegebenenfalls bis zu einer Ausfallsequenz von sechs aufeinanderfolgenden Datenausfällen. Mit dem erforderlichen sechsmonatigen zeitlichen Vorlauf begrenzen sich die Analyseergebnisse auf den Zeitraum zwischen März 2011 und März 2014. Um ein Zufallsergebnis auszuschließen, wurde die Simulation in fünf unabhängigen Simulationsläufen wiederholt. Die Ergebnisse konvergierten hierdurch ausreichend gut.

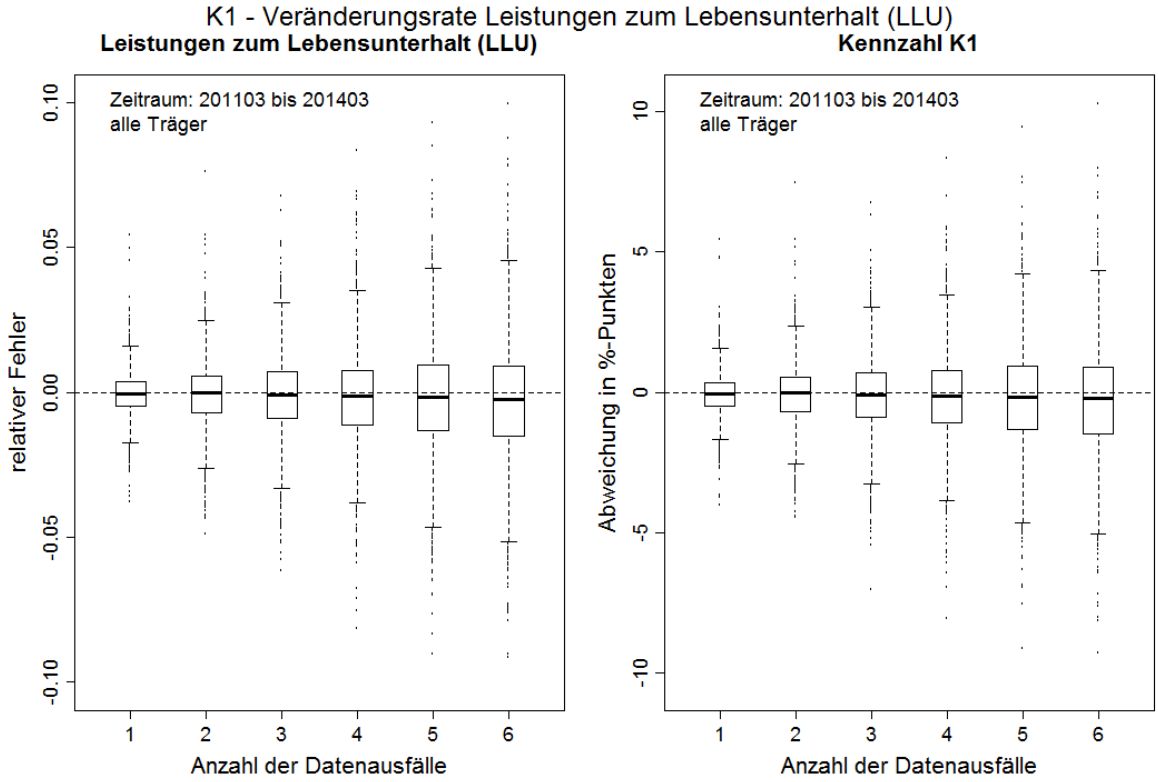
Grundsätzlich können Mehrfachausfälle sowohl im Zähler (LLU im Berichtsmonat) als auch im Nenner (LLU im Vorjahresmonat) auftreten. Es ist aber davon auszugehen, dass sich die Simulationsergebnisse für Ausfälle im Nenner ähnlich verhalten wie die Ergebnisse für Ausfälle im Zähler. Ergänzende Analysen zeigen außerdem, dass die Konstellation eines gleichzeitigen (Mehrfach-)Ausfalls in Zähler und Nenner keinen nennenswerten Einfluss auf die Schätzgüte haben (vgl. Abschnitt 7.6). Die nachfolgende Analyse konzentriert sich deshalb ausschließlich auf Ausfälle im Zähler. Die betreffende Kennzahl K1 berechnet sich anschließend aus den simulierten Schätzwerten im Zähler und den dazugehörigen nichtsimulierten Vorjahreswerten im Nenner. Im Ergebnis erhält man:

- den relativen Fehler für die Zählergröße LLU
- die daraus resultierende Abweichung zur Kennzahl K1 „Veränderung der Summe der LLU“ in Prozentpunkten.

7.2.3 Ergebnisse

Abbildung 33 zeigt, wie sich die Fehler-Boxplots (vgl. Anhang 8.1, Lesehilfe zu Boxplots) mit zunehmender Anzahl an gereihten Mehrfachausfällen (x-Achse) verbreitern, das heißt der Schätzfehler steigt mit Mehrfachausfällen sichtbar an.

Abbildung 33: Mehrfachausfälle: Relativer Fehler bei LLU und die daraus resultierende Abweichung zur Kennzahl K1 in Prozentpunkten



Relativer Fehler bei den LLU: Die Whisker der Boxplots (Verteilung ohne Ausreißer = jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1) liegen bei Einzelausfällen im Bereich von +/- 1,7 Prozent, bei sechs hintereinander folgenden Datenausfällen zwischen –5,2 und +4,5 Prozent. Der Fehlerbereich hat sich also bei sechs Datenausfällen beinahe verdreifacht. Im Extremfall liegt der relative Fehler bei LLU bei sechsmonatigen Ausfallsequenzen zwischen –9,1 und +10,0 Prozent. Dabei zeigt sich im Median mit zunehmender Anzahl von Ausfällen eine leichte Unterschätzung.

Abweichung zur Kennzahl K1: Der relative Fehler der Zählergröße überträgt sich erwartungsgemäß annähernd eins zu eins auf den Kennzahlenfehler. Die Whisker der Boxplots liegen bei Einzelausfällen im Bereich zwischen –1,7 und +1,6 Prozentpunkten, bei sechs hintereinander folgenden Datenausfällen zwischen –5,1 und +4,4 Prozentpunkten. Der Fehlerbereich hat sich also bei sechs Datenausfällen ebenfalls beinahe verdreifacht. Im Extremfall liegt die Abweichung zur Kennzahl K1 zwischen –9,3 und +10,3 Prozentpunkten. Dabei zeigt sich im Median mit zunehmender Anzahl von Ausfällen ebenfalls eine leichte Unterschätzung.

Die Kennzahl K1 umfasst im Betrachtungszeitraum Werte zwischen –25,3 und +15,3 Prozent. Vertiefte Analysen der Ausreißer zeigen, dass besonders große Schätzfehler auf entweder kleine JC oder vor allem auf untypische Ausgangswerte im zeitlichen Umfeld der Schätzung zurückzuführen sind. Bei den hier verwendeten Fortschreibungsverfahren setzen sich zum Beispiel Niveaufehler aus ungewöhnlich niedrigen oder hohen Vormonats-werten bei längeren Ausfallsequenzen monatelang fort.

7.2.4 Fazit, Ausblick

Die Analyse zeigt, dass Mehrfachausfälle erwartungsgemäß mit einer erhöhten Schätzunsicherheit einhergehen: Bei sechs aufeinanderfolgenden Datenausfällen sind die Kennzahl K1 und ihre dazugehörigen Grunddaten LLU von einem beinahe drei Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen und werden im Median leicht unterschätzt. Ausgehend von den bisher beobachteten Ausfallmustern (maximal fünf hintereinander folgende Ausfälle, die mit dem beschriebenen statistischen Schätzverfahren imputiert werden) ist jedoch von akzeptablen Unsicherheiten auszugehen. Ausreißer in der Fehlerverteilung sind in der Regel ungewöhnlich niedrigen oder hohen Ausgangswerten im zeitlichen Umfeld der Schätzung geschuldet. Insgesamt können die Schätzergebnisse als robust beurteilt werden.

7.3 Ergänzungsgröße K1E2 „Veränderung der Zahl der erwerbsfähigen Leistungsberechtigten“

7.3.1 Ergänzungsgröße K1E2, ihre Grunddaten und dazugehörigen Imputationsverfahren

Die Ergänzungsgröße K1E2 „Veränderung der Zahl der erwerbsfähigen Leistungsberechtigten" bietet ergänzende Informationen zum Ziel „Verringerung der Hilfebedürftigkeit“. Sie misst die Veränderung des Bestands von ELB eines JC zwischen dem Berichtsmonat (Zähler) und dem entsprechenden Vorjahresmonatswert (Nenner).

Die Imputation der Datenlücken erfolgt durch ein Fortschreibungsverfahren: fehlende Werte bei der Bestandsgröße der ELB in einem Berichtsmonat werden durch das Produkt des Vormonatswertes mit einem Fortschreibungsfaktor imputiert. Als Referenzgruppe dienen die JC des SGB-II-Vergleichstyps, dem das betreffende JC angehört. Der Fortschreibungsfaktor ist also die geschätzte mittlere Veränderungsrate zum Vormonat im gemeinsamen Vergleichstyp (vgl. Abschnitt 4.1.2). Veränderungsmuster in wirtschaftsstrukturell ähnlichen Gebieten werden hierdurch gut abgebildet. Bei der Simulation von Einzelausfällen wies das Verfahren eine hervorragende Schätzgüte auf.

7.3.2 Simulation von Mehrfachausfällen

Die Simulation von Mehrfachausfällen beruht auf den Daten der Berichtsmonate August 2009 bis März 2014. Mit dem erforderlichen sechsmonatigen zeitlichen Vorlauf begrenzen sich die Analyseergebnisse für ELB auf den Zeitraum zwischen Februar 2010 und März 2014, für die Ergänzungsgröße K1E2 zwischen September 2010 und März 2014. Aus den Schätzwerten der Simulation werden relative Fehler für ELB bzw. die K1E2 berechnet.

Grundsätzlich können Mehrfachausfälle sowohl im Zähler (ELB im Berichtsmonat) als auch im Nenner (ELB im Vorjahresmonat) auftreten. Es ist aber davon auszugehen, dass sich die Simulationsergebnisse für Ausfälle im Nenner ähnlich verhalten wie die Ergebnisse für Ausfälle im Zähler. Ergänzende Analysen zeigen außerdem, dass die Konstellation eines gleichzeitigen (Mehrfach-)Ausfalls in Zähler und Nenner keinen nennenswerten Einfluss auf die Schätzgüte haben (vgl. Abschnitt 7.6). Die nachfolgende Analyse konzentriert sich deshalb ausschließlich auf Ausfälle im Zähler. Die betreffende Ergänzungsgröße K1E2 berechnet sich also aus den simulierten Schätzwerten im Zähler und den dazugehörigen nichtsimulierten Vorjahreswerten im Nenner. Im Ergebnis erhält man:

- den relativen Fehler für die Zählergröße ELB
- die daraus resultierende Abweichung zur Ergänzungsgröße K1E2 „Veränderung der Zahl der ELB“ in Prozentpunkten.

7.3.3 Ergebnisse

Abbildung 34: Mehrfachausfälle: Relativer Fehler bei ELB und die daraus resultierende Abweichung zur Ergänzungsgröße K1E2 in Prozentpunkten

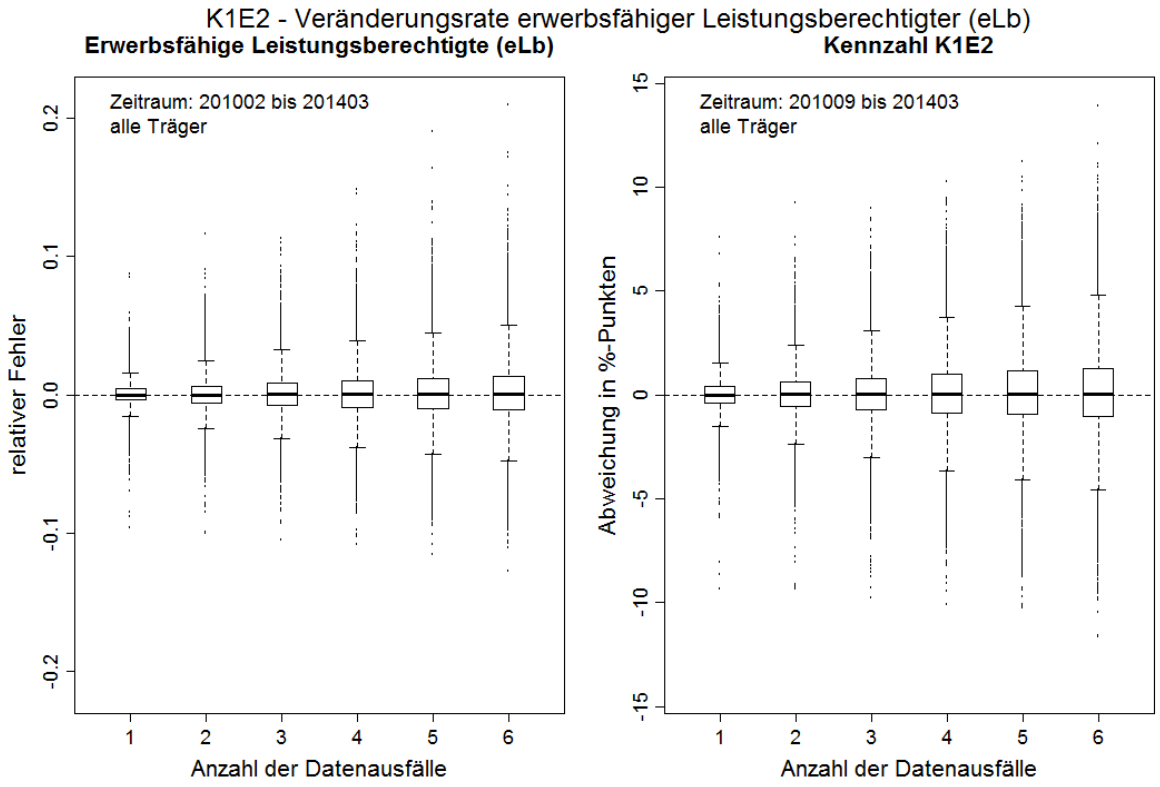


Abbildung 34 zeigt, wie sich die Fehler-Boxplots (vgl. Anhang 8.1, Lesehilfe zu Boxplots) mit zunehmender Anzahl an gereihten Mehrfachausfällen (x-Achse) verbreitern, das heißt der Schätzfehler steigt mit Mehrfachausfällen sichtbar an.

Relativer Fehler bei ELB: Die Whisker der Boxplots (Verteilung ohne Ausreißer = jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1) liegen bei Einzelausfällen im Bereich von +/- 1,6 Prozent, bei sechs hintereinander folgenden Datenausfällen zwischen –4,8 und +5,0 Prozent. Der Fehlerbereich hat sich also bei sechs Datenausfällen etwa verdreifacht. Im Extremfall liegt der relative Fehler bei ELB bei sechsmonatigen Ausfallsequenzen zwischen –12,7 und +21,0 Prozent. Mit zunehmender Anzahl von Ausfällen zeigt sich das Verfahren erwartungstreu, das heißt es liegt weder eine Unter- noch Überschätzung vor.

Abweichung zur Ergänzungsgröße K1E2: Der relative Fehler der Zählergröße überträgt sich erwartungsgemäß annähernd eins zu eins auf den Kennzahlenfehler. Die Whisker der Boxplots liegen bei Einzelausfällen im Bereich zwischen –1,5 und +1,6 Prozentpunkten, bei sechs hintereinander folgenden Datenausfällen zwischen –4,6 und +4,8 Prozentpunkten. Der Fehlerbereich hat sich also bei sechs Datenausfällen ebenfalls etwa verdreifacht. Im Extremfall liegt die Abweichung zur Ergänzungsgröße K1E2 zwischen –11,6 und +13,9 Prozentpunkten. Die Erwartungstreue des Schätzverfahrens bei ELB überträgt sich auf die Daten bei der Ergänzungsgröße K1E2, das heißt auch hier liegt weder eine Unter- noch Überschätzung vor.

Die Ergänzungsgröße K1E2 umfasst im Betrachtungszeitraum Werte zwischen –25,8 und 14,3 Prozent. Vertiefte Analysen der Ausreißer zeigen, dass besonders große Schätzfehler vor allem auf untypische Ausgangswerte im zeitlichen Umfeld der Schätzung zurückzuführen sind. Bei den hier verwendeten Fortschreibungsverfahren setzen sich zum Beispiel Niveaufehler aus ungewöhnlich niedrigen oder hohen Vormonatswerten bei längeren Ausfallsequenzen monatelang fort.

7.3.4 Fazit, Ausblick

Die Analyse zeigt, dass Mehrfachausfälle erwartungsgemäß mit einer erhöhten Schätzunsicherheit einhergehen: Bei sechs aufeinanderfolgenden Datenausfällen sind die Ergänzungsgröße K1E2 und ihre dazugehörigen Grunddaten ELB von einem etwa drei Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen. Ausgehend von den bisher beobachteten Ausfallmustern (maximal drei hintereinander folgende Ausfälle) ist jedoch von akzeptablen Unsicherheiten auszugehen. Ausreißer in der Fehlerverteilung sind in der Regel ungewöhnlich niedrigen oder hohen Ausgangswerten im zeitlichen Umfeld der Schätzung geschuldet. Insgesamt können die Schätzergebnisse als robust beurteilt werden.

7.4 Kennzahl K2 „Integrationsquote“

7.4.1 Kennzahl K2, ihre Grunddaten und dazugehörigen Imputationsverfahren

Die Kennzahl K2 „Integrationsquote“ bildet ab, in welchem Umfang ELB in Erwerbstätigkeit integriert werden konnten. Die Kennzahl ist als rollierende Jahresquote konstruiert und misst die Summe der Integrationen in den vergangenen zwölf Monaten (Zähler) im Verhältnis zum durchschnittlichen Bestand an ELB im entsprechenden Zeitraum (Nenner). Da sich die Integrationen eines Berichtsmonats immer auf den Bestand der ELB aus dem Vormonat beziehen, ist der Berechnungszeitraum der Nennergröße immer um einen Monat versetzt.

Für die Grunddaten in Zähler und Nenner werden jeweils verschiedene Imputationsverfahren verwendet. Diese Verfahren weisen untereinander keine Rückkopplungen auf, die bei der Simulation von Mehrfachausfällen berücksichtigt werden müssten.

Die Imputation für Datenlücken der Nennergröße der ELB erfolgt mit dem in Abschnitt 4.1.2 beschriebenen Fortschreibungsverfahren: fehlende Werte bei der Bestandsgröße der ELB in einem Berichtsmonat werden durch das Produkt des Vormonatswertes mit einem Fortschreibungsfaktor imputiert. Als Referenzgruppe dienen die JC des SGB-II-Vergleichstyps, dem das betreffende JC angehört. Der Fortschreibungsfaktor ist also die geschätzte mittlere Veränderungsrate zum Vormonat im gemeinsamen Vergleichstyp. Veränderungsmuster in wirtschaftsstrukturell ähnlichen Gebieten werden hierdurch gut abgebildet. Bei der Simulation von Einzelausfällen wies das Verfahren eine hervorragende Schätzgüte auf.

Die Imputation für Datenlücken der Zählergröße der Integrationen erfolgt mit einem zweistufigen Verfahren: in der ersten Stufe werden die Integrationen für das betroffene JC mit einem k-nächsten-Nachbarn-Verfahren fortgeschrieben. Hier dienen diejenigen JC als Referenzgruppe, bei denen sich die Integrationen in der Vergangenheit ähnlich entwickelt haben. Dieser Fortschreibungswert findet zusammen mit einem weiteren Prädiktor – Abgänge Arbeitsloser im Rechtskreis SGB II – in der zweiten Stufe Eingang in ein Regressionsmodell, mit dem die endgültigen Schätzwerte für die ausgefallenen Integrationen eines JC bestimmt werden (vgl. Abschnitt 6.1.2).

7.4.2 Simulation von Mehrfachausfällen

Die Simulation von Mehrfachausfällen erfolgt auf den Daten der Berichtsmonate September 2010 bis März 2014. In diesem Zeitraum liegen simultan Daten für Zähler (Integrationen) und Nenner (ELB) vor. Mit dem erforderlichen sechsmonatigen zeitlichen Vorlauf begrenzen sich die Analyseergebnisse für diese Grunddaten auf den Zeitraum zwischen März 2011 und März 2014, für die daraus abgeleiteten rollierenden Jahressummen und die Kennzahl K2 auf den Zeitraum zwischen März 2012 und März 2014. Abweichend vom regulären Imputationsverfahren wurde beim k-nächsten-Nachbarn-Verfahren für die Zählergröße der Integrationen die Zahl k der nächsten Nachbarn fest auf 20 gesetzt. Dieses Vorgehen reduziert die Simulationszeiten erheblich bei nur geringem Genauigkeitsverlust.

Aus den Schätzwerten der Simulationen werden – analog zur Konstruktion der Kennzahl K2 als rollierende Jahresquote – Jahresfehler sowohl für die Zähler- als auch für die Nennergröße berechnet. Hierfür wird angenommen, dass die untersuchten Mehrfachausfälle (von einem bis zu sechs aufeinander folgenden Datenausfällen) aus Datenausfällen bei der Nennergröße der ELB resultieren, die im jeweiligen Nachfolgemonat bzw. in den jeweiligen Nachfolgemonaten automatisch Datenlücken bei der Zählergröße der Integrationen nach sich ziehen. Die Ausfallsequenzen werden dabei jeweils am aktuellen Rand des betrachteten Zwölfmonatszeitraums simuliert. Somit sind Zähler und Nenner immer zugleich und von gleich langen Ausfallsequenzen betroffen. Grundsätzlich können Mehrfachausfälle auch ausschließlich im Zähler auftreten. In Bezug auf den Schätzfehler stellen gleichzeitige Mehrfachausfälle in Zähler und Nenner jedoch den schlechteren Fall dar. Somit bieten die hier vorliegenden Analysen eine Maximalabschätzung des Schätzfehlers.

Für jede simulierte Ausfallsequenz werden über alle JC und Berichtsmonate mittlere relative Jahresfehler für Zähler bzw. Nenner berechnet, das heißt die Abweichungen in den rollierenden Jahressummen werden über alle JC und Berichtsmonate gemittelt. Im Ergebnis erhält man:

- den mittleren relativen Jahresfehler für
  - die Zählergröße der Integrationen
  - die Nennergröße der ELB
- die daraus resultierende mittlere Abweichung zur Kennzahl K2 „Integrationsquote“ in Prozentpunkten.

7.4.3 Ergebnisse

Abbildung 35: Mehrfachausfälle: Relativer Jahresfehler bei Integrationen und ELB und die daraus resultierende Abweichung zur Kennzahl K2 in Prozentpunkten

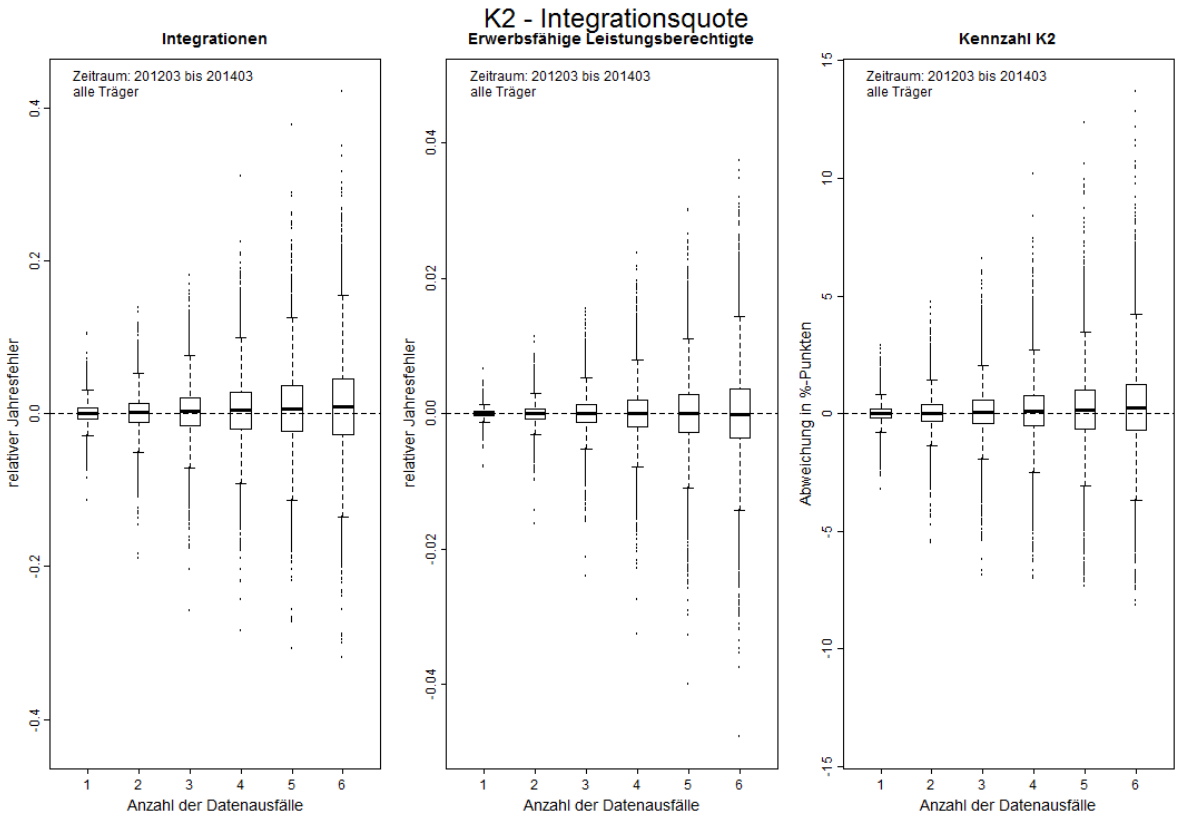


Abbildung 35 zeigt, wie sich die Fehler-Boxplots (vgl. Anhang 8.1, Lesehilfe zu Boxplots) über alle drei Größen mit zunehmender Anzahl an gereihten Mehrfachausfällen (x-Achse) verbreitern, das heißt der Schätzfehler steigt mit Mehrfachausfällen sichtbar an.

Relativer Jahresfehler bei Integrationen: Die Whisker der Boxplots (Verteilung ohne Ausreißer = jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1) liegen bei Einzelausfällen im Bereich von –2,9 bis +3,0 Prozent, bei sechs hintereinander folgenden Datenausfällen zwischen –13,6 und +15,4 Prozent.

Der Fehlerbereich hat sich also bei sechs Datenausfällen beinahe verfünffacht. Im Extremfall liegt der relative Jahresfehler bei Integrationen bei sechsmonatigen Ausfallsequenzen zwischen –31,8 und +42,2 Prozent. Dabei zeigt sich im Median mit zunehmender Anzahl von Ausfällen eine leichte Überschätzung.

Relativer Jahresfehler bei ELB: Bei der Nennergröße der ELB ist der relative Jahresfehler sehr viel kleiner und die Whisker der Boxplots liegen bei Einzelausfällen bei etwa +/-0,1 Prozent, bei sechs hintereinander folgenden Datenausfällen bei etwa +/-1,4 Prozent. Trotz dieser nominell sehr kleinen Fehlerwerte vergrößert sich der Fehlerbereich bei sechs Datenausfällen etwa um den Faktor 14. Im Extremfall liegt der relative Jahresfehler bei ELB bei sechsmonatigen Ausfallsequenzen zwischen –4,8 und +3,7 Prozent. Mit zunehmender Anzahl von Ausfällen zeigt sich das Verfahren hier allerdings erwartungstreu, das heißt es liegt weder eine Unter- noch Überschätzung vor.

Abweichung zur Kennzahl K2: Bei der Bestimmung des Kennzahlenfehlers treten teilweise Ausgleicheffekte zwischen den Fehlern der Zähler- und Nennergröße auf, aber auch Verstärkungseffekte: der betragsmäßig große relative Jahresfehler bei Integrationen im Zähler überträgt sich in Relation zum kleinen relativen Jahresfehler bei ELB im Nenner in starkem Maße auf die Abweichungen zur Kennzahl in Prozentpunkten und dominiert somit den Kennzahlenfehler.[116] Die Whisker der Boxplots liegen bei Einzelausfällen zwischen +/-0,8 Prozentpunkten, bei sechs hintereinander folgenden Datenausfällen zwischen –3,7 und +4,2 Prozentpunkten. Der Fehlerbereich hat sich also bei sechs Datenausfällen – wie beim Jahresfehler für die Zählergröße Integrationen – beinahe verfünffacht. Im Extremfall liegt die Abweichung zur Kennzahl K2 zwischen –8,1 und +13,7 Prozentpunkten. Dabei zeigt sich im Median – wie bei den Integrationen im Zähler – mit zunehmender Anzahl von Ausfällen ebenfalls eine leichte Überschätzung.

Die Kennzahl K2 umfasst im Betrachtungszeitraum Werte zwischen +13,6 und +56,0 Prozent. Insgesamt treten große Abweichungen eher bei JC mit großer Kennzahl K2 auf. Vertiefte Analysen der Ausreißer zeigen, dass besonders große Schätzfehler vor allem untypischen Ausgangswerten im zeitlichen Umfeld der Schätzung geschuldet sind. Bei den hier verwendeten Fortschreibungsverfahren setzen sich zum Beispiel Niveaufehler aus ungewöhnlich niedrigen oder hohen Vormonatswerten bei längeren Ausfallsequenzen monatelang fort und können sich im Zusammenspiel von Zähler und Nenner gegenseitig verstärken. Solche ungewöhnlichen Werte können zudem bei der k-nächste-Nachbarn-Fortschreibung für die Zählergröße der Integrationen direkt zu einer verzerrten Schätzung der Veränderungsrate oder indirekt zu einem verzerrten Abgleich mit anderen JC führen.

7.4.4 Fazit, Ausblick

Die Analyse zeigt, dass Mehrfachausfälle erwartungsgemäß mit einer erhöhten Schätzunsicherheit einhergehen. Bei sechs aufeinanderfolgenden Datenausfällen ist die Kennzahl K2 von einem etwa fünf Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen und wird im Median leicht überschätzt. Ausgehend von den bisher beobachteten Ausfallmustern (maximal fünf hintereinander folgende Ausfälle) ist jedoch von akzeptablen Unsicherheiten auszugehen. Ausreißer in der Fehlerverteilung sind in der Regel

[116] Der relative Kennzahlenfehler kann über die relativen Jahressummenfehler der Zähler- und der Nennergröße berechnet werden gemäß: $(f_{Int} - f_{ELB})/(1 + f_{ELB})$. Da der relative Fehler bei Integrationen deutlich breiter streut als jener bei ELB, wird der relative Kennzahlenfehler durch den Fehler bei Integrationen dominiert. Die relative Fehlerverteilung bei Integrationen ist jener des Kennzahlenfehlers sehr ähnlich.

ungewöhnlich niedrigen oder hohen Ausgangswerten im zeitlichen Umfeld der Schätzung geschuldet. Insgesamt können die Ergebnisse als robust angesehen werden.

7.5 Kennzahl K3 „Veränderung des Bestands an Langzeitleistungsbeziehenden “

7.5.1 Kennzahl K3, ihre Grunddaten und dazugehörigen Imputationsverfahren

Die Kennzahl K3 „Veränderung des Bestands an Langzeitleistungsbeziehern“ bietet zentrale Informationen zum Ziel „Vermeidung von langfristigem Leistungsbezug“. Sie misst die Veränderung des Bestands an LZB eines Jobcenters zwischen dem Berichtsmonat (Zähler) und dem entsprechenden Vorjahresmonat (Nenner).

Nachfolgend geht es ausschließlich um sogenannte isolierte Datenlücken. Diese betreffen die Phase, in der für das JC keine plausiblen Daten vorliegen. Es gibt bei LZB auch sogenannte Langzeitlücken. Sie entstehen, wenn zurückliegende isolierte Datenlücken infolge ausgefallener Personenbestandsdaten für die Ermittlung der Bezugsdauern und somit für die Zählung der LZB nicht mehr zu kompensieren sind. Für Langzeitlücken bei LZB wurde ein gesondertes Schätzverfahren entwickelt (vgl. Abschnitt 4.3.2.2). Es wird dann eingesetzt, wenn nach isolierten Datenausfällen wieder plausible Monatsdaten vorliegen, aber eine vollständige Dauermessung nicht mehr bzw. noch nicht wieder möglich ist.

Die Imputation der nachfolgend fokussierten isolierten Datenlücken erfolgt durch das gleiche Fortschreibungsverfahren wie bei ELB (vgl. Abschnitt 4.3.2.1): fehlende Werte bei der Bestandsgröße der LZB in einem Berichtsmonat werden durch das Produkt des Vormonatswertes mit einem Fortschreibungsfaktor imputiert. Als Referenzgruppe dienen die JC des SGB-II-Vergleichstyps, dem das betreffende JC angehört. Der Fortschreibungsfaktor ist also die geschätzte mittlere Veränderungsrate zum Vormonat im gemeinsamen Vergleichstyp. Veränderungsmuster in wirtschaftsstrukturell ähnlichen Gebieten werden hierdurch gut abgebildet. Bei der Simulation von Einzelausfällen wies das Verfahren eine hervorragende Schätzgüte auf.

7.5.2 Simulation von Mehrfachausfällen

Die Simulation von Mehrfachausfällen erfolgt analog zum Vorgehen bei ELB (vgl. Abschnitt 7.3.2). Sie beruht auf den Daten der Berichtsmonate August 2009 bis März 2014. Mit dem erforderlichen sechsmonatigen zeitlichen Vorlauf begrenzen sich die Analyseergebnisse für LZB auf den Zeitraum zwischen Februar 2010 und März 2014, für die Kennzahl K3 zwischen September 2010 und März 2014. Aus den Schätzwerten der Simulation werden relative Fehler für LZB bzw. die K3 berechnet.

Grundsätzlich können Mehrfachausfälle sowohl im Zähler (LZB im Berichtsmonat) als auch im Nenner (LZB im Vorjahresmonat) auftreten. Es ist aber davon auszugehen, dass sich die Simulationsergebnisse für Ausfälle im Nenner ähnlich verhalten wie die Ergebnisse für Ausfälle im Zähler. Ergänzende Analysen zeigen außerdem, dass die Konstellation eines gleich-zeitigen (Mehrfach-)Ausfalls in Zähler und Nenner keinen nennenswerten Einfluss auf die Schätzgüte haben (vgl. Abschnitt 7.6). Die nachfolgende Analyse konzentriert sich deshalb ausschließlich auf Ausfälle im Zähler. Die betreffende Kennzahl K3 berechnet

sich anschließend aus den simulierten Schätzwerten im Zähler und den dazugehörigen nichtsimulierten Vorjahreswerten im Nenner. Im Ergebnis erhält man:

- den relativen Fehler für die Zählergröße LZB
- die daraus resultierende Abweichung zur Kenngröße K3 „Veränderung des Bestands an LZB“ in Prozentpunkten.

7.5.3 Ergebnisse

Abbildung 36: Mehrfachausfälle: Relativer Fehler bei LZB und die daraus resultierende Abweichung zur Kennzahl K3 in Prozentpunkten

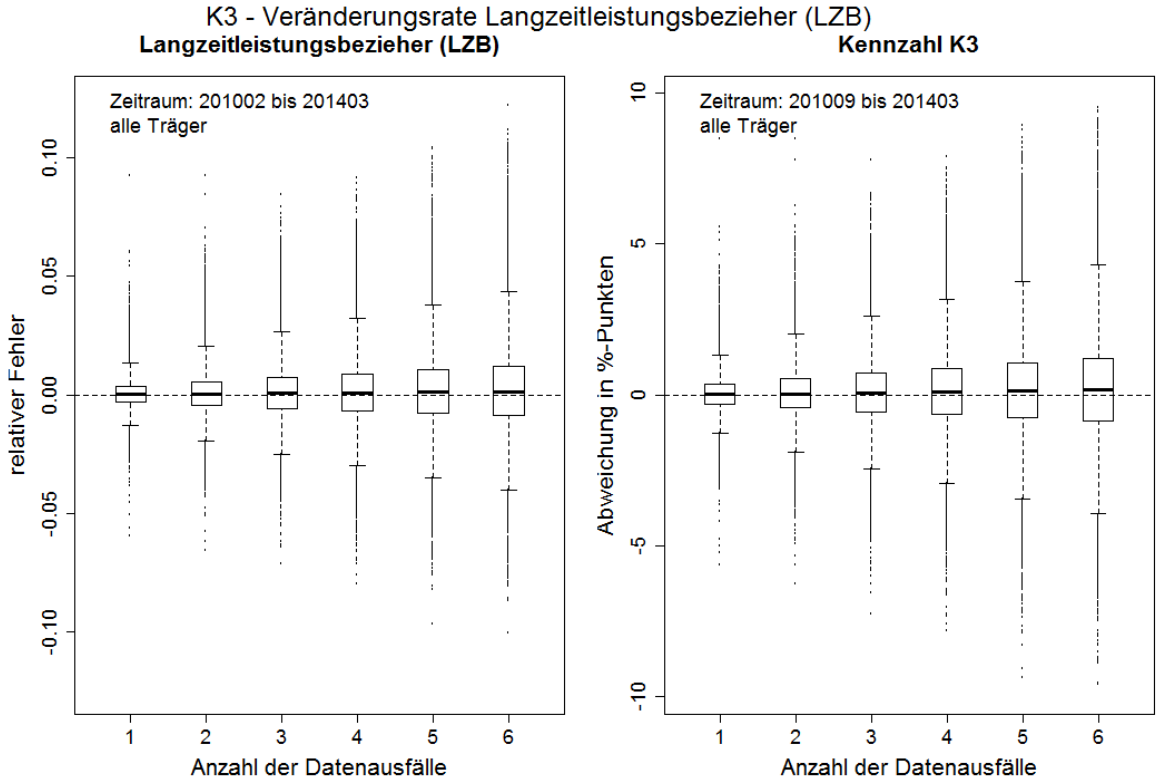


Abbildung 36 zeigt, wie sich die Fehler-Boxplots (vgl. Anhang 8.1, Lesehilfe zu Boxplots) mit zunehmender Anzahl an gereihten Mehrfachausfällen (x-Achse) verbreitern, das heißt der Schätzfehler steigt mit Mehrfachausfällen sichtbar an.

Relativer Fehler bei LZB: Die Whisker der Boxplots (Verteilung ohne Ausreißer = jeweiliges Quartil +/- 1,5-facher Interquartilsabstand, vgl. Anhang 8.1) liegen bei Einzelausfällen im Bereich von –1,3 bis +1,4 Prozent, bei sechs hintereinander folgenden Datenausfällen zwischen –4,0 und +4,4 Prozent. Der Fehlerbereich hat sich also bei sechs Datenausfällen etwa verdreifacht. Im Extremfall liegt der relative Fehler bei LZB bei sechsmonatigen Ausfallsequenzen zwischen –10,0 und +12,2 Prozent. Dabei zeigt sich im Median mit zunehmender Anzahl von Ausfällen eine geringfügige Überschätzung.

Abweichung zur Kennzahl K3: Der relative Fehler der Zählergröße überträgt sich erwartungsgemäß annähernd eins zu eins auf den Kennzahlenfehler. Die Whisker der Boxplots liegen bei Einzelausfällen im Bereich zwischen +/1,3 Prozentpunkten, bei sechs hintereinander folgenden Datenausfällen zwischen –3,9 und +4,3 Prozentpunkten. Der Fehlerbereich hat sich also bei sechs Datenausfällen ebenfalls etwa verdreifacht. Im Extremfall liegt die Abweichung zur Kennzahl K3 zwischen –9,6 und +9,5 Prozentpunkten. Dabei zeigt sich im Median mit zunehmender Anzahl von Ausfällen ebenfalls eine geringfügige Überschätzung.

Die Kennzahl K3 umfasst im Betrachtungszeitraum Werte zwischen –20,4 und 12,3 Prozent. Vertiefte Analysen der Ausreißer zeigen, dass besonders große Schätzfehler auf entweder kleine JC oder vor allem auf untypische Ausgangswerte im zeitlichen Umfeld der Schätzung zurückzuführen sind. Bei den hier verwendeten Fortschreibungsverfahren setzen sich zum Beispiel Niveaufehler aus ungewöhnlich niedrigen oder hohen Vormonatswerten bei längeren Ausfallsequenzen monatelang fort.

7.5.4 Fazit, Ausblick

Die Analyse zeigt, dass Mehrfachausfälle erwartungsgemäß mit einer erhöhten Schätzunsicherheit einhergehen: Bei sechs aufeinanderfolgenden Datenausfällen sind die Kennzahl K3 und ihre dazugehörigen Grunddaten LZB von einem etwa drei Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen und werden im Median geringfügig überschätzt. Ausgehend von den bisher beobachteten Ausfallmustern (maximal fünf hintereinander folgende isolierte Ausfälle, die mit dem beschriebenen statistischen Schätzverfahren imputiert werden) ist jedoch von akzeptablen Unsicherheiten auszugehen. Ausreißer in der Fehlerverteilung sind in der Regel ungewöhnlich niedrigen oder hohen Ausgangswerten im zeitlichen Umfeld der Schätzung geschuldet. Insgesamt können die Schätzergebnisse als robust beurteilt werden.

7.6 Abschätzung des Schätzfehlers bei gleichzeitigen Mehrfachausfällen in Zähler und Nenner

Für die Kenngrößen K1, K1E2 und K3 wurden in den Abschnitten 7.2, 7.3 und 7.5 die Auswirkungen von Mehrfachausfällen in den jeweiligen Zählergrößen analysiert. Dieser Abschnitt zeigt die ergänzende Analyse von Auswirkungen gleichzeitiger Mehrfachausfälle in Zähler und Nenner. Es wurde für die Kenngrößen K1E2 und K3 geprüft, wie sich hierdurch der jeweilige Schätzfehler in Prozentpunkten (Abweichung von der jeweiligen Kenngröße ohne Ausfälle) verändert. Es wurde analog vorgegangen wie in den Abschnitten 7.3.2 (zu ELB und K1E2) und 7.5.2 (zu LZB und K3).

Tabelle 8 stellt die Ergebnisse für gereihte sechsfache Mehrfachausfälle nur im Zähler denjenigen für gleichzeitige gereihte sechsfache Mehrfachausfälle in Zähler und Nenner gegenüber. Sie zeigt für beide Kenngrößen ähnliche Ergebnisse im Vergleich zwischen den beiden Ausfallvarianten. Sind Zähler und Nenner ausgefallen, so erhöhen sich die Extremwerte der Schätzfehler leicht, der Median hingegen tendiert stärker zu Null. Auch die übrigen Verteilungsparameter weisen nur geringfügige Verschiebungen auf.

Empirisch sind Fälle bislang sehr selten aufgetreten, in denen sowohl Zähler als auch Nenner aus geschätzten Werten bestanden. Für nur ein einziges JC ergab sich bisher die Konstellation, dass zu einem Einzelausfall im Zähler ein zweifach gereihter Mehrfachausfall im Nenner hinzukam.

Tabelle 8: Auswirkungen von Mehrfachausfällen nur im Zähler gegenüber gleichzeitigen Mehrfachausfällen in Zähler und Nenner bei ausgewählten Kennzahlen

	Schätzfehler in Prozentpunkten (Februar 2012 bis März 2014)			
	K1E2		K3	
	Veränderung der Zahl der ELB		Veränderung des Bestands an LZB	
Zähler	6 Ausfälle	6 Ausfälle	6 Ausfälle	6 Ausfälle
Nenner	kein Ausfall	6 Ausfälle	kein Ausfall	6 Ausfälle
Minimum	-11,6	-14,3	-9,6	-12,1
Maximum	13,9	14,2	9,5	12,4
Median	0,1	0,0	0,2	0,0
25-%-Quantil	-1,1	-1,4	-0,8	-1,1
25-%-Quantil – 1,5*IQA	-4,6	-5,4	-2,3	-2,2
75-%-Quantil	1,3	1,3	1,3	1,1
75-%-Quantil + 1,5*IQA	4,9	5,4	4,4	4,3

Die Veränderung der Schätzfehler durch die Berücksichtigung von gleichzeitigen Mehrfachausfällen in Zähler und Nenner ist in der Simulation von so geringem Ausmaß, dass die Ergebnisse aus der Analyse von Mehrfachausfällen nur im Zähler als hinreichend aussagekräftig erscheinen. Zudem sind die dargestellten Ergebnisse für K1E2 und K3 so eindeutig, dass eine analoge Analyse für die Kennzahl K1 (bzw. die zugrundeliegenden Leistungen zum Lebensunterhalt) als entbehrlich angesehen wurde. Die Ergebnisse zu K1E2 und K3 lassen nicht befürchten, dass die Analyse lediglich von Zählerausfällen zu einer systematischen Fehleinschätzung führt.

7.7 Fazit aus den Mehrfachausfallanalysen

Während der Entwicklungsarbeiten an den Verfahren zur Datenimputation für Kennzahlen nach § 48a SGB II wurden alle Imputationsverfahren standardmäßig quantitativ-simulativen Bewertungsanalysen unterzogen. Hierbei wurden zunächst nur Einzelwerte als fehlend simuliert („leave-one-out“, vgl. Abschnitt 2.1.2). Tatsächlich liegen aber manchmal – wenn auch sehr selten – Mehrfachausfälle vor, sodass sich die Datenlücken bei einzelnen JC häufen. Die vorangegangenen Abschnitte stellen die Ergebnisse einer Analyse zu den Auswirkungen von Mehrfachausfällen auf die Schätzfehler bei den drei Kennzahlen K1 (Veränderung der Summe der Leistungen zum Lebensunterhalt), K2 (Integrationsquote) und K3 (Veränderung des Bestands an LZB) sowie der Ergänzungsgröße K1E2 (Veränderung der Zahl der ELB) vor. Es wurden bis zu sechs aufeinanderfolgende Datenausfälle simuliert.

Die Analyse zeigt, dass Mehrfachausfälle erwartungsgemäß mit einer erhöhten Schätzunsicherheit einhergehen: Bei simulierten sechs aufeinanderfolgenden Datenausfällen…

- …sind die Kennzahl K1 und ihre dazugehörigen Grunddaten LLU von einem beinahe drei Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen und werden im Median geringfügig unterschätzt.

- …sind die Ergänzungsgröße K1E2 und ihre dazugehörigen Grunddaten ELB von einem etwa drei Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen. Mit zunehmender Anzahl von Ausfällen zeigt sich das Verfahren erwartungstreu, das heißt es liegt weder eine Unter- noch Überschätzung vor.
- …ist die Kennzahl K2 von einem etwa fünf Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen und wird im Median geringfügig überschätzt.
- …sind die Kennzahl K3 und ihre dazugehörigen Grunddaten LZB von einem etwa drei Mal so breiten Fehlerbereich betroffen wie bei Einzelausfällen und werden im Median geringfügig überschätzt.

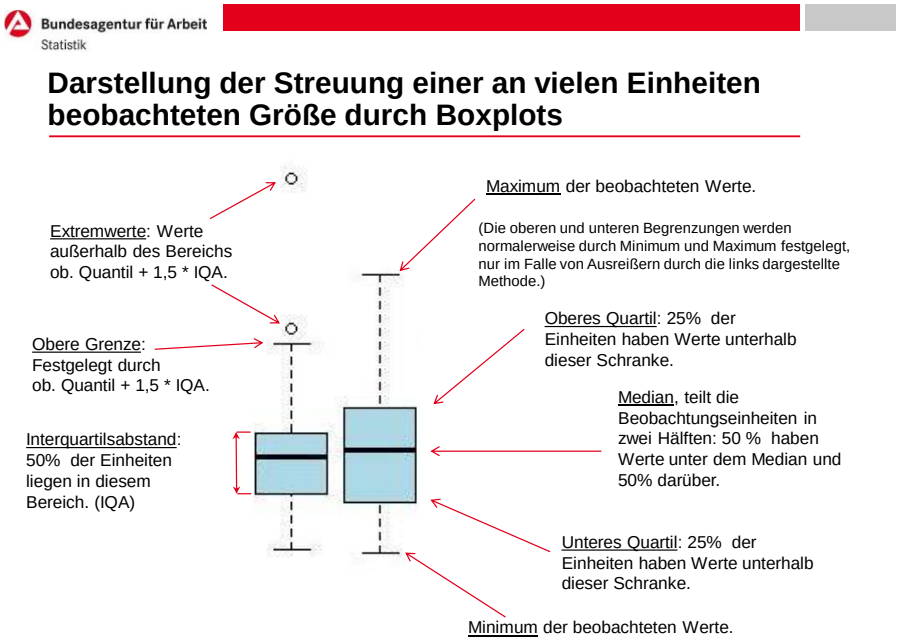
Ausgehend von den bisher beobachteten Ausfallmustern (maximal fünf hintereinander folgende Ausfälle über alle genannten Themen hinweg) ist jedoch von akzeptablen Unsicherheiten auszugehen. Ausreißer in der Fehlerverteilung sind in der Regel ungewöhnlich niedrigen oder hohen Ausgangswerten im zeitlichen Umfeld der Schätzung geschuldet. Insgesamt können die Ergebnisse der Imputationsverfahren – auch bei Mehrfachausfällen – als robust angesehen werden.

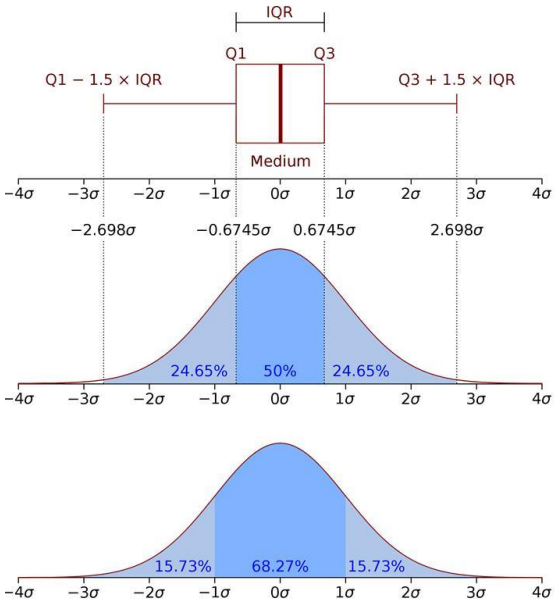
8 Anhang

8.1 Lesehilfe zu Boxplots

Oft liegen Messergebnisse für verschiedene Einheiten (z. B. JC) vor und eine Aufgabe besteht darin, die Ergebnisse einerseits zusammenfassend darzustellen, aber andererseits auch die Verschiedenartigkeit der Ergebnisse für die Einheiten aufzuzeigen. Das spielt insbesondere dann eine Rolle, wenn Ergebnisse auf bestimmten Aggregationsebenen (zum Beispiel Länder oder Vergleichstypen) miteinander verglichen werden sollen. Dann stellt sich nicht nur die Frage nach dem Mittelwert für die einzelnen Länder, sondern auch nach der Spanne der Werte für die JC in diesem Land oder die Spanne für die mittleren 50 %.

In den 70-er Jahren wurde dafür eine standardisierte Darstellungsform entwickelt, die den Namen Boxplot („Kastengrafik“) trägt. Die nebenstehende Abbildung erläutert die einzelnen Elemente einer solchen Darstellung. Der mittlere Kasten wird begrenzt durch unteres und oberes Quartil und geteilt durch den Median. Im Falle einer symmetrischen Verteilung sind die oberen und unteren Elemente gleich weit vom Median entfernt.





Quelle: http://en.wikipedia.org/wiki/Box_plot
IQR (interquartile range) = IQA, Medium = Median

Die nebenstehende Abbildung zeigt für den Fall normalverteilter Messergebnisse den Zusammenhang zwischen den Elementen des Boxplots und den Verteilungsmassen unter der Glockenkurve der Normalverteilung. Dargestellt ist der Zusammenhang zwischen Interquartilsabstand (IQA) und der Standardabweichung

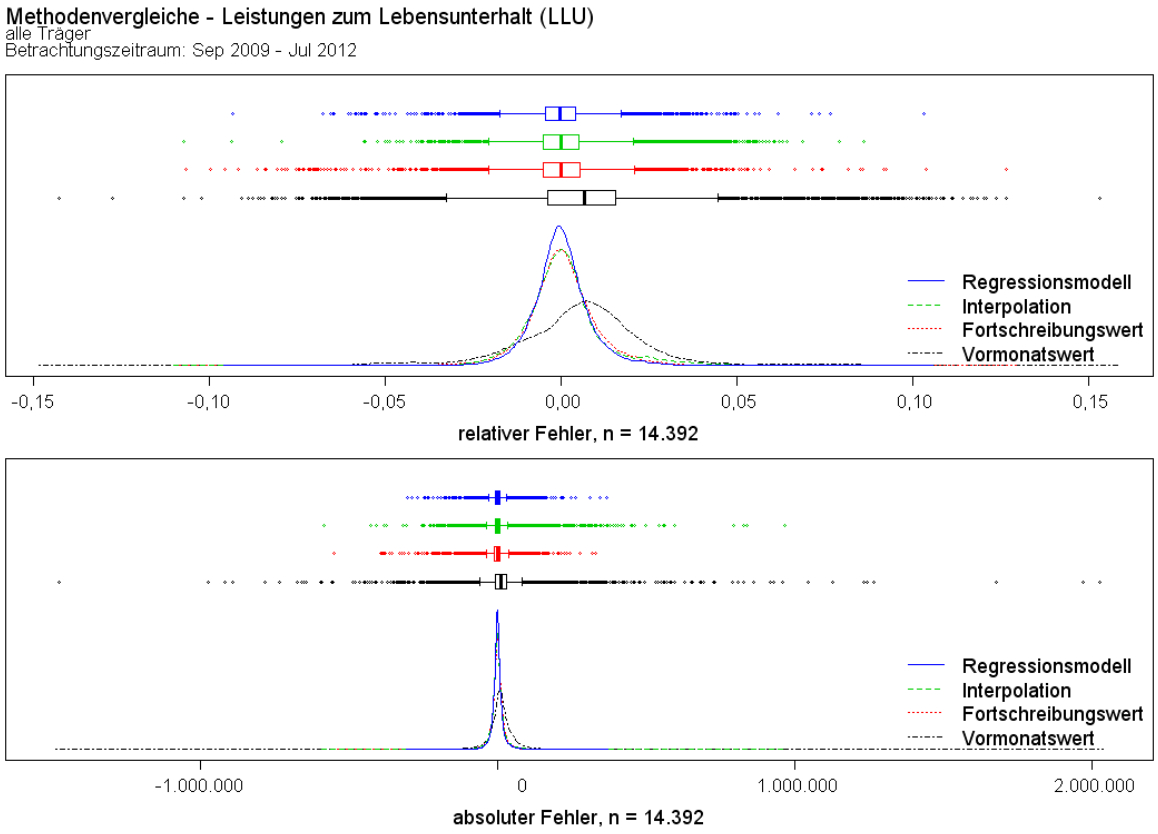
$$IQA = 2 \cdot 0{,}6745 \cdot \sigma \approx 1{,}35 \cdot \sigma$$

Die Abbildung legt den Hintergrund für die Festlegung der Kriterien für Ausreißer offen: Innerhalb der unteren und oberen Grenzwerte liegen bei einer Normalverteilung 99,3 % aller Fälle, außerhalb sind nur 0,7 % zu erwarten, also zum Beispiel bei rund 400 Jobcentern nur ca. 3 Fälle. Im Falle „schiefer“ Verteilungen liegt die Anzahl als potentielle Ausreißer markierter Fälle häufig etwas höher.

8.2 Lesehilfe zu kombinierten Boxplot-Dichte-Grafiken

Der im Abschnitt 2.1.3 erläuterte Standardvergleich erfolgt unter anderem anhand kombinierter Boxplot-Dichte-Grafiken, die in der Regel im jeweiligen Bewertungs-Abschnitt zu den Imputationsverfahren abgebildet sind. Dargestellt sind die Verteilungen der Abweichungen zwischen tatsächlichen Werten und simulierten Schätzwerten, normalerweise immer für das vorgeschlagene Schätzverfahren (blau, liniert) und die Vergleichsverfahren „Vormonatswertersetzung“ (schwarz, teil-gepunktet), „Interpolationsverfahren“ (grün, gestrichelt) und „Fortschreibungswertverfahren“ (rot, gepunktet) – im oberen Abschnitt in der Relativ-, im unteren in der Absolutbetrachtung.

Abbildung 37: Kombinierte Boxplot-Dichte-Grafik für den Methodenvergleich am Beispiel LLU



Geringe Schätzfehler schlagen sich bei den Boxplots (vgl. vorangehender Anhang 8.1) in schmalen Boxen und kurzen Whiskern, bei den Dichtekurven[117] in schmalen und hohen Kurven nieder. Gröbere Schätzfehler, also Ausreißer, sind gut bei den Boxplots durch Punkte außerhalb der Whisker zu erkennen – je mehr,

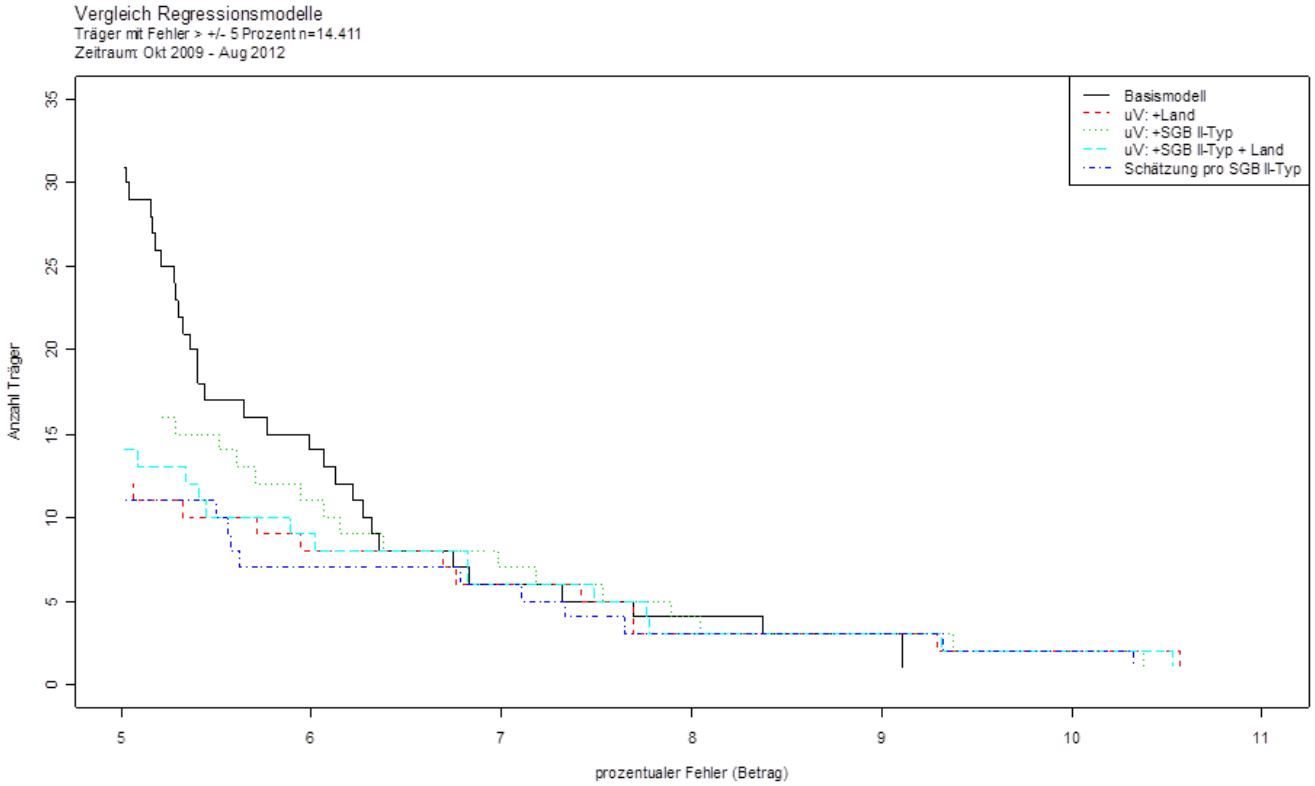
[117] Die in dieser Methodenübersicht betrachteten Schätzfehler sind stetige Daten, deren relative Häufigkeitsverteilungen mittels Dichtekurven dargestellt werden können. Die Dichtekurve ist eng verwandt mit dem Histogramm. Dort werden die Daten in Größenklassen unterteilt. Die Klassen werden dann grafisch in direkt aneinander angrenzende Rechtecke übersetzt, deren Flächeninhalt die relative Klassenhäufigkeit darstellt. Die erhaltene – oft sprunghafte – Treppenfunktion, kann durch mathematische Umformung in eine glatte Dichtekurve übersetzt werden – eine Darstellungsform, die angesichts der hier vorliegenden umfangreichen Datensätze (Fallzahlen im zweistelligen Tausenderbereich) besonders gut geeignet ist.

desto mehr gröbere Schätzfehler gibt es, je weiter außen gelegen, umso größer ist der jeweilige Schätzfehler. In der Beispielabbildung für Schätzfehler bei der Schätzung von Leistungen zum Lebensunterhalt sind beim vorgeschlagenen Regressionsmodell die Boxen am schmalsten, die Whisker am kürzesten, die Ausreißer streuen am wenigsten nach außen, die Dichtekurven sind die schmalsten und höchsten und ihr Maximum liegt über dem Nullwert. Das vorgeschlagene Regressionsmodell ist den Standardvergleichsmodellen also bereits nach der grafischen Sichtung überlegen. Exemplarisch wird hier auch die mangelnde Erwartungstreue des Standardvergleichsmodells der Vormonatswertersetzung deutlich (vgl. Fußnote 12, Seite 29): im betrachteten Zeitraum unterlagen die Leistungen zum Lebensunterhalt einem abnehmenden Trend. Das Vormonatswertverfahren würde die Messgröße systematisch überschätzen. Die Verzerrung zeigt sich in einer leichten Verschiebung sowohl des Medianstriches, der Boxen als auch der Dichtekurven in den positiven Wertebereich hinein.

An der Skala der x-Achse sind einige Verteilungsparameter ablesbar. Zum Beispiel enden beim vorgeschlagenen Regressionsmodell die Whisker des Boxplots der relativen Abweichungen bei ca. +/- 2% – ein sehr geringer Schätzfehler im Allgemeinen. Es gibt zwar Ausreißer, aber meist im Bereich bis +/-5 %, in Einzelfällen bis maximal +/-10 %. Mit dem Regressionsmodell treten demnach nie Schätzfehler über +/-10 % auf – auch in Bezug auf Ausreißer eine hervorragende Schätzgüte.

8.3 Schrittweise Erweiterung des Regressionsmodells für die LLU-Schätzung[118]

Abbildung 38: Modellvergleich – schrittweise Erweiterung des Regressionsmodells für die LLU-Schätzung



Die Auswahl des Schätzmodells zur Fortschreibung der LLU erfolgt schrittweise: Ein erstes Basismodell umfasst die Größen ELB, NEF und BG. Die quantitative Bewertung der Modelle erfolgt anhand der betragsmäßigen Abweichungen der Schätzungen vom tatsächlichen Wert, den betragsmäßigen Residuen. Im Fokus steht hierbei der Zeitraum Oktober 2009 bis August 2012. Bereits das Basismodell liefert im Mittel gute Werte; für etliche JC weisen die Schätzungen allerdings vergleichsweise hohe Abweichungen im niedrigen einstelligen Prozentbereich auf.[119] Die Erweiterung des Modells um die prozentuale Veränderung der LLU im Bundesland sowie im SGB-II-Vergleichstyp ist sinnvoll, da dadurch nicht nur der mittlere prozentuale Fehler, sondern auch die Zahl der Ausreißer, für die das Schätzergebnis stark vom tatsächlichen Wert abweicht, deutlich gesenkt werden kann.[120] Mit der prozentualen Veränderung der Zahl der ELB, NEF und BG sowie der prozentualen Veränderung der LLU im jeweiligen Bundesland und SGB-II-Vergleichstyp steht ein Set an unabhängigen Variablen zu Verfügung, mit dem sich die Veränderung

[118] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[119] Im Median liegt der betragsmäßige Fehler für dieses Basismodell bei 7.954 Euro, dies entspricht einem mittleren prozentualen Fehler von 0,46 % (Median). Bei 31 JC beträgt die Abweichung allerdings 5 % oder mehr (vgl. obige Abbildung), bei 478 JC liegt der Fehler über 3 %.

[120] Die Veränderung der LLU im Bundesland und die Veränderung der LLU im SGB-II-Vergleichstyp des jeweiligen JC messen oft ähnliche, teils aber auch unterschiedliche Kontext-Effekte (zum Beispiel konjunkturelle Entwicklungen, Antrags- und Erfassungsverhalten infolge von Ferien- und Feiertagen). Das Modell mit beiden Kontext-Variablen weist einen höheren Anteil erklärter Varianz (R-Quadrat) und einen geringeren mittleren prozentualen Fehler auf als die Modelle mit nur einer Kontext-Variablen und wird deswegen favorisiert.

der LLU im Mittel gut erklären lässt. Nun impliziert die Formulierung eines einheitlichen Schätzmodells für alle JC, dass die Beziehung zwischen den unabhängigen und der abhängigen Variablen für alle JC gleich ist. Die Berechnung je einer Schätzgleichung pro JC würde diese Annahme zwar überflüssig machen, allerdings auch die Zahl der zur Schätzung zur Verfügung stehenden Fälle auf die Zahl der Messzeitpunkte t reduzieren. Um möglichen JC-spezifischen Unterschieden im Gewicht der unabhängigen Variablen gerecht zu werden und zugleich eine ausreichende Zahl von Datenpunkten zur Verfügung zu haben, passiert im letzten Schritt ein Split: die Regressionsschätzung erfolgt pro SGB-II-Vergleichstyp. Dieses Modell zeigt sich sowohl hinsichtlich des betragsmäßigen mittleren Fehlers (Median 7.365 Euro bzw. 0,42 %) als auch der Zahl der Ausreißer den vorherigen Regressionsschätzungen überlegen.[121]

[121] Es gibt hier nur noch elf Ausreißer mit einer Abweichung von 5 % oder mehr vom tatsächlichen Wert. Damit ist dieses Modell auch in Bezug auf Ausreißer das beste (vgl. obige Abbildung und Abbildung 3).

8.4 Datenimputation mit dem Optimierungsmodell für Bewegungsdaten (Zu- und Abgänge von ELB bzw. LZB)

Als Datenimputationsverfahren für Bewegungsdaten von ELB (vgl. Abschnitt 5.1.2) bzw. LZB (vgl. Abschnitt 5.2.2) kommt ein Optimierungsmodell zum Einsatz, das eigens auf das spezifische Schätzproblem zugeschnitten wurde. Die Modellierung beruht im Kern auf der sogenannten Stock-Flow-Gleichung, die allgemein den Zusammenhang zwischen Stromgrößen und Beständen beschreibt: der Bestand eines Berichtsmonats ergibt sich rechnerisch aus dem Bestand des Vormonats minus der Abgänge plus der Zugänge zwischen den Stichtagen. Hier wird die Gleichung noch um einen Korrekturterm ρ (Rho) erweitert.

- In Bezug auf die ELB-Bewegungszählung steht der Term ρ für vermeintliche Bewegungen, die aus technischen oder operativen Gründen gemessen werden, obwohl ein ELB seinen Leistungsbezug eigentlich nicht unterbrochen hat. Somit werden zwar Bewegungen gemessen, aber der ELB-Bestand bleibt unverändert. Diese überzähligen Bewegungen werden gemessen, wenn ein ELB zum Beispiel aufgrund verspäteter Antragstellung bei der Weiterbewilligung seinen Leistungsbezug mehr als sieben Tage unterbricht (vgl. Abschnitt 5.1.1). Diese Unterbrechungsregel führt zu einer Verzerrung des Stock-Flow-Modells, die durch den Korrekturterm ρ aufgefangen wird.
- In Bezug auf die LZB-Bewegungszählung steht der Term ρ seit einer Revision der Zugangsmessung für Bestandsveränderungen, die nicht durch Zugänge in die bzw. Abgänge aus der Gruppe der LZB, sondern durch JC-Wechsel erfolgen. Personen, die in ein anderes JC-Gebiet gezogen sind, fließen nicht in die Bewegungszählung bei LZB ein, da sie vorher wie nachher LZB sind. Hierdurch verschieben sich die LZB-Bestände der JC, ohne dass LZB-Bewegungen gemessen werden, weil der Indikator Mindest-Anwesenheitsdauer (21 aus 24 Monaten) schon vorher erfüllt war (vgl. Abschnitt 5.2.2).[122]

Die allgemeine um den Korrekturterm ρ erweiterte Stock-Flow-Gleichung lautet:

$$Bestand_t^k = Bestand_{t-1}^k - Abgang_t^k + Zugang_t^k + \rho_t^k \qquad (15)$$

mit:
k JC
t Zeitpunkt

Der Term ρ fungiert in der Gleichung als korrigierende Saldogröße. Er gleicht sich auf Bundesebene zwar aus, fängt aber bei Bedarf auf Ebene der JC etwaige Verzerrungen auf.

Wegen des Stock-Flow-Zusammenhangs sind die jeweiligen Zu- und Abgänge immer gleichzeitig von Datenlücken betroffen. Postuliert man den Stock-Flow-Zusammenhang als zentrales Gütekriterium und Kernbedingung für ein gutes Schätzmodell bei Bewegungsdaten, dann müssen die jeweiligen Zu- und Abgänge immer simultan geschätzt werden, um die wechselseitige Abhängigkeit zu berücksichtigen.

[122] Die genannte Revision der LZB-Zugangsmessung erfolgte im September 2014 und begründete Version 3.1 dieser Methodenübersicht (November 2014). Zuvor waren es außerdem altersbedingte Übergänge in den LZB-Bestand, die nicht in die Zugangsmessung mit einflossen und den Korrekturterm ρ quantitativ dominierten. ρ hat seit dieser Revision also einen anderen Bedeutungsgehalt und fängt bei Bedarf nur noch JC-spezifische Umzugssalden auf.

Beim Datenausfall der Bewegungsdaten zum Zeitpunkt *t* liegen zwar die jeweiligen Bestandswerte vor – nötigenfalls imputiert (vgl. Abschnitt 4.1.2 für ELB bzw. 4.3.2 für LZB). Doch drei Größen der Gleichung sind zunächst unbekannt: die Zugänge, die Abgänge sowie das *ρ*.

Zuerst soll es um *ρ* gehen: Analysen zeigen, dass sich *ρ* in der Regel proportional zum Bestand verhält: je höher der Bestand, desto größer ist auch *ρ* und umgekehrt. Zudem ist *ρ* dann quantitativ ausschlaggebend, wenn es über längere Zeit hinweg für ein bestimmtes JC größer oder kleiner als null ist. *ρ* ist zum Beispiel in Bezug auf LZB-Bewegungen dann von Bedeutung, wenn die Umzugssalden über längere Zeit hinweg durchgehend positiv oder negativ sind. Somit erscheint für die Schätzung von *ρ* ein Anteilswertverfahren sinnvoll, das die JC-individuellen Anteilswertinformationen aus der Vergangenheit nutzt. Die Schätzwerte für *ρ* ergeben sich als Produkt aus diesem Anteil und dem aktuellen ELB-Bestand bzw. LZB-Bestand. Zur Absicherung gegen zufällige Schwankungen wird der Anteilswert über die letzten zwölf Monate gemittelt.[123]

$$\hat{\rho}_t^k = \frac{\sum_{m=t-12}^{t-1} \rho_m^k}{\sum_{m=t-12}^{t-1} Bestand_m^k} \cdot Bestand_t^k \qquad (16)$$

Nun soll es um die verbleibenden beiden Unbekannten gehen, also um die Zu- und die Abgänge. Zeitreihen aggregierter Bewegungsdaten weisen ausgeprägte Verlaufsmuster auf. Diese folgen überregionalen saisonalen und konjunkturellen Einflüssen und ähneln sich in wirtschaftsstrukturell ähnlichen Gebieten. Auf Ebene der JC sind diese Muster infolge häufig kleiner Fallzahlen und relativ starker Datensprünge oft nicht mehr verlässlich zu identifizieren und zu modellieren. Deshalb ist es sinnvoll, einen Umweg über die zugehörigen Bewegungsraten zu nehmen. Infolge der Normierung mit dem Bestand sind sie besser geeignet, zeitstabile Niveaueffekte und zeitvariable Trendeffekte zu identifizieren.

Zunächst zu den zeitstabilen Niveaueffekten: Die auf die jeweiligen Bestandsdaten bezogenen Zu- und Abgangsraten weisen im Zeitablauf eine gewisse Stabilität auf, variieren dabei aber deutlich zwischen den JC. Für die Schätzung eines JC-spezifischen Niveaus bei den Zu- bzw. Abgangsraten (nachfolgend Z_{niv}, A_{niv}) wird – wie bei der obigen Schätzung für *ρ* – auf JC-individuelle Informationen aus der Vergangenheit zurückgegriffen. Es wird die Annahme getroffen, dass die jeweils zu schätzende Niveaurate des JC *k* zum Zeitpunkt *t* ähnlich seiner mittleren Bewegungsrate der vergangenen zwölf Monate ist:[124]

$$\hat{Z}_{niv,t}^k = \frac{\sum_{m=t-12}^{t-1} Zugang_m^k}{\sum_{m=t-12}^{t-1} Bestand_m^k} \approx \frac{Zugang_t^k}{Bestand_t^k} \text{ bzw. } \hat{A}_{niv,t}^k = \frac{\sum_{m=t-12}^{t-1} Abgang_m^k}{\sum_{m=t-12}^{t-1} Bestand_{m-1}^k} \approx \frac{Abgang_t^k}{Bestand_{t-1}^k} \qquad (17)$$

Würde man nur mit diesen Raten fortfahren, erhielte man systematische Schätzfehler. Überregionale aktuelle Trendeffekte zum Beispiel infolge von Saison oder Konjunktur blieben unberücksichtigt. Es käme für einzelne Monate zu einem allgemeinen Trend der Unter- oder Überschätzung.

[123] Der Durchschnitt wird über die Division aus Summen in Zähler und Nenner gebildet. Alternativ könnte man den Durchschnitt über die Monatsraten bilden. Die gewählte Variante erwies sich in Tests jedoch als robuster gegenüber Ausreißern, wie sie bei Bewegungsgrößen häufig vorkommen.

[124] Erneut wird aus Gründen der Robustheit der Durchschnitt über die Division aus Summen in Zähler und Nenner gebildet. Für die Zugangsrate wird die Logik der Kennzahlen nach § 48a SGB II verwendet, das heißt der Zugang eines Monats wird dem Bestand desselben Monats gegenübergestellt. In der amtlichen Statistik zur Grundsicherung wird dagegen der Bestand des Vormonates genutzt.

Deshalb nun zu den zeitvariablen Trendeffekten: Wie beim Fortschreibungsverfahren für Datenlücken bei ELB und LZB (vgl. Abschnitt 4.1.2 bzw. 4.3.2.1) wird diese Trendinformation mit Hilfe entsprechender Informationen aller JC im gleichen SGB-II-Vergleichstyp VT_k (vgl. Abschnitt 2.2) gewonnen, dem das JC *k* angehört. Die gesuchte allgemeine aktuelle Zu- bzw. Abgangsrate (nachfolgend Z_{akt}, A_{akt}) ist die mittlere Abweichung der monatsaktuellen Bewegungsrate im Vergleichstyp zur Niveaurate im Vergleichstyp. Letztere wird gleichartig zur Niveaurate des JC als mittlere Bewegungsrate der vergangenen zwölf Monate bestimmt. Aus Gründen der Robustheit wird bei der Mittelung der Median und nicht das arithmetische Mittel verwendet.

$$\hat{Z}^{VT^k}_{akt,t} = median\left\{\frac{Zugang^j_t}{Bestand^j_t} - \hat{Z}^j_{niv,t} \mid j \epsilon VT^k \; und \; j \neq k\right\} \text{ bzw.}$$
$$\hat{A}^{VT^k}_{akt,t} = median\left\{\frac{Abgang^j_t}{Bestand^j_{t-1}} - \hat{A}^j_{niv,t} \mid j \epsilon VT^k \; und \; j \neq k\right\} \quad (18)$$

Die so geschätzte allgemeine aktuelle Trendrate wird auf das betreffende JC übertragen. Die zugrundeliegende Annahme ist, dass die Bewegungsraten von JC desselben SGB-II-Vergleichstyps denselben zeitlichen Trends unterliegen.

Schließlich zu den geschätzten Gesamtraten: Für die Gesamtschätzung der Zu- bzw. Abgangsraten (nachfolgend Z_{ges}, A_{ges}) werden die beiden Elemente – jeweilige Niveaurate und Trendrate für Bewegungen – summarisch verknüpft:

$$\hat{Z}^k_{ges,t} = \hat{Z}^k_{niv,t} + \hat{Z}^{VT^k}_{akt,t} \text{ bzw. } \hat{A}^k_{ges,t} = \hat{A}^k_{niv,t} + \hat{A}^{VT^k}_{akt,t} \quad (19)$$

Somit berücksichtigen die geschätzten Zu- und Abgangsraten mit den Bewegungsraten eines JC aus dem zeitlichen Vorfeld einerseits JC-individuelle Besonderheiten im Niveau (erster Summand) und mit der Entwicklung der Bewegungsraten im zugehörigen SGB-II-Vergleichstyp am aktuellen Rand andererseits auch den allgemeinen aktuellen Trend (zweiter Summand).

Zurück zu den verbleibenden beiden Unbekannten, also den Zu- und Abgängen. Nach den bisherigen Überlegungen gibt es für den Zeitpunkt der Datenlücke zwar Schätzwerte für die Zu- und Abgangsrate und den Korrekturterm *ρ*, aber noch immer keine simultanen Schätzwerte für die Zu- und Abgänge selbst. Aus statistischer Sicht kann man die geforderte Simultanschätzung als Optimierungsproblem auffassen. Optimierungsprobleme werden zum Beispiel mit Hilfe eines Minimierungsansatzes gelöst. Man sucht diejenige simultane Ausprägung von Zu- und Abgängen, bei der der Schätzfehler minimal ist. Dieser lässt sich verschieden messen. Im vorliegenden Fall wird versucht, die Zu- und Abgänge so zu wählen, dass der Abstand der sich daraus ergebenden Bewegungsraten zu den oben geschätzten Bewegungsraten möglichst klein ist. Die Minimierung erfolgt nach dem üblichen Standardkriterium in solchen Fällen: minimiert werden soll die Summe der quadrierten Abweichungen. Das Quadrieren stellt sicher, dass positive

und negative Abweichungen in gleicher Weise einfließen.[125] Die Stock-Flow-Gleichung fungiert als sogenannte mathematische Nebenbedingung:

$$min\left(\hat{Z}^k_{ges,t} - \frac{Zugang^k_t}{Bestand^k_t}\right)^2 + \left(\hat{A}^k_{ges,t} - \frac{Abgang^k_t}{Bestand^k_{t-1}}\right)^2 \qquad (20)$$

$$u.d.Nb.: Bestand^k_t = Bestand^k_{t-1} - Abgang^k_t + Zugang^k_t + \rho^k_t$$

Die vorliegende Notation mag es zwar ein wenig verschleiern, doch kommen die im vorangegangenen Zwischenschritt geschätzten Bewegungsraten nun ähnlich wie Fortschreibungsfaktoren zum Einsatz. Insofern kann das Optimierungsmodell als Fortschreibungsverfahren interpretiert werden – egal welche mathematischen Umformungen und modellverfeinernden Schritte noch folgen.

Sobald für eine der beiden Bewegungsarten ein Imputationswert vorliegt, wird der Wert für die andere Bewegungsart simultan mitgeliefert. Über welche Größe man die Schätzung definiert – über Zugänge oder Abgänge – hat keinen Einfluss. Hier werden die Zugänge immer in Abhängigkeit von den Abgängen definiert.[126] Nach einigen mathematischen Umformungen lautet die Optimierungsfunktion:[127]

$$opt.Abgang^k_t = \left(\frac{\frac{\hat{A}^k_{ges,t}}{Bestand^k_{t-1}} + \frac{\hat{Z}^k_{ges,t} \cdot Bestand^k_t + Bestand^k_{t-1} - Bestand^k_t + \hat{\rho}^k_t}{\left(Bestand^k_t\right)^2}}{\frac{1}{\left(Bestand^k_t\right)^2} + \frac{1}{\left(Bestand^k_{t-1}\right)^2}}\right) \qquad (21)$$

Nun zu unteren Schranken für die Definition von Mindestfallzahlen: Würde man es bei dem bisher abgeleiteten Modell belassen, erhielte man vereinzelt ungewöhnlich niedrige oder gar negative Werte bei der Schätzung der Zugangs- und Abgangszahlen. Das Phänomen tritt vor allem im Falle hoher Bestandsveränderungen auf. Mathematisch ist ein solches Ergebnis korrekt, doch es entspricht nicht der Zählweise von Bewegungen und muss deswegen aus dem Modell ausgeschlossen werden. Dies wird durch Mindestfallzahlen sowohl für Ab- als auch für Zugänge realisiert.

Erneut wird der Weg über die jeweiligen Bewegungsraten genommen. Im ersten Schritt werden für JC, für die im betreffenden Berichtsmonat keine Datenlücken bei den Zu- und Abgangsraten vorliegen, diese Raten normiert, indem sie durch die jeweiligen Gesamtraten $\hat{Z}^k_{ges,t}$ bzw. $\hat{A}^k_{ges,t}$ dividiert werden. Die Normierung soll für den nachfolgenden Schritt sicherstellen, dass die Daten über die JC annähernd der gleichen Verteilung entstammen. Sie erfolgt daher im Sinne ausreichend hoher Fallzahlen über alle JC hinweg und nicht getrennt nach SGB-II-Vergleichstyp. Im zweiten Schritt werden aus den jeweiligen normierten

[125] Der Umstand, dass auf den Bestand normierte Zu- und Abgangsraten statt der jeweiligen Absolutwerte herangezogen werden, sorgt für eine Gleichbehandlung kleiner JC mit den großen, was im vorliegenden Anwendungsfall zweckmäßig ist.

[126] Hier werden zuerst die Abgänge geschätzt, weil sie weniger von den oben beschriebenen nicht gezählten Bewegungen durch Merkmalsausprägungen (ρ) betroffen sind.

[127] Um für eine Funktion ein Minimum zu finden, werden die erste und die zweite Ableitung gebildet. Im Minimum ist die erste Ableitung gleich Null, die zweite größer Null. Die ausführliche formal-mathematische Herleitung wird auf Anfrage zur Verfügung gestellt.

Verteilungen die Grenzen für Ausreißer bestimmt. Als Ausreißer gelten – in Anlehnung an die Boxplotmethode (vgl. Anhang 8.1, Lesehilfe zu Boxplots) – jene Punkte, die von den Quartilen der Verteilung weiter als der 1,5-fache Interquartilsabstand entfernt sind. Da es hier um die Bestimmung von Mindestfallzahlen geht, genügt es, die jeweilige Untergrenze für Ausreißer zu betrachten. Im dritten Schritt wird diese normierte untere Rate auf die von Datenlücken betroffenen JC übertragen: ihre jeweilige JC-individuelle Mindestfallzahl für den Berichtsmonat wird ermittelt, indem die Normierung der Bewegungsraten wieder aufgehoben wird und die jeweiligen Ergebnisse mit den dazugehörigen Bestandswerten multipliziert werden.

$$Min.Zugang_t^k = norm.unt.Zug.rate_t \cdot \hat{Z}_{ges,t}^k \cdot Bestand_t^k \text{ bzw.}$$
$$Min.Abgang_t^k = norm.unt.Abg.rate_t \cdot \hat{A}_{ges,t}^k \cdot Bestand_{t-1}^k \quad (22)$$

Schließlich: Simultane Definition von Mindestfallzahlen: Aufgrund der Simultanschätzung und der wechselseitigen Abhängigkeit von Zu- und Abgängen muss die Mindestfallzahl einer Größe jene der anderen Größe berücksichtigen. Es kann also vorkommen, dass zum Beispiel die Mindestfallzahl für Abgänge, wenn man sie nach der erweiterten Stock-Flow-Gleichung (15) ermittelt, von der mit Formel (22) berechneten Mindestfallzahl abweicht:

$$Min2.Abgang_t^k = norm.unt.Zug.rate_t \cdot \hat{Z}_{ges,t}^k \cdot Bestand_t^k + Bestand_{t-1}^k - Bestand_t^k + \hat{\rho}_t^k \quad (23)$$

Beide Fallkonstellationen müssen berücksichtigt werden. Die Regelungen zu Mindestfallzahlen müssen mit einer zweiten Absicherung versehen werden: der für die Datenimputation endgültig zu verwendende Schätzwert muss mindestens so groß sein wie die größere der beiden Mindestfallzahlen.

$$endgültigerAbgang_t^k = max\left(opt.Abgang_t^k; Min.Abgang_t^k; Min2.Abgang_t^k\right) \quad (24)$$

Der endgültige Imputationswert der zweiten Bewegungsart (hier die Zugänge) wird mittels der erweiterten Stock-Flow-Gleichung simultan mitgeliefert.

$$endgültigerZugang_t^k = Bestand_t^k - Bestand_{t-1}^k + endgültigerAbgang_t^k - \hat{\rho}_t^k \quad (25)$$

8.5 Fortschreibung auf Basis des k-nächste-Nachbarn-Verfahrens am Beispielthema Integrationen

Das Datenimputationsverfahren, das unter anderem für das Thema Integrationen von ELB angewendet wird (vgl. Abschnitt 6.1.2), beinhaltet ein spezielles Fortschreibungsverfahren, das auf einem k-nächste-Nachbarn-Verfahren beruht. Damit wird die Entwicklung der Integrationszahl ähnlicher Nachbarn auf das betreffende JC übertragen. Der Schätzwert ergibt sich als Produkt des Vormonatswertes und der geschätzten mittleren Veränderung zum Vormonat der k ähnlichsten JC.

Ähnlichkeit drückt sich über ähnliche saisonale Verlaufsmuster bei der Zahl der Integrationen aus. Sie wird anhand der monatlichen Veränderungsraten der Integrationszahlen im zeitlichen Umfeld zu demjenigen Monat gemessen, für den eine Imputation erforderlich ist. So werden latente regionalspezifische Entwicklungsmuster eingefangen, die JC-übergreifend in der Zeitreihe zum Tragen kommen.

Die Datensituation stellt sich wie folgt dar: Grundlage ist ein rechteckiger Datensatz *X* mit JC in den Zeilen (*j*) und Zeitpunkten in den Spalten (*s*):

$$X = \left(x_{js}\right)_{\substack{j=1,\dots,n \\ s=t-m,\dots t}} = \begin{bmatrix} x_{1(t-m)} & \cdots & x_{1t} \\ \vdots & \ddots & \vdots \\ x_{n(t-m)} & \cdots & x_{nt} \end{bmatrix} \tag{26}$$

Die Zahl der Integrationen für JC *j* zum Zeitpunkt *s* liegt also bei x_{js}. Um für ein JC *j* die nächsten *k* Nachbarn zu finden, werden Informationen über die relativen Veränderungen der Integrationen zum Vormonat benötigt. Hierfür wird die Datenmatrix *X* transformiert, sodass die neue Datenmatrix X_V die jeweiligen Vormonatsänderungen enthält. Die Transformationsregel lautet:

$$X_V = \left(v_{js}\right)_{\substack{j=1,\dots,n \\ s=t-m+1,\dots t}} := \left(\frac{x_{js}}{x_{j(s-1)}}\right)_{\substack{j=1,\dots,n \\ s=t-m+1,\dots t}} \tag{27}$$

Der Nachteil dieser Transformation ist, dass ein Datenausfall in den ursprünglichen Daten *X* zwei Datenausfälle in der Veränderungsmatrix X_V nach sich zieht, zum einen im selben Berichtsmonat, zum anderen im Nachfolgemonat. Vorteilhafte Effekte der Transformation liegen in der einheitlichen Skalierung und Vergleichbarkeit von JC, sodass man in der Schätzung unabhängig vom Absolutniveau der Integrationen je JC wird. Folglich werden kleine und große JC gleichartig behandelt.

In der konkreten Situation gebe es einen fehlenden Wert für JC *j* zum Zeitpunkt *s*. Nachfolgend wird das **Vorgehen zur Schätzung** Schritt für Schritt vorgestellt.

1. Für den nachfolgenden Ähnlichkeitsabgleich nutzt man aus der Veränderungsmatrix X_V nur das zeitliche Umfeld vor dem Monat, für den eine Imputation erforderlich ist: es werden die Vormonatsveränderungen der drei letzten verfügbaren Zeitpunkte und des Vorjahresmonats herangezogen.[128]
2. Vom betreffenden JC wird zu jedem JC, das zum Zeitpunkt *s* in der Veränderungsmatrix keinen fehlenden Wert hat, der quadratische Abstand ermittelt Das Quadrieren stellt sicher, dass positive und negative Abweichungen in gleicher Weise einfließen.[129]
3. Auf dieser Basis werden die *k* nächsten Nachbarn mit dem jeweils kleinsten quadratischen Abstand von *j* ausgewählt.
4. Nach dem vorangegangenen Ähnlichkeitsabgleich wird der Mittelwert der Vormonatsveränderung im Monat *s* für die *k* Nachbarn $\hat{v}_{js}^{k}$ bestimmt.
5. Der Schätzwert bestimmt sich im einfachen Dreisatz als Produkt des Vormonatswertes und der mittleren Vormonatsveränderung aus Schritt 4:

$$\hat{x}_{js}^{k} \coloneqq \hat{v}_{js}^{k} \cdot x_{j(s-1)} \tag{28}$$

Falls der Vormonatswert $x_{j(s-1)}$ nicht vorliegt, wird stattdessen der Schätzwert verwendet – sowohl in der Fortschreibung als auch im vorangegangenen Ähnlichkeitsabgleich. Letzteres bedeutet, dass Nachbarn auch auf Basis von imputierten Werten ermittelt werden. Treten beim selben JC mehrere Ausfälle nacheinander auf, steigt also die Schätzunsicherheit, sowohl für den Vorgang der Fortschreibung als auch für die Suche nach Nachbarn. Bei den aktuell vorliegenden Ausfallmustern mit seltenen Mehrfachausfällen ist das Vertrauen in die festgeschriebenen Imputationswerte in vorangegangenen Berichtsmonaten auch gerechtfertigt und nur konsequent.[130] Den Ähnlichkeitsabgleich, also die Suche nach Nachbarn, könnte

[128] Es werden hier keine Berichtsmonate herangezogen, die nach dem Monat folgen, für den eine Imputation erforderlich ist (sogenannte „Leads“). So muss man nicht auf festgeschriebene Lead-Werte mit dreimonatiger Wartezeit warten und damit auch nicht auf endgültige imputierte Werte. Die Berücksichtigung von Lead-Werten war im Simulationstest ohnehin nicht wesentlich überlegen und wurde verworfen. Auch das Heranziehen von mehr als drei Vormonaten und dem Vorjahresmonat bietet keine Vorteile. Mehr Beobachtungen würden die Dimensionen und das Volumen des mathematischen Raums stark ansteigen lassen, doch wäre der Gewinn jeder weiteren Dimension marginal, im ungünstigen Fall würde die Trennschärfe sogar verschlechtert werden – ein Phänomen, das in der Fachwelt als „Fluch der Dimensionalität“ bezeichnet wird. Drei Vormonate und der Vorjahresmonat bieten nach unserem Erachten eine treffende Momentaufnahme für die Entwicklung und Saisonalität des jeweiligen JC zum jeweiligen Beobachtungszeitpunkt, um ähnliche JC zu identifizieren.

[129] Wenn man die Wurzel aus dem quadratischen Abstand ziehen würde, erhielte man den euklidischen Abstand – ein weiteres gebräuchliches Distanzmaß. Diese mathematische Umformung würde hier dieselben Ergebnisse erzeugen.

[130] Analysen der Wirkung von Mehrfachausfällen zeigen, dass der Schätzfehler erwartungsgemäß steigt. Wenn sich aber die bisher beobachteten Ausfallmuster nicht dramatisch verschlechtern, ist von akzeptablen Unsicherheiten auszugehen.

man stattdessen auch auf die tatsächlich vorliegenden Ursprungsdaten beschränken. Doch fiele durch die Verwendung der mittleren Entwicklung ähnlicher Nachbarn die Auswahl ohnehin weitgehend auf dieselben JC. Außerdem wird eine Fortschreibung vom Niveaueffekt, also dem Ausgangsniveau des Vormonats, dominiert. Insofern wäre die Beschränkung auf tatsächlich vorhandene Ursprungswerte annähernd ergebnisgleich. Beruht der Ähnlichkeitsabgleich hingegen gleichberechtigt auf Original- und Imputationswerten, besteht zudem die Chance, Sondereffekte, die das betroffene JC mit seinen Nachbarn teilt, in jedem einzelnen Berichtsmonat mitzeichnen zu können, sogar, wenn Ausfälle und vorhandene Monatswerte im direkten Wechsel miteinander auftreten.

Da die Zeitpunkte aufsteigend abgearbeitet werden, liefert das Verfahren überall ein Ergebnis für nicht gefüllte Zellen. Dem Datennutzer müssen Imputationswerte ab Berichterstattungsbeginn Januar 2011 verfügbar gemacht werden. Wegen des Ähnlichkeitsabgleichs gemäß Schritt 1 bis 3 erfolgen die ersten Schätzungen bereits ab dem zweiten Berichtsmonat in der Datenmatrix, also ab Oktober 2010.[131] Sie werden bis einschließlich Dezember 2010 jedoch nur temporär vorgehalten und dann verworfen. Da in den ersten beiden Berichtsmonaten keine Datenlücken vorliegen, ist glücklicherweise ein einheitlicher Zeitreihenstart für alle JC gewährleistet.

Die Grundfrage bei der Auswahl der Nachbarn lautet: wie viele Nachbarn sind optimal? Man könnte zwar die singuläre Veränderungsinformation von nur einem einzigen nächsten Nachbarn nutzen. Doch man wäre unsicher, ob nicht ein starker Ausreißer und damit ungeeigneter Fall für die Fortschreibung verwendet würde. Im anderen Extrem könnte man die durchschnittliche Veränderungsinformation von allen anderen, also rund 400 JC, nutzen. Doch wäre das JC, für das man einen Schätzwert sucht, von sehr vielen dieser JC sehr weit entfernt, also nicht wirklich ein Nachbar. In der Folge würde man die Fortschreibung zu stark nivellieren. Als Kompromiss könnte man zum Beispiel immer die drei nächsten Nachbarn für die Fortschreibung heranziehen – intuitiv eine bessere Variante als die beiden zuvor beschriebenen extremen Varianten.

Aus statistischer Sicht kann man die obige Grundfrage als Optimierungsproblem auffassen. Optimierungsprobleme werden zum Beispiel mit Hilfe eines Minimierungsansatzes gelöst. Man sucht diejenige Anzahl *k* der nächsten Nachbarn, bei der der gesamte Schätzfehler minimal ist. Dieser lässt sich verschieden messen. Hier fällt die Wahl auf das übliche Standardkriterium in solchen Fällen, den mittleren quadratischen Fehler bezogen auf die relativen Abweichungen bei Anwendung des Verfahrens. Es handelt sich also um den Mittelwert aus den aufsummierten quadrierten Abweichungen zwischen simulierten Schätzwerten und dem tatsächlichen Ursprungswerten. Wieder stellt das Quadrieren sicher, dass positive und negative Abweichungen in gleicher Weise einfließen.[132]

Zu lösen ist das Optimierungsproblem mit Hilfe einer Simulation: die JC werden einzeln und nacheinander als fehlend behandelt und stattdessen Schätzwerte ermittelt wie in Schritt 1 bis 5 beschrieben. Dabei

[131] Wie in Abschnitt 6.1.1 erläutert, beginnt die Zeitreihe bei Integrationen für ELB mit September 2010.

[132] Das gewählte Minimierungskriterium des mittleren quadratischen Fehlers bringt als Standardinstrument eine hoch erwünschte Eigenschaft mit sich: Minimiert wird der Schätzfehler sowohl in der mittleren Tendenz als auch in seiner Streuungsbreite, sodass die Zahl und die Intensität insbesondere von großen Schätzfehlern (Ausreißerfälle) verringert wird. Zudem sorgt der Umstand, dass auch hier wieder relative Schätzfehler herangezogen werden, für eine Gleichbehandlung kleiner JC mit den großen, was im vorliegenden Anwendungsfall zweckmäßig ist.

verwendet man zuerst nur einen nächsten Nachbarn, dann zwei, dann drei usw. bis rund 400. Dies geschieht berichtsmonatsweise. In Berichtsmonaten, in denen sich alle JC sehr ähnlich entwickeln, weil zum Beispiel alle gleichermaßen in einer winterlichen Saisondelle verharren, genügt eine kleinere Anzahl *k* an nächsten Nachbarn, um im Mittel das typische Entwicklungsmuster zu beschreiben. In Berichtsmonaten, in denen die Entwicklung sehr uneinheitlich ist, wird eine größere Anzahl *k* an nächsten Nachbarn benötigt, um die Unsicherheit in der Schätzung aufzufangen. Im Rahmen der Simulation wird für jeden Berichtsmonat jenes *k* als optimal ausgewählt, das den mittleren quadratischen Fehler minimiert. Es gibt demnach genauso viele Simulationsdurchgänge, wie tatsächlich Daten bei den JC in diesem Monat vorliegen. Da die Anzahl *k* der nächsten Nachbarn monatsweise optimiert wird, liefert das Verfahren endgültige Ergebnisse, die festgeschrieben werden können.

8.6 Modellauswahl und -vergleich für ein regressionsbasiertes Schätzverfahren bei Integrationen von Alleinerziehenden[133]

Für „Integrationen von Alleinerziehenden“ und „Integrationen von LZB“ wird als Datenimputationsverfahren ein regressionsbasiertes Schätzverfahren angewendet. Dieser Anhang beschreibt die Modellauswahl für die Integrationen von Alleinerziehenden (vgl. Abschnitt 6.2.2). Je nach Verfügbarkeit gehen WZ2-, WZ1-, WZ0- oder Vormonatsinformationen mit weiteren Größen in das Regressionsmodell ein. Als Auswahlkriterien für die Spezifikation werden der Median des betragsmäßigen prozentualen Fehlers und die Anzahl von JC mit großen Fehlern über alle Berichtsmonate von Februar 2011 bis Juli 2013 (12.251 Beobachtungszeitpunkte) herangezogen.[134] Der jeweilige Fehler errechnet sich, indem der regressionsbasierte Schätzwert ermittelt und mit dem tatsächlichen Beobachtungswert abgeglichen wird. Bei über der Hälfte der Beobachtungen (54,9 %) liegt die Zahl der Integrationen Alleinerziehender bei 20 und weniger Fällen.[135] Geringe Fallzahlen ergeben sich insbesondere für die JC in Bayern, Rheinland-Pfalz und teilweise auch in Thüringen. Hier können kleine absolute Abweichungen bereits zu vergleichsweise hohen relativen Fehlern führen. Deswegen wurde zur Bestimmung großer Fehler eine Kombination von absoluter und relativer Abweichung gewählt: Ein großer Fehler liegt vor, wenn die betragsmäßige relative Abweichung 20 % oder mehr beträgt und dabei einen absoluten betragsmäßigen Fehler von fünf oder mehr Fällen aufweist.

Dem Grundmodell mit den Integrationen Alleinerziehender mit Informationen aus verkürzter Wartezeit werden nach und nach der Grad der Untererfassung der Integrationen insgesamt, der Vormonatswert der alleinerziehenden ELB und die Abgänge Arbeitsloser im Rechtskreis SGB II zugefügt. Aus diesen Modellen wird die Kombination mit dem niedrigsten Median und der niedrigsten Zahl großer Fehler für die weitere Testung der Variablen Bundesland und SGB-II-Vergleichstyp gewählt. Beide werden jeweils einmal separat und schließlich beide zusammen im Modell geschätzt. Dieses Verfahren wird für alle Wartezeitstände durchgeführt.

<u>Prädiktoren/Modelle</u>

A: Information mit verkürzter bzw. ohne Wartezeit (WZ2, WZ1 bzw. WZ0) bzw. Vormonatswerte der Integrationen Alleinerziehender $(int_allein)_{t-1}$

B: Modell A und Nacherfassungsquote der Integrationen insgesamt (NEQ)

---

[133] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[134] Zur Modellspezifikation werden häufig eine stufenweise Regression oder Informationskriterien wie das AIC (Akaike Information Criterion) herangezogen. Prinzipiell sollen dabei die Ausprägungen eines Merkmals durch eine möglichst geringe Anzahl anderer Merkmale bzw. deren Ausprägungen erklärt werden. Im vorliegenden Anwendungsfall interessiert weniger, wie gut einzelne Koeffizienten geschätzt werden. Das Verfahren wird hier als prädiktives Modell eingesetzt, um im Einzelfall Schätzwerte möglichst nah an den Ursprungswerten zu erhalten, so dass der mittlere Fehler sowie die Anzahl großer Fehler maßgeblich sind.

[135] Auch wenn JC mit kleinen Fallzahlen deutlich überwiegen, weisen einzelne große JC Werte von mehreren Hundert Integrationen Alleinerziehender je Berichtsmonat auf. Da hieraus resultierende Größeneffekte die Schätzung dominieren können, wurde ein gewichtetes Kleinste-Quadrate-Verfahren angewendet (Weighted Least Square-Methode, WLS). Dabei wird für jedes JC der Kehrwert der Varianz der Integrationen Alleinerziehender in die Schätzung einbezogen.

C1: Vormonatswert der alleinerziehenden ELB $(elb_allein)_{t-1}$

C2: Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II (alo_abg)

C3: Kombination $(elb_allein)_{t-1}$ und alo_abg

C$: Auswahl aus C1, C2 und C3

D1: C$ + Bundesland (Land)

D2: C$ + SGB-II-Vergleichstyp 2011 (VT)

D3: C$ + Land und VT

Für den Fall, dass keine Informationen aus verkürzter Wartezeit vorliegen, wird als weitere Niveauvariable der Vormonatswert der Integrationen Alleinerziehender einbezogen. Abschließend werden die Ergebnisse des regressionsbasierten Schätzverfahrens denen des Ersatzwertverfahrens (zweite Entwicklungsstufe der Imputationsverfahren) gegenübergestellt.

Wartezeit 2

Daten mit einer Wartezeit von zwei Monaten liegen bereits nah an den endgültigen Werten mit dreimonatiger Wartezeit, so dass die weiteren Größen keinen so starken Beitrag zur Verringerung der Fehler liefern, wie im Fall der Schätzung auf Basis der Daten mit Wartezeit von einem Monat bzw. jener ohne Wartezeit. In den nachfolgenden Tabellen werden die Modelle bzw. Komponenten, die nicht einbezogen werden, grau hinterlegt. Das Modell C2 zeigt im Ergebnis, dass die Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II den mittleren Fehler zwar im Dezimalbereich verringern. Insgesamt lässt sich jedoch für die Schätzungen auf Grundlage der Daten mit verkürzter Wartezeit feststellen, dass die Abgänge aus Arbeitslosigkeit kaum zur Verringerung der Fehler beitragen und daher nicht weiter berücksichtigt werden. Für den Wartezeitstand von zwei Monaten führt die Kombination aus Bundesland und Vergleichstyp zum gleichen Ergebnis wie das Modell, in das nur das Bundesland eingeht. Da die Kombination in den anderen Wartezeitständen die Fehler erkennbar reduziert, wird im Sinne eines möglichst einheitlichen Vorgehens über alle Wartezeitstände hinweg Modell D3 gewählt.

Tabelle 9: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen mit Daten mit zwei Monaten Wartezeit

WZ2	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	WZ2	2,9	33
B	WZ2 + NEQ	2,9	18
C1	WZ2 + NEQ + $(elb_allein)_{t-1}$	3,0	15
C2	WZ2 + NEQ + alo_abg	2,9	17
C3	WZ2 + NEQ + $(elb_allein)_{t-1}$ + alo_abg	3,0	15
D1	WZ2 + NEQ + $(elb_allein)_{t-1}$ + Land	2,9	14
D2	WZ2 + NEQ + $(elb_allein)_{t-1}$ + VT	2,9	15
D3	WZ2 + NEQ + $(elb_allein)_{t-1}$ + Land + VT	2,9	14

Wartezeit 1

Im Fall der Regressionsrechnung mit Daten mit einer Wartezeit von einem Monat zeigt sich noch deutlicher, dass Abgänge Arbeitsloser im Rechtskreis SGB II nicht zu einer Reduzierung der Fehler beitragen.

Dagegen führt die Berücksichtigung von Bundesland und SGB-II-Vergleichstyp zu einer Fehlerreduzierung, so dass im Gesamtmodell lediglich die Abgänge aus Arbeitslosigkeit nicht berücksichtigt werden.

Tabelle 10: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen mit Daten mit einem Monat Wartezeit

WZ1	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	WZ1	7,3	168
B	WZ1 + NEQ	5,9	105
C1	WZ1 + NEQ + $(elb_allein)_{t-1}$	5,7	91
C2	WZ1 + NEQ + alo_abg	5,8	95
C3	WZ1 + NEQ + $(elb_allein)_{t-1}$ + alo_abg	5,7	93
D1	WZ1 + NEQ + $(elb_allein)_{t-1}$ + Land	5,6	80
D2	WZ1 + NEQ + $(elb_allein)_{t-1}$ + VT	5,6	78
D3	WZ1 + NEQ + $(elb_allein)_{t-1}$ + Land + VT	5,5	68

Daten ohne Wartezeit

Bei der regressionsbasierten Schätzung mit Daten ohne Wartezeit tragen die Abgänge aus Arbeitslosigkeit im Rechtskreis SGB II ebenfalls nicht zu einer Fehlerreduzierung bei. Auch hier führt die Kombination aus Bundesland und SGB-II-Vergleichstyp zu den besten Ergebnissen.

Tabelle 11: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen mit Daten ohne Wartezeit

WZ0	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	WZ0	34,9	6.632
B	WZ0 + NEQ	12,3	1.240
C1	WZ0 + NEQ + $(elb_allein)_{t-1}$	10,5	753
C2	WZ0 + NEQ + alo_abg	10,9	887
C3	WZ0 + NEQ + $(elb_allein)_{t-1}$ + alo_abg	10,5	756
D1	WZ0 + NEQ + $(elb_allein)_{t-1}$ + Land	9,9	638
D2	WZ0 + NEQ + $(elb_allein)_{t-1}$ + VT	10,2	676
D3	WZ0 + NEQ + $(elb_allein)_{t-1}$ + Land + VT	9,7	560

Ohne Informationen mit geringerer bzw. ohne Wartezeit

Wenn keine Daten mit verkürzter Wartezeit vorliegen, wird als weitere Niveauvariable der Vormonatswert der Integrationen Alleinerziehender aufgenommen. Hier leisten die Abgänge aus Arbeitslosigkeit einen Beitrag zur Verringerung der Schätzfehler. Es ist dabei erkennbar, dass der mittlere Fehler sowie die Anzahl großer Fehler höher ausfallen, als bei der regressionsbasierten Schätzung mit Daten aus verkürzter Wartezeit. Gleichwohl zeigt die in Tabelle 13 abgebildete Gegenüberstellung, dass das Regressionsmodell zu besseren Ergebnissen führt als die ersatzweise Verwendung des Vormonatswertes der Integrationen Alleinerziehender (zweite Entwicklungsstufe).

Tabelle 12: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen, wenn keine Daten mit verkürzter Wartezeit vorliegen

VM	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	$(int_allein)_{t-1}$	22,3	4.360
B1	$(int_allein)_{t-1}$ + $(elb_allein)_{t-1}$	19,2	3.464
B2	$(int_allein)_{t-1}$ + alo_abg	19,5	3.616
B3	$(int_allein)_{t-1}$ + $(elb_allein)_{t-1}$ + alo_abg	19,2	3.440
C1	$(int_allein)_{t-1}$ + $(elb_allein)_{t-1}$ + alo_abg + Land	17,8	3.044
C2	$(int_allein)_{t-1}$ + $(elb_allein)_{t-1}$ + alo_abg + VT	18,1	3.137
C3	$(int_allein)_{t-1}$ + $(elb_allein)_{t-1}$ + alo_abg + Land + VT	17,1	2.827

Vergleich zwischen Ersatzwertverfahren und regressionsbasiertem Schätzverfahren

In der Gegenüberstellung erweist sich das regressionsbasierte Schätzverfahren (dritte Stufe) dem Ersatzwertverfahren (zweite Stufe) deutlich überlegen. Lediglich auf Basis von WZ2-Daten ist der mittlere Fehler beim Ersatzwertverfahren geringer – jedoch bei einer höheren Anzahl großer Fehler.

Tabelle 13: Modellvergleich – Integrationen von Alleinerziehenden: Median und große Fehler bei Schätzungen: Ersatzwertverfahren (Stufe 2) versus regressionsbasiertes Schätzverfahren (Stufe 3)

Vergleich zwischen Ersatzwertverfahren und regressionsbasiertem Schätzverfahren		Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
Ersatzwertverfahren (Stufe 2): Ersatzwert auf Basis des…			
1.	WZ2-Wertes	2,3	46
2.	angepassten WZ1-Wertes	5,5	92
3.	angepassten WZ0-Wertes	12,6	1.525
4.	Vormonatswertes $(int_allein)_{t-1}$	26,5	5.589
Regressionsbasierte Schätzung (Stufe 3) mit Regressoren...			
1.	WZ2 + NEQ + $(elb_allein)_{t-1}$ + Land + VT	2,9	14
2.	WZ1 + NEQ + $(elb_allein)_{t-1}$ + Land + VT	5,5	68
3.	WZ0 + NEQ + $(elb_allein)_{t-1}$ + Land + VT	9,7	560
4.	$(int_allein)_{t-1}$ + $(elb_allein)_{t-1}$ + alo_abg + Land + VT	17,1	2.827

8.7 Modellauswahl und -vergleich für ein regressionsbasiertes Schätzverfahren bei Integrationen von LZB[136]

Für „Integrationen von Alleinerziehenden“ und „Integrationen von LZB“ wird als Datenimputationsverfahren ein regressionsbasiertes Schätzverfahren angewendet. Dieser Anhang beschreibt die Modellauswahl für die Integrationen von LZB (vgl. Abschnitt 6.3.2.1). Je nach Verfügbarkeit gehen WZ2-, WZ1-, WZ0- oder Vormonatsinformationen mit weiteren Größen in das Regressionsmodell ein. Als Auswahlkriterien für die Spezifikation werden der Median des betragsmäßigen prozentualen Fehlers und die Anzahl von JC mit großen Fehlern über alle Berichtsmonate von Februar 2011 bis Juli 2013 (12.228 Beobachtungszeitpunkte) herangezogen.[137] Der jeweilige Fehler errechnet sich, indem der regressionsbasierte Schätzwert ermittelt und mit dem tatsächlichen Beobachtungswert abgeglichen wird. Bei rund einem Fünftel der Beobachtungen (23,2 %) liegt die Zahl der Integrationen von LZB bei 30 und weniger Fällen.[138] Geringe Fallzahlen ergeben sich insbesondere für die JC in Bayern, Baden-Württemberg und teilweise auch in Rheinland-Pfalz. Hier können kleine absolute Abweichungen bereits zu vergleichsweise hohen relativen Fehlern führen. Deswegen wurde zur Bestimmung großer Fehler eine Kombination von absoluter und relativer Abweichung gewählt: Ein großer Fehler liegt vor, wenn die betragsmäßige relative Abweichung 20 % oder mehr beträgt und dabei einen absoluten betragsmäßigen Fehler von mehr als zwei Fällen aufweist.

Dem Grundmodell mit den Integrationen von LZB mit Informationen aus verkürzter Wartezeit werden nach und nach der Grad der Untererfassung der Integrationen insgesamt sowie alternativ der Vormonatswert der Integrationen von LZB und die Abgänge Arbeitsloser im Rechtskreis SGB II zugefügt. Aus diesen Modellen wird die Kombination mit dem niedrigsten Median und der niedrigsten Zahl großer Fehler für die weitere Testung der Variablen Bundesland und SGB-II-Vergleichstyp gewählt. Beide werden jeweils einmal separat und schließlich beide zusammen im Modell geschätzt. Dieses Verfahren wird für alle Wartezeitstände durchgeführt. In den nachfolgenden Tabellen wird das Modell, das letztlich verwendet wird, fett gedruckt.

[136] Die Analyseergebnisse dieses Abschnitts beruhen auf Daten vor der Revision der Grundsicherungsstatistik SGB II (April 2016). Die Analyse wurde nicht anhand von revidierten, verlängerten Zeitreihen aktualisiert. Es ist nicht anzunehmen, dass sich die Ergebnisse grundlegend ändern würden (vgl. Abschnitt 2.1.1).

[137] Wie bereits im vorangehenden Anhangskapitel 8.6 benannt, interessiert im vorliegenden Anwendungsfall weniger, wie gut einzelne Koeffizienten geschätzt werden. Das Verfahren wird hier als prädiktives Modell eingesetzt, um im Einzelfall Schätzwerte möglichst nah an den Ursprungswerten zu erhalten, so dass der mittlere Fehler sowie die Anzahl großer Fehler maßgeblich sind.

[138] Auch wenn JC mit kleinen Fallzahlen deutlich überwiegen, weisen einzelne große JC Werte von mehreren Hundert Integrationen von LZB je Berichtsmonat auf. Da hieraus resultierende Größeneffekte die Schätzung dominieren können, wurde ein gewichtetes Kleinste-Quadrate-Verfahren angewendet (Weighted Least Square-Methode, WLS). Dabei wird für jedes JC der Kehrwert der Varianz der Integrationen von LZB in die Schätzung einbezogen.

Prädiktoren/Modelle

A: Information mit verkürzter bzw. ohne Wartezeit (WZ2, WZ1 bzw. WZ0)

B: Modell A und Nacherfassungsquote der Integrationen insgesamt (NEQ)

C1: B+ Vormonatswert der Integrationen von LZB $(int_lzb)_{t-1}$

C2: B+ Abgänge Arbeitsloser im SGB II (alo_abg)

C3: B+ Kombination $(int_lzb)_{t-1}$ und (alo_abg)

C$: Auswahl aus C1, C2 und C3

D1: C$ + Bundesland (Land)

D2: C$ + SGB-II-Vergleichstyp 2011 (VT)

D3: C$ + Land und VT

E: alle Variablen

Wartezeit 2

Abgänge aus Arbeitslosigkeit im SGB II und die SGB-II-Typisierung üben fast keinen Einfluss auf die Integrationen von LZB aus. Im Median den niedrigsten Fehler hat das Modell mit allen anderen getesteten Variablen. Die Variable Bundesland und Vergleichstyp verändern die Zahl großer Fehler und den mittleren Fehler nicht. JC mit kleinen Fallzahlen bei Integrationen LZB konzentrieren sich in Baden-Württemberg und vor allem Bayern sowie in den Vergleichstypen neun und zehn. Hohe Fehler blenden hohe relative Abweichungen bei kleinen absoluten Verschätzungen aus, damit beeinflussen Land und Vergleichstypen zwar den mittleren Fehler, aber nicht mehr die Zahl großer Fehler. Wenn der t-2-Wert des ausgefallenen JC plausibel war, wird der Ersatzwert aus der Nacherfassungsquote der Integrationen insgesamt, dem Vormonatswert der Integrationen LZB sowie dem Bundesland ermittelt (Modell D1).

Tabelle 14: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen mit Daten mit zwei Monaten Wartezeit

WZ2	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	WZ2	1,69	30
B	WZ2 + NEQ	1,54	21
C1	WZ2 + NEQ + $(int_lzb)_{t-1}$	1,54	18
C2	WZ2 + NEQ + alo_abg	1,54	21
C3	WZ2 + NEQ + $(int_lzb)_{t-1}$ + alo_abg	1,55	17
D1	**WZ2 + NEQ + $(int_lzb)_{t-1}$ + Land**	**1,47**	**17**
D2	WZ2 + NEQ + $(int_lzb)_{t-1}$ + VT	1,50	17
D3	WZ2 + NEQ + $(int_lzb)_{t-1}$ + Land + VT	1,47	17
E	Alle Variablen	1,47	17

Wartezeit 1

Der Erklärungsbeitrag der Integrationen LZB mit einem Monat Wartezeit fällt niedriger als bei zwei Monaten aus, damit erhalten die anderen Variablen ein höheres Gewicht in der Schätzung und verbessern sie. Den besten Erklärungsgehalt hat das Modell mit allen getesteten Variablen (Modell D3).

Tabelle 15: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen mit Daten mit einem Monat Wartezeit

WZ1	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	WZ1	3,87	201
B	WZ1 + NEQ	3,45	117
C1	WZ1 + NEQ + $(int_lzb)_{t-1}$	3,33	107
C2	WZ1 + NEQ + alo_abg	3,33	115
C3	WZ1 + NEQ + $(int_lzb)_{t-1}$ + alo_abg	3,32	103
D1	WZ1 + NEQ + $(int_lzb)_{t-1}$ + alo_abg + Land	3,09	90
D2	WZ1 + NEQ + $(int_lzb)_{t-1}$ + alo_abg + VT	3,13	96
D3	**WZ1 + NEQ + $(int_lzb)_{t-1}$ + alo_abg + Land+VT**	**3,03**	**93**

Einen zusätzlichen Gewinn durch die SGB-II-Typisierung ergibt sich nur mit Blick auf den Median des Fehlers, wenn Bundesland und Vergleichstyp im Modell berücksichtigt werden.[139] Die Zahl hoher Fehler steigt leicht an. Die SGB-II-Typisierung 2014 hat in allen Wartezeitvarianten einen stärkeren Einfluss auf die Schätzungen als die ältere Version. Da eine Umstellung bevorsteht, wird die Variable beibehalten.

Daten ohne Wartezeit

Die Erklärungsgüte bei der Schätzung mit WZ0-Daten fällt gegenüber jener mit WZ1- oder WZ2-Daten deutlich ab. Jede weitere Variable kann die Fehler spürbar reduzieren. Daher werden auch hier alle getesteten Variablen in die Ermittlung des Ersatzwertes einbezogen (Modell D3).

Tabelle 16: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen mit Daten ohne Wartezeit

WZ0	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
A	WZ0	10,00	2.295
B	WZ0 + NEQ	8,25	1.473
C1	WZ0 + NEQ + $(int_lzb)_{t-1}$	7,35	1.094
C2	WZ0 + NEQ + alo_abg	7,30	1.220
C3	WZ0 + NEQ + $(int_lzb)_{t-1}$ + alo_abg	7,04	1.060
D1	WZ0 + NEQ + $(int_lzb)_{t-1}$ + alo_abg + Land	6,40	919
D2	WZ0 + NEQ + $(int_lzb)_{t-1}$ + alo_abg + VT	6,65	968
D3	WZ0 + NEQ + $(int_lzb)_{t-1}$ + alo_abg + Land + VT	6,25	872

[139] Alternativ wurden alle Modelle mit der SGB-II-Typisierung 2014 gerechnet. Sie verbessert die Vorhersagegüte aller Schätzungen sichtbar, vor allem auch für alle zurückliegenden Zeiträume seit Januar 2011.

Ohne Informationen mit geringerer bzw. ohne Wartezeit

Hier können nur die Modelle C1 bis D3 getestet werden, da keine Werte mit geringerer Wartezeit der Integrationen LZB, sondern nur die Vormonatswerte verfügbar sind. Erwartungsgemäß verbessert sich die Schätzung mit jeder weiteren Variable, sodass auch hier alle Größen einbezogen werden (Modell D3).

Tabelle 17: Modellvergleich – Integrationen von LZB: Median und große Fehler bei Schätzungen, wenn keine Daten mit verkürzter Wartezeit vorliegen

VM	Regressoren	Median – betragsmäßiger relativer Fehler in %	Anzahl große Fehler
C1	NEQ + $(int_lzb)_{t-1}$	15,39	4.922
C2	NEQ + alo_abg	16,67	5.301
C3	NEQ + $(int_lzb)_{t-1}$ + alo_abg	13,33	4.098
D1	NEQ + $(int_lzb)_{t-1}$ + alo_abg + Land	11,96	3.601
D2	NEQ + $(int_lzb)_{t-1}$ + alo_abg + VT	12,92	3.870
D3	**NEQ + $(int_lzb)_{t-1}$ + alo_abg + Land + VT**	**11,77**	**3.511**

Statistik-Infoseite

Im Internet stehen statistische Informationen, unterteilt nach folgenden Themenbereichen, zur Verfügung:

Arbeitsmarkt und Grundsicherung im Überblick
Arbeitslose, Unterbeschäftigung und Arbeitsstellen
Ausbildungsmarkt
Beschäftigung
Förderung und berufliche Rehabilitation
Grundsicherung für Arbeitsuchende (SGB II)
Leistungen SGB III
Berufe
Bildung
Daten zu den Eingliederungsbilanzen
Einnahmen/Ausgaben
Familien und Kinder
Frauen und Männer
Langzeitarbeitslosigkeit
Migration
Regionale Mobilität
Wirtschaftszweige
Zeitreihen
Amtliche Nachrichten der BA
Kreisdaten

Die Methodischen Hinweise der Statistik bieten ergänzende Informationen.

Die Qualitätsberichte der Statistik erläutern die Entstehung und Aussagekraft der jeweiligen Fachstatistik.

Das Glossar enthält Erläuterungen zu allen statistisch relevanten Begriffen, die in den verschiedenen Produkten der Statistik der BA Verwendung finden.

Abkürzungen und Zeichen, die in den Produkten der Statistik der Bundesagentur für Arbeit vorkommen, werden im Abkürzungsverzeichnis bzw. der Zeichenerklärung der Statistik der BA erläutert.