

00

Statistische Grundlagen und Übersichten

338-0072

Methodenbericht

Wertschöpfungsstatistik

Revision 2009: Statistische Datenaufbereitung und Hochrechnung

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Neuchâtel, 2014

Die vom Bundesamt für Statistik (BFS)
herausgegebene Reihe «Statistik der Schweiz»
gliedert sich in folgende Fachbereiche:

0 Statistische Grundlagen und Übersichten

1 Bevölkerung

2 Raum und Umwelt

3 Arbeit und Erwerb

4 Volkswirtschaft

5 Preise

6 Industrie und Dienstleistungen

7 Land- und Forstwirtschaft

8 Energie

9 Bau- und Wohnungswesen

10 Tourismus

11 Mobilität und Verkehr

12 Geld, Banken, Versicherungen

13 Soziale Sicherheit

14 Gesundheit

15 Bildung und Wissenschaft

16 Kultur, Medien, Informationsgesellschaft, Sport

17 Politik

18 Öffentliche Verwaltung und Finanzen

19 Kriminalität und Strafrecht

20 Wirtschaftliche und soziale Situation der Bevölkerung

21 Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene

Methodenbericht
Wertschöpfungsstatistik
Revision 2009: Statistische Datenaufbereitung und Hochrechnung

Autor Daniel Assoulin

Herausgeber Bundesamt für Statistik (BFS)

Office fédéral de la statistique (OFS)
Neuchâtel, 2014

Herausgeber: Bundesamt für Statistik (BFS)
Auskunft: Daniel Assoulin, Tel. 032 713 67 51, E-Mail: Daniel.Assoulin@bfs.admin.ch
Bearbeitung: Sektion Statistische Methoden, BFS
Vertrieb: Bundesamt für Statistik, CH-2010 Neuchâtel
Tel. 032 713 60 60 / Fax 032 713 60 61 / E-Mail: order@bfs.admin.ch
Internet: www.statistik.admin.ch
Bestellnummer: 338-0072
Preis: Gratis
Reihe: Statistik der Schweiz
Fachbereich: 0 Statistische Grundlagen und Übersichten
Originaltext: Deutsch
Titelgrafik: BFS; Konzept: Netthoevel & Gaberthüel, Biel; Foto: © NorthShoreSurfPhotos – Fotolia.com
Grafik/Layout: Sektion DIAM, Prepress/Print
Copyright: BFS, Neuchâtel 2014
Abdruck – ausser für kommerzielle Nutzung –
unter Angabe der Quelle gestattet
ISBN: 978-3-303-00516-3

Inhaltsverzeichnis

1 Einleitung **7**

2 Statistische Datenaufbereitung **7**
2.1 Die Behandlung des reduzierten Fragebogens 7
2.1.1 Vergleich mit Alternativen 8
2.1.2 Diskussion der Resultate 9
2.2 Einsetzung fehlender Bilanzwerte 13
2.3 Die Behandlung von Sammelantworten 15
2.4 Behandlung einzigartiger Unternehmen 15
2.4.1 Identifikation einzigartiger Unternehmen 15
2.4.2 Illustration - Einfluss auf Varianz 17
2.4.3 Einsetzungen bei den identifizierten Unternehmen 18

3 Die Hochrechnung der Resultate **18**
3.1 Korrektur für Antwortausfälle 18
3.2 Robustifizierung 21
3.3 Kalibrierung 23
3.4 Die Totalschätzung 24
3.4.1 Die Varianzschätzung 24
3.4.2 Einige Hochrechnungsbeispiele 26

4 Schlussfolgerungen **28**

Anhang **31**

A Wirtschaftszweige gemäss zweiter Stufe NOGA 2008 **31**

Literatur **33**

Methodenberichte der Sektion Statistische Methoden **35**

Vorwort

Die Sektion Statistische Methoden wurde mit der Entwicklung von statistischen Methoden für die Datenaufbereitung und die Hochrechnung der Wertschöpfungsstatistik beauftragt. Der Hauptteil dieser Arbeiten wurde im Zeitraum 2010-2011 durch Daniel Assoulin ausgeführt. Dabei wurde er durch Veronica Roth und Markus Däppen von der Sektion Monetäre Unternehmensstatistik (MON) begleitet. Der Autor dankt Veronica Roth und Markus Däppen für die gute und effiziente Zusammenarbeit. Ein Dank geht auch an

- Jean-Pierre Renfer und Paul-André Salamin von der Sektion Statistische Methoden für die fachliche Unterstützung.
- Markus Däppen sowie Jean-Pierre Renfer für die aufmerksame Lektüre und hilfreichen Tipps im Rahmen der Erstellung dieses Methodenberichts.

Zusammenfassung

Im Rahmen der Revision der Wertschöpfungsstatistik 2009 wurden verschiedene methodischen Anpassungen vorgenommen, welche nebst der eigentlichen Hochrechnung auch Bereiche der Datenaufbereitung betreffen. In diesem Bericht werden diese Methoden einerseits beschrieben, andererseits wird auch Wert darauf gelegt, die Methodenwahl zu begründen. Die Wertschöpfungsstatistik (WS) basiert auf einer Stichprobenerhebung in welcher eine Vielzahl betriebswirtschaftlicher Variablen erfasst werden. Nach der Datensammlung durchlaufen die Daten eine Aufbereitungsphase, in welcher sie kontrolliert und nötigenfalls bearbeitet werden, so dass sie dann in der Hochrechnung möglichst unverzerrte Resultate liefern können. Im Zusammenhang mit der Revision 2009 wurden verschiedene statistische Einsetzungsmethoden eingeführt, welche im Rahmen der Datenaufbereitung eingesetzt werden, um verschiedene Situationen wie den Einsatz eines reduzierten Fragebogens bei kleinen Unternehmen, Sammelantworten (Antworten von Unternehmensgruppen), fehlende Angaben zu gewissen Positionen oder auch Antwortausfälle bei sehr grossen Unternehmen zu behandeln. Nebst statistischen Einsetzungen wurden auch Methoden für die Identifikation und Behandlung von Unternehmen eingeführt, welche aufgrund ihrer Grösse in ihrer Branche als einzigartig betrachtet werden. Im Rahmen der Schätzung von Totalen werden diese Unternehmen mit Eins gewichtet. Für die übrigen Unternehmen besteht die Berechnung der Hochrechnungsgewichte ausgehend von den Stichprobengewichten aus drei Schritten: Korrektur für Antwortausfälle, Robustifizierung und Kalibrierung mittels der im Stichprobenrahmen zur Verfügung stehenden Hilfsvariablen Vollzeitstellenäquivalent. Im Rahmen der Hochrechnung werden Schätzungen für verschiedene Untersuchungsbereiche (z.B. Wirtschaftszweige) benötigt. Dies wird mit der Definition spezieller Variablen bewerkstelligt, welche für Unternehmen innerhalb des Untersuchungsbereiches der Zielvariablen entsprechen und ausserhalb des Untersuchungsbereiches den Wert Null annehmen. Mittels diesen Variablen und den Hochrechnungsgewichten ergeben sich die für unterschiedliche Untersuchungsbereiche verwendeten Schätzungen durch gewichtete Summen über die ganze Stichprobe. Nebst Schätzungen für Totale werden im Bericht auch Methoden dargestellt, mit denen entsprechende Varianzen und Variationskoeffizienten geschätzt werden können. Testhochrechnungen der Hauptvariablen Bruttoproduktion, Vorleistungen, Personalkosten und Wertschöpfung auf Ebene NOGA Stufe 2 führten auf Präzisionen (Variationskoeffizienten) die im Grossen und Ganzen in jenem Bereich liegen, in welchem sie aufgrund der Arbeiten rund um den Stichprobenplan erwartet werden durften. Eine Ausnahme bildet die Variable Vorleistungen, für welche die erzielten Präzisionen etwas unter den Erwartungen lagen.

1 Einleitung

Die Wertschöpfungsstatistik ist eine jährlich stattfindende Unternehmenserhebung. Sie wird durch die Sektion Monetäre Unternehmensstatistik (MON) durchgeführt und liefert Schätzungen zu ökonomisch wichtigen Daten wie der Bruttoproduktion oder der Wertschöpfung. Die Rahmenbedingungen für die Erhebung 2009 und der nachfolgenden Jahre unterscheiden sich in gewissen Punkten von jenen für die Vorjahre:

- Änderung in der Definition der Wirtschaftszweige. Die frühere Einteilung basierte auf der NOGA[1] des Jahres 2002 und die neue Einteilung auf der NOGA des Jahres 2008. Ein expliziter Beschrieb der Wirtschaftszweige gemäss NOGA-Stufe 2 des Jahres 2008 (kurz: NOGA2) befindet sich in Tabelle 11 des Anhangs.
- Eine neue Stichprobe, welche das Panel der Jahre 2002-08 ersetzt und im Stichprobenverwaltungssystem (SVS) gezogen wurde. Dabei ist zu beachten, dass vorgesehen ist, die Stichprobe jährlich zu erneuern, womit deren Aktualität gewährleistet ist[2].
- Für Eurostat und im Rahmen der Volkswirtschaftlichen Gesamtrechnung werden Resultate auf verschiedenen Niveaus wie NOGA Stufe 3 (kurz: NOGA3) oder NOGA Stufe 2 gekreuzt mit Rechtsformgruppen benötigt. Die Hochrechnungsmethodik sollte also eine gewisse Flexibilität bezüglich der betrachteten Untersuchungsbereiche bieten.

Die hinsichtlich der statistischen Datenaufbereitung sowie der Hochrechnung vorgenommenen Anpassungen werden in diesem Bericht eingehend beschrieben.

2 Statistische Datenaufbereitung

Im Vorfeld der eigentlichen Hochrechnung werden die Daten aufbereitet. Diese Datenaufbereitung umfasst vielseitige Arbeiten wie Plausibilitätschecks, Nacherfassungen, telefonische Abklärungen und die Berechnung von Einsetzungen anhand von Sekundärdatenquellen. Die Datenaufbereitung wird durch MON vorgenommen. Dieser Bericht beschränkt sich auf den Beschrieb von Methoden, welche durch die Sektion Statistische Methoden (METH) entwickelt wurden und im Rahmen der Datenaufbereitung verwendet werden.

2.1 Die Behandlung des reduzierten Fragebogens

Im Fragebogen der Wertschöpfungsstatistik werden Angaben zu Buchhaltungspositionen erhoben, welche kleine Unternehmen in ihrer Buchhaltung oft nicht separat ausweisen. Aus diesem Grund werden Unternehmen mit weniger als 20 Beschäftigten mit einem reduzierten Fragebogen befragt [3]. Dabei bezieht sich der kleine Fragebogen grösstenteils auf Aggregate von Posten aus dem grossen Fragebogen. Um ein einheitliches Vorgehen bei der weiteren Auswertung zu ermöglichen, werden Positionen, welche im kleinen Fragebogen nicht abgedeckt werden, mittels Einsetzungen behandelt. Im Folgenden wird die entsprechende Technik vorgestellt. Wir betrachten den Fall einer Variablen y_j, die Teil des Aggregates $A \in \{A_1, ..., A_g\}$ ist. Für ein Unternehmen i ist das Aggregatstotal durch

$$y_{i,A} := \sum_{j \in A} y_{i,j}$$

[1] Allgemeine Systematik der Wirtschaftszweige.

[2] Details zur Stichprobe der WS2009 findet man in [9].

[3] Eine Ausnahmen bildet die Branche Energieversorgung, in welcher kein reduzierter Fragebogen verwendet wird.

gegeben. Im Falle eines reduzierten Fragebogens ist der Aggregatswert $y_{i,A}$ bekannt, die Komponenten $y_{i,j},\ j \in A$, müssen hingegen geschätzt / eingesetzt werden. Für die Einsetzung betrachten wir eine Aufteilung:

$$\hat{y}_{i,j} \quad = \quad a_{j,A} y_{i,A} \tag{1}$$

mit $a_{j,A} \geq 0$ und $\sum_{j \in A} a_{j,A} = 1$. Nun müssen die Komponenten $a_{j,A}$ geschätzt werden. Hierzu verwenden wir folgende Notation:

- c entspricht einer Wirtschaftsbranche, definiert durch die zweite Stufe der NOGA 2008 (kurz NOGA2) oder (bei NOGAs mit nur wenig Unternehmen - siehe weiter unten) einem NOGA2-Aggregat.
- SC ist die Menge der Unternehmen mit einem Beschäftigtentotal gemäss Stichprobenrahmen im Intervall $[20; 249]$. Diese Unternehmen erhalten im Allgemeinen einen vollständigen Fragebogen.
- r = Menge der im Rahmen der Erhebung antwortenden Unternehmen.

Um die Einsetzungen entsprechend (1) vornehmen zu können, wird nun der Anteil $a_{j,A}$, separat für jeden Wirtschaftssektor c, wie folgt geschätzt:

$$\hat{a}_{j,A}^{(c)} \quad = \quad \frac{\sum_{k \in r \cap SC \cap c} y_{k,j}}{\sum_{k \in r \cap SC \cap c} y_{k,A}} \tag{2}$$

Bemerkungen:

- Damit die Schätzungen $\hat{a}_{j,A}^{(c)}$ einigermassen stabil sind, werden benachbarte NOGAs so zusammengefasst, dass $r \cap SC \cap c$ mindestens 20 Unternehmen umfasst.
- Die Beschränkung auf Unternehmen mit höchstens 249 Beschäftigten soll verhindern, dass die Schätzung (2) von einzelnen sehr grossen Unternehmen dominiert wird. Dabei ist zu beachten, dass entsprechende Analysen keinen Grund zur Annahme lieferten, dass die Anteile $a_{j,A}$ von der Grösse der betrachteten Unternehmen abhängig wären.

2.1.1 Vergleich mit Alternativen

Die Wahl des vorgestellten Einsetzungsverfahrens basierte auf Vergleichen mit alternativen Einsetzungsmethoden. Diese werden im Folgenden kurz illustriert.

- HD (Hot Deck): Die Anteile für die Verteilung des Aggregates werden hier unternehmensspezifisch festegelegt. Die Einsetzung für Unternehmen i beruht auf:

  $$\hat{y}_{i,j} = \frac{y_{i^*,j}}{y_{i^*,A}} y_{i,A},$$

  wobei i^* ein zufällig gewähltes Unternehmen aus $r \cap SC \cap c$ bezeichnet.

- NNBR (Nearest Neighbor): gleich wie HD, aber hier wird das Unternehmen i^* anhand

  $$i^* = \mathbf{argmin}_{k \in r \cap SC \cap c} D_{L2}(\mathbf{s}(\mathbf{y_{i,\mathcal{A}}}), \mathbf{s}(\mathbf{y_{k,\mathcal{A}}}))$$

  gewählt. Dabei bezeichnet D_{L2} die L2-Distanz und $\mathbf{s}(\mathbf{y_{i,\mathcal{A}}}) = (\mathbf{s}(\mathbf{y})_{\mathbf{i,A_1}}, \ldots, \mathbf{s}(\mathbf{y})_{\mathbf{i,A_g}})$ den Vektor bestehend aus standardisierten[4] Aggregatstotalen der betrachteten Aggregate $\mathcal{A} = \{A_1, ..., A_g\}$.

[4] Zentrierung mittels Mittelwert und Division mittels Standardabweichung der Antwortenden der Branche c mit weniger als 250 Beschäftigten.

- ONE: Das ganze Aggregat wird in eine - aus fachspezifischen Kenntnissen ermittelte - Hauptposition gesetzt. Die anderen Positionen werden Null gesetzt. Diese Einsetzungsvariante ist von daher interessant, als sie dem Datenzustand nach der elektronischen Erfassung der Daten entspricht [5].

Die Methoden wurden im Rahmen der Datenaufbereitung der WS2009 entwickelt und getestet. Aus zeitlichen Gründen konnten die Analysen nicht erst nach Abschluss der Datensammlungsperiode durchgeführt werden. Das für die Analysen verwendete Datenmaterial beruhte daher auf:

- definitiven Daten der WS2008.
- provisorischen Daten der WS2009 (Stand Oktober 2009).

Die durchgeführten Tests beruhten darauf, dass man eingesetzte Werte mit effektiv erhobenen Werten vergleicht. Darum wurde das Einsetzungsverfahren auf Unternehmen mit betot in $[20;30]$ angewandt, welche ja effektiv mit dem vollständigen Fragebogen befragt worden sind. Die Grössenklasse der Unternehmen, mit welchen die Anteile a geschätzt werden, wurde entsprechend auf die Unternehmen in $SC^* = \{i; betot_i \in [31;249]\}$ eingeschränkt. Zur Beurteilung der Einsetzung wurden in jedem Wirtschaftssektor c für jede Einsetzungsmethode folgende Grössen betrachtet:

$$m_{c,j} = \frac{|\sum_{\{i;\ i \in r \cap c \wedge betot_i \in [20;30]\}} (\hat{y}_{i,j} - y_{i,j})|}{\sum_{\{i;\ i \in r \cap c \wedge betot_i \geq 20\}} y_{i,j}} \quad (3)$$

$$m_{w,c,j} = \frac{|\sum_{\{i;\ i \in r \cap c \wedge betot_i \in [20;30]\}} w_i(\hat{y}_{i,j} - y_{i,j})|}{\sum_{\{i;\ i \in r \cap c \wedge betot_i \geq 20\}} w_i y_{i,j}} \quad (4)$$

Es handelt sich also um eine Art Schätzung des relativen durch die Einsetzung verursachten Schätzfehlers. Die Gewichte w_i entsprechen den Stichprobengewichten, wobei im Falle der WS2008 Daten noch die Gewichtskorrektur für Antwortausfälle berücksichtigt wurde. Betrachtet wurden drei Variablengruppen, welche jeweils zu einem Aggregat gehören. Bei der Gruppe 1 (oder Aggregat 1) handelt es sich um 10 Aktivposten der Bilanz, bei der Gruppe 2 um vier Passivposten aus der Bilanz und bei Gruppe 3 um 7 Ertragsposten[6]. Es gibt je Kombination $methode \times c \times variable$ einen m-Wert. Alles in Allem also sehr viele Zahlen. Darum wurden die Vergleiche über die Wirtschaftsbereiche c hinweg mittels Rangierungsmethoden und Median zusammengefasst. Das heisst, es resultiert für jede Kombination $Methode \times Variable$ ein Median sowie ein mittlerer Rang. Die mittleren Ränge für die gewichteten Vergleiche findet man in der Tabelle 1. Ein mittlerer Rang von Eins würde bedeuten, dass eine Methode in sämtlichen Wirtschaftsbereichen die kleinsten m-Werte aufweist. Entsprechend bedeutet ein mittlerer Rang von Vier, dass eine Methode immer die grössten m-Werte aufweist. Die Mediane sind in Tabelle 2 abgebildet [7].

2.1.2 Diskussion der Resultate

In erster Linie geht es bei den Tests um einen Vergleich der verschiedenen Methoden. Hierzu konzentrieren wir uns als Erstes auf die mittleren Ränge in Tabelle 1. Die fett markierten

[5] Die Methode ONE entspricht auch dem Vorgehen der Vorjahre.

[6] Die hier verwendeten Variablengruppen (oder Aggregate) dienen der Illustration. Die Definition der Variablengruppen, auf welche das Einsetzungsverfahren effektiv angewandt wird, wird durch MON vorgenommen.

[7] Für die Positionen s100 und s102 waren aufgrund der Datenlage der WS2008 keine Analysen möglich, weshalb entsprechende Felder in den Ergebnistabellen 1 und 2 leer sind.

Tabelle 1 Statistik m1 gewichtet: mittlere Ränge aus der Simulation.

	2008				2009			
Pos.	HD	NNBR	QUOT	ONE	HD	NNBR	QUOT	ONE
	Variablengruppe 1							
s100					2,4	2,3	2,6	2,7
s101	2,4	2,2	2,4	3,0	2,3	2,1	1,9	3,6
s102					2,4	2,6	2,9	2,2
s103	2,5	2,1	2,9	2,5	2,3	2,5	2,8	2,4
s104	**2,0**	**2,4**	**1,6**	**4,0**	**2,0**	**2,0**	**1,9**	**4,0**
s107	2,6	2,3	2,1	3,0	2,2	2,3	2,6	2,9
s111	2,7	2,3	2,6	2,4	2,0	2,5	2,2	3,2
s108	2,1	2,0	2,2	3,7	2,4	2,4	2,4	2,8
s109	2,5	2,5	2,5	2,6	2,4	2,5	2,6	2,5
s110	2,3	2,4	2,2	3,1	2,4	2,5	2,4	2,7
	Variablengruppe 2							
s204	2,5	2,7	2,2	2,6	2,7	2,3	2,2	2,8
s205	**2,4**	**3,0**	**2,3**	**2,3**	**2,5**	**2,6**	**1,9**	**3,0**
s207	2,5	2,5	2,7	2,3	2,4	2,4	2,1	3,2
s208	2,5	3,0	2,2	2,2	2,9	2,6	2,2	2,2
	Variablengruppe 3							
s600	**2,4**	**2,7**	**2,4**	**2,5**	**2,0**	**2,0**	**2,2**	**3,7**
s601	2,5	2,5	2,5	2,4	2,4	2,3	2,5	2,8
s602	2,6	2,7	2,6	2,1	2,6	2,1	2,2	3,1
s604	2,5	2,5	2,9	2,1	2,5	2,3	3,0	2,1
s605	2,3	2,4	2,8	2,5	2,4	2,4	2,5	2,7
s609	2,8	2,5	3,0	1,7	2,3	2,5	3,1	2,1
s610	2,2	2,4	2,1	3,3	2,7	2,5	2,3	2,6

Tabelle 2 Statistik m1: Medianwerte aus der Simulation.

	m1 ungewichtet in %								m1 gewichtet in %							
	2008				2009				2008				2009			
Pos.	HD	NNBR	QUOT	ONE	HD	NNBR	QUOT	ONE	HD	NNBR	QUOT	ONE	HD	NNBR	QUOT	ONE
	Variablengruppe 1															
s100					4,4	7,7	8,9	3,4					1,7	1,8	3,2	0,9
s101	3,2	3,1	4,2	5,9	2,6	2,2	3,0	6,5	0,9	0,5	0,7	2,4	1,2	1,1	1,1	2,7
s102					1,8	1,8	6,2	0,0					0,7	0,6	2,3	0,0
s103	3,8	3,9	8,2	2,8	2,1	3,2	6,7	1,3	1,3	1,1	1,6	1,5	1,0	1,9	2,8	0,3
s104	**4,8**	**4,4**	**4,1**	**28,8**	**1,3**	**1,5**	**1,8**	**17,2**	**0,8**	**1,0**	**0,6**	**5,0**	**0,5**	**0,7**	**0,7**	**5,2**
s107	10,0	8,5	8,2	13,0	1,7	2,0	2,3	2,3	3,2	2,5	2,5	6,8	0,7	0,7	1,0	0,9
s108	11,3	11,9	11,5	16,1	4,4	2,7	3,8	3,6	1,1	1,0	1,2	3,0	1,8	1,2	1,5	1,4
s109	16,2	12,3	15,4	33,5	3,9	5,2	5,0	3,0	1,3	1,7	1,6	3,6	1,5	2,5	2,1	1,7
s110	7,8	7,9	6,7	11,3	3,0	2,2	3,6	1,5	1,2	1,4	2,0	2,6	0,9	1,0	1,1	1,0
s111	6,3	4,6	4,3	4,0	2,3	3,8	2,5	5,1	0,7	1,0	0,8	1,3	1,1	1,2	1,2	2,1
	Variablengruppe 2															
s204	5,0	6,8	6,5	7,2	4,4	3,5	2,7	4,4	1,5	1,1	1,1	2,0	4,4	3,5	2,7	4,4
s205	**1,5**	**2,7**	**1,2**	**1,3**	**4,1**	**4,3**	**3,1**	**6,9**	**0,6**	**0,6**	**0,6**	**0,4**	**4,1**	**4,3**	**3,1**	**6,9**
s207	4,3	3,0	4,8	1,2	2,9	2,8	1,9	3,7	0,7	0,5	1,7	0,4	2,9	2,8	1,9	3,7
s208	3,5	5,7	2,8	1,8	5,7	4,8	2,4	2,2	1,1	1,7	1,1	0,5	5,7	4,8	2,4	2,2
	Variablengruppe 3															
s600	**0,6**	**0,7**	**0,8**	**0,7**	**0,1**	**0,3**	**0,7**	**6,4**	**0,1**	**0,2**	**0,2**	**0,3**	**0,1**	**0,3**	**0,7**	**6,4**
s601	34,2	34,0	33,9	7,1	3,1	2,2	6,6	3,7	0,4	0,4	0,4	0,8	3,1	2,2	6,6	3,7
s602	29,2	32,7	32,4	5,4	1,8	0,6	1,4	4,4	1,0	1,3	1,1	2,0	1,8	0,6	1,4	4,4
s604	10,2	6,9	22,8	0,0	1,5	1,6	6,2	0,0	1,1	1,2	3,0	0,0	1,5	1,6	6,2	0,0
s605	10,9	9,9	12,6	11,3	3,6	3,5	4,9	1,9	0,8	0,9	1,7	0,8	3,6	3,5	4,9	1,9
s609	10,7	2,8	33,8	0,0	0,0	0,1	6,3	0,0	1,4	0,7	5,3	0,0	0,0	0,1	6,3	0,0
s610	16,4	18,8	18,6	39,1	10,3	5,0	9,4	3,7	6,8	5,8	4,6	6,0	10,3	5,0	9,4	3,7

Werte entsprechen den Testergebnissen für die jeweiligen Hauptpositionen der drei betrachteten Aggregate (s104=Forderungen aus Lieferungen und Leistungen; s205=langfristige Bank- und Hypothekarverbindlichkeiten; s600=Nettoproduktionsertrag) . Die Hauptpositionen sind jene Positionen, welche üblicherweise die höchsten Werte aufweisen und in welche im Rahmen der elektronischen Datenerfassung des kleinen Fragebogens der gesamte Aggregatswert geschrieben wird (entspricht Einsetzungsvariante ONE)[8].

- In den Hauptpositionen und anderen betragsmässig wichtigen Positionen zeigt es sich, dass sämtliche Alternativmethoden tendenziell tiefere mittlere Ränge aufweisen als ONE. Für die Hauptpositionen erzielt die Methode QUOT in fünf der sechs betrachteten Situationen den tiefsten mittleren Rang - aber die Unterschiede zwischen den Einsetzungsalternativen sind nicht gross.
- In den restlichen Variablen ist aufgrund von m durch eine Einsetzung generell keine wesentliche Verbesserung zu erwarten (Variante ONE schneidet nicht schlechter ab als die betrachteten Alternativen). Dies kann damit erklärt werden, dass die beobachteten Werte in diesen Positionen effektiv oft klein (auch Null) sind, weshalb eine Einsetzung mittels der Konstanten Null vergleichsweise nicht schlecht abschneidet. Hingegen ist zu beachten, dass die Zielvariablen im Allgemeinen keine negativen Werte aufweisen können und somit die Einsetzung mit Null zu einer systematischen Unterschätzung führt. Dieser Tatsache wird mittels der Betrachtung von m nicht Rechnung getragen.

Betrachtet man die Medianwerte in Tabelle 2, so stellt man folgendes fest:

- In Hauptpositionen und anderen betragsmässig wichtigen Positionen erzielen die Alternativen im Vergleich zu ONE bessere Resultate. In anderen Variablen kann ONE durchaus mithalten.
- Die drei betrachteten Alternativen schneiden - insbesondere in den Hauptpositionen - ähnlich ab.
- Die Mediane weisen darauf hin, dass die Grösse der relativen Fehlern von der Position abhängig ist. Dabei ist es so, dass die alternativen Methoden für betragsmässig wichtige Positionen (z.B. Hauptpositionen) tendenziell kleinere relative Fehler aufweisen als für betragsmässig kleinere Positionen, in welchen die relativen Fehler gross werden können. Dies kann damit erklärt werden, dass die Bezugsgrössen (Nenner) für die Berechnung der relativen Fehler bei diesen Positionen häufig eben klein sind.

Bemerkung: Bei den Masszahlen ist zu beachten, dass sie den relativen Fehler in Bezug auf die Testsituation messen. Die Grössenordnung in der effektiven Einsetzung wird eine andere sein, da wir ja nicht die Unternehmen mit betot zwischen 20 und 30 sondern jene mit betot in [3;19] einsetzen. Für unsere Tests ist dies hingegen nicht so tragisch, da es vorallem darum geht, mehrere Varianten zu vergleichen und die geeignetste auszuwählen.

Schlussfolgerungen: In betragsmässig wichtigen Positionen (insbesondere Hauptpositionen) erzielen die betrachteten Einsetzungsvarianten gegenüber ONE (Zustand der Daten nach elektronischer Erfassung) im Allgemeinen Vorteile. Die Variante QUOT ist gut interpretierbar und aufgrund der Evaluation der Tests gibt es keine Einsetzungsvariante, welche klar als die Beste hervorsticht. Darum wurden im Rahmen der WS09 die kleinen Fragebogen mittels der Einsetzungsmethode QUOT behandelt. Die Tests liessen vermuten, dass die Einsetzungen bei

[8] Hauptpositionen sind aufgrund Ihrer betragsmässigen Wichtigkeit auch für separate Hochrechnungen von Interesse, während die anderen Positionen oft nur aggregiert betrachtet werden.

branchenweisen Hochrechnungsresultaten von betragsmässig weniger wichtigen Positionen zu grossen relativen Schätzfehlern führen können. Dies kann damit erklärt werden, dass der Schätzfehler in Bezug zu einem kleinen Total gesetzt wird. Ist das Hochrechnungsresultat einer solchen Einzelposition wichtig, empfiehlt es sich, die Position im kleinen Fragebogen explizit zu erheben. Bei der Interpretation der m-Werte ist hingegen zu beachten, dass es sich bei den Hauptvariablen wie t017 (Bruttoproduktion), t022 (Vorleistungen), t053 (Personalkosten) oder t029 (Wertschöpfung) um betragsmässig wichtige Positionsaggregate und eben nicht um Einzelpositionen handelt. Abbildung 1 stellt hierzu die relativen Differenzen zwischen zwei Testhochrechnungen der Daten auf Niveau NOGA 2002[9] Stufe 2 vor und nach statistischen Einsetzungsverfahren in Form von Boxplots dar. Die berücksichtigten statistischen Einsetzungsverfahren umfassen nebst den in diesem Abschnitt diskutierten Positionseinsetzungen (betraf im Rahmen der WS2009 1613 Unternehmen) auch die in späteren Abschnitten vorgestellten Einsetzungsverfahren:

- die Behandlung konsolidierter Antworten (betraf in der WS2009 24 Unternehmen)
- unternehmensspezifische Einsetzung für sehr grosse Unternehmen (betraf 6 Unternehmen).

Die Daten vor Einsetzung entsprechen in der Testhochrechnung dem Datensatz vor den statistischen Einsetzungsverfahren und somit auch dem Verfahren vor der Revision:

- Aggregatswerte werden in Hauptposition geschrieben (Methode ONE).
- Eine konsolidierte Antwort wird dem Hauptunternehmen zugeordnet.
- Antwortausfälle bei sehr grossen Unternehmen werden durch Gewichtskorrektur der übrigen Unternehmen (Kalibrierung) kompensiert.

Aufgrund der im Boxplot dargestellten Resultate schliesst man, dass der Einfluss der Einsetzungen auf die Schätzresultate für die Hauptvariablen generell kleiner als 1% ist. Ausnahmefälle tauchen insbesondere in den Variablen t022 und t029 auf. Diese werden kurz erläutert:

- In der Variablen t022 fallen mehrere Differenzen mit Absolutwerten von über 1% auf. Bis auf eine Ausnahme handelt es sich dabei um Branchen mit einem kleinen Totalwert in t022 - weshalb bereits eine kleine Veränderung in der Schätzung zu einer relativen Veränderung von über einem Prozent führen kann. Die Ausnahme bildet das Gastgewerbe. Dieses befindet sich ebenfalls unter den Branchen mit einem Einsetzungseinfluss von über einem Prozent, obwohl das Total t022 recht gross ist. Dies könnte damit zusammenhängen, dass diese Branche einen hohen Anteil kleinerer Unternehmen aufweist und somit die Art der Behandlung des reduzierten Fragebogens eine grössere Rolle spielt als in anderen Branchen.
- In der Variablen t029 (Wertschöpfung) fällt eine relative Differenz von -3.25% auf. Diese Differenz gehört zur Branche Rückgewinnung (NOGA2002-Code: 37), welche für die Variable t029 ein sehr kleines Total aufweist, was die grosse relative Differenz erklären kann.

[9] Nebst der NOGA 2008 stand für sämtliche Unternehmen im Rahmen der WS2009 auch die NOGA 2002 zur Verfügung. Für die früheren Erhebungen ist diese Doppelcodierung nicht vorhanden, jedes Unternehmen besitzt einfach die NOGA 2002 Branchenzuteilung. Um den Einfluss der methodischen Änderungen auf die Resultate besser beurteilen zu können, bediente man sich darum für Tests oft der NOGA 2002, weil für diese Branchen aufgrund der früheren Erhebungen Vergleichswerte zur Verfügung standen.

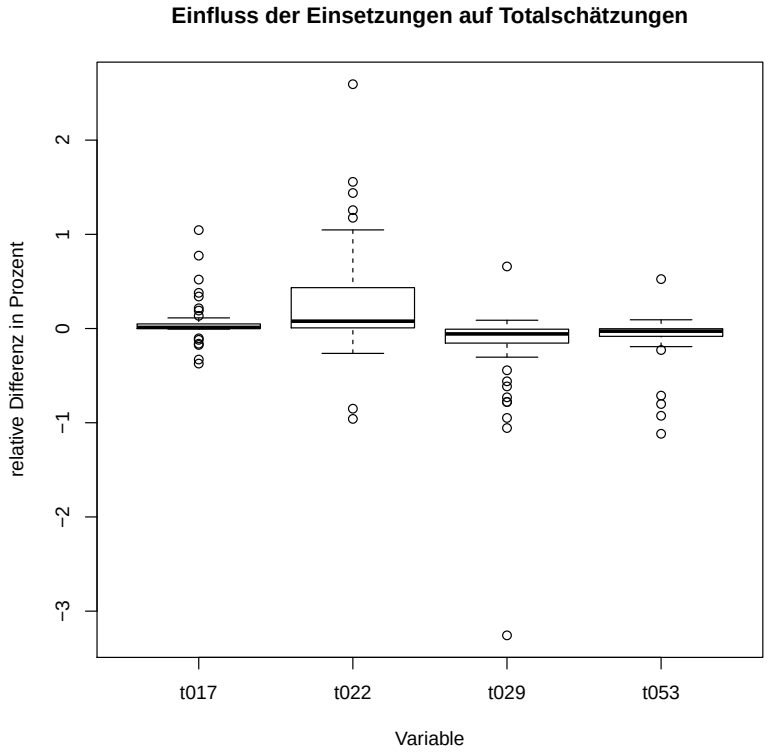


Abbildung 1 Untersuchung des Einflusses der Einsetzungen auf Totalschätzungen Niveau NOGA 2002, Stufe 2, anhand Testhochrechnungen WS2009.

2.2 Einsetzung fehlender Bilanzwerte

Es gibt eine gewisse Anzahl Unternehmen mit kleinem Fragebogen, welche keine Angaben zu den Bilanzpositionen liefern. Diese Merkmalausfälle werden mit einem zweistufigen Einsetzungsverfahren behandelt:

- in einem ersten Schritt wird mittels einer Regressionsprädiktion die Bilanzsumme eingesetzt
- in einem zweiten Schritt wird dann die Bilanzsumme analog zur Methodik für die Behandlung des reduzierten Fragebogens auf die im kleinen Fragebogen erhobenen Aktiv- und Passivposten aufgeteilt.

Grundsätzlich bestehen die Einsetzungsklassen für dieses Verfahren aus den Unternehmen mit kleinem Fragebogen einer gewissen NOGA2. Für die Imputationsklassen wird verlangt, dass sie mindestens 30 Beobachtungen mit Bilanzwerten enthalten. Ist dies nicht erfüllt, werden benachbarte NOGAs zusammengefasst. Diese Zusammenfassung wird mittels der Variablen imp_{nog} codiert. Die Regression wird im Folgenden für eine gewisse Einsetzungsklasse vorgestellt:

$$ln(BS_i + 1) = \beta_1 ln(s600_i + 1) + \beta_2 ln(s4430_i + 1) + \epsilon_i \quad (5)$$

Dabei bezeichnen

- BS_i die Bilanzsumme
- $s600_i$ den Nettoproduktionsertrag

Tabelle 3 Geschätzte Regressionskoeffizienten für die Bilanzeinsetzung.

imp_{nog}	Parameter	Estimate	tValue	Probt
47	s600	0.77	76.92	< 0.0001
47	s4430	0.22	10.30	< 0.0001
68	s600	0.84	57.63	< 0.0001
68	s4430	0.35	23.73	< 0.0001
69	s600	0.83	43.08	< 0.0001
69	s4430	0.26	4.27	$= 0.0002$
88	s600	0.81	62.46	< 0.0001
88	s4430	0.27	5.13	< 0.0001
3738	s600	0.94	24.55	< 0.0001
3738	s4430	0.23	4.71	< 0.0001
5556	s600	0.71	54.56	< 0.0001
5556	s4430	0.35	15.07	< 0.0001
7172	s600	0.88	40.52	< 0.0001
7172	s4430	0.35	5.29	< 0.0001
7374	s600	0.85	37.46	< 0.0001
7374	s4430	0.30	2.65	$= 0.0114$
8587	s600	0.87	28.99	< 0.0001
8587	s4430	0.26	7.11	< 0.0001
9596	s600	0.73	55.88	< 0.0001
9596	s4430	0.35	10.04	< 0.0001
131415	s600	0.84	46.35	< 0.0001
131415	s4430	0.21	2.77	$= 0.0081$

- $s4430_i$ den Immobilienbestand
- ϵ_i das Residuum von Unternehmen i mit Erwartungswert Null.

Die Addition der Konstanten 1 ist nötig, weil die betrachteten Variablen in einigen (seltenen) Fällen den Wert Null annehmen können und dann die zur Stabilisierung der Residuenvarianz verwendete Logarithmierung nicht definiert wäre. Tests haben gezeigt, dass die Hinzunahme weiterer erklärender Variablen oder eines Intercepts die Modellgüte nicht mehr wesentlich verbessert. Die Schätzung der Regressionsparameter erfolgt mittels der Methode der kleinsten Quadrate, wobei im Sinne einer Robustifizierung vorgängig eine Ausreisserelimination vorgenommen wird. Kriterium für die Ausreisseridentifikation:

$$|\hat{\epsilon}_{s,i}| \geq 3 \tag{6}$$

wobei $\hat{\epsilon}_{s,i}$ eine robuste Schätzung (SAS-Prozedur Robustreg) des standardisierten Residuums von Beobachtung i ist. Einen Anhaltspunkt zur Modellgüte verschafft Tabelle 3, in welcher die in den einzelnen Einsetzungsklassen erzielten Parameter- und t-Werte (tValue) für den provisorischen Datensatz vom Februar 2010 gelistet sind. Dabei bezeichnet imp_{nog} 3738 beispielsweise die Einsetzungsklasse, welche aus der Zusammenlegung der NOGAs 37 und 38 hervorgeht. Zu beachten ist, dass das Modell nur in NOGAs betrachtet wurde, in welchen tatsächlich fehlende Bilanzwerte auftraten. Da kein Intercept verwendet wurde, liefert das Programm R^2-Werte ohne Mittelwertabzug. Diese fallen allesamt sehr hoch aus - im Bereich von 99%. Aber bei der Interpretation dieser Werte ist (wegen des fehlenden Mittelwertabzuges) Vorsicht geboten, weshalb auf eine Listung der Werte verzichtet wird.
Für ein Unternehmen i mit fehlender Bilanzangabe wird der Einsetzungswert $\widehat{BS}_i$ dann mittels Rücktransformation aus der Schätzung von $ln(BS_i + 1)$, $\widehat{lnBS}_i$ ermittelt:

$$\widehat{BS}_i = \exp(\widehat{lnBS}_i) - 1 \tag{7}$$

Die Rücktransformaton ist nicht linear und es ist somit zu erwarten, dass sie zu einer leicht verfälschten Schätzung führt. Aufgrund der relativ kleinen Anzahl betroffener Unternehmen (im Rahmen der WS09 insgesamt 37 Unternehmen) sowie der Tatsache, dass es sich allesamt um

kleine Unternehmen handelt, wurde dies in Kauf genommen. (Als Verbesserung könnte eine Biaskorrektur, z.B. gemäss [8], ins Auge gefasst werden.) Analog zum Vorgehen bei reduzierten Fragebogen wird anschliessend die Bilanzsumme auf die Aktiv- und Passivposten des kleinen Fragebogens verteilt.

2.3 Die Behandlung von Sammelantworten

Das Problem der Sammelantworten entsteht, wenn für eine Unternehmensgruppe, g, nur eine Antwort vorliegt. In diesem Fall verteilt die Sektion MON die Werte der Gruppenantwort proportional zur Anzahl Beschäftigten (betot) gemäss Stichprobenrahmen auf sämtliche Unternehmen in g. Falls also y_g der Wert der Zielvariablen y auf Gruppenebene darstellt, so erhält jedes Unternehmen $i \in g$ einen Wert

$$\hat{y}_i = \frac{betot_i}{\sum_{i \in g} betot_i} y_g \tag{8}$$

Diese Einsetzung betrifft im Rahmen der WS09 die Werte von 24 Unternehmen.

2.4 Behandlung einzigartiger Unternehmen

Naturgemäss sind viele interessante Variablen in Unternehmenserhebungen positiv und rechtsschief verteilt. Dabei zeigt es sich auch, dass es je nach Aktivitätssektor einzigartige Unternehmen geben kann, die einen wesentlichen Teil an einem interessierenden Total ausmachen. Diese Unternehmen befinden sich aufgrund ihrer Grösse üblicherweise in Schichten, die vollerhoben werden. Erfahrungsgemäss ist es schwierig Antwortausfälle in diesen Take-All Schichten vollständig zu verhindern. Ohne besondere Massnahmen führen Antwortausfälle in Take-All Schichten zu folgenden Problemen:

- der gewünschte Effekt (Varianzelimination durch Vollerhebung) wird nicht erzielt.
- Es gibt einzelne Unternehmen, für welche Antwortausfälle die Schätzungen stark beeinflussen und Vergleiche über die Zeit erschweren.

In der WS soll dieser Problematik mit folgender Strategie begegnet werden:

- Identifikation potentiel problematischer Unternehmen bereits in der Datensammlungs-/ aufbereitungsphase.
- Verhinderung von Antwortausfällen in dieser Gruppe.
- Im Rahmen der Hochrechnung werden die identifizierten Unternehmen als separate vollerhobene Schicht mit Antwortrate 100% behandelt.

Das Vorgehen kann als Spezialfall von Selective Editing [6] angesehen werden.

2.4.1 Identifikation einzigartiger Unternehmen

Die Identifikation beruht auf dem bedingten Bias des Horvitz-Thompson Schätzers und wird im Folgenden vorgestellt:

- R_{TA_c}= Menge der m_c antwortenden Unternehmen in TA_c (Take-All Schicht Wirtschaftszweig c);
- Behandlung von R_{TA_c} als einfache Zufallsstichprobe aus TA_c;

- X_{TA_c}: Anzahl Beschäftigte in TA_c gemäss Stichprobenrahmen;
- Bedingter Bias des Horvitz-Thompson Schätzers $\hat{X}_{TA_c}$ gegeben i^* hat nicht geantwortet ([5]):

$$\mathbf{E}(\hat{X}_{TA_c} - X_{TA_c} | i^* \in TA_c \setminus R_{TA_c}) = \bar{X}_{TA_c \setminus i^*} - x_{i^*} \tag{9}$$

Berechnung von (9):

$$\begin{aligned}
\mathbf{E}(\hat{X}_{TA_c} | i^* \in TA_c \setminus R_{TA_c}) &= \sum_{i \in TA_c} \frac{N_{TA_c}}{m_c} \mathbf{E}(I(i \in R_{TA_c}) | i^* \in TA_c \setminus R_{TA_c}) x_i \\
&= \sum_{i \in TA_c \setminus i^*} \frac{N_{TA_c}}{m_c} \overbrace{\mathbf{E}(I(i \in R_{TA_c}) | i^* \in TA_c \setminus R_{TA_c})}^{\frac{m_c}{N_{TA_c}-1}} x_i \\
&= N_{TA_c} \bar{X}_{TA_c \setminus i^*} \qquad (10) \\
X_{TA_c} &= x_{i^*} + (N_{TA_c} - 1) \bar{X}_{TA_c \setminus i^*} \qquad (11)
\end{aligned}$$

Die Differenz zwischen (10) und (11) führt auf $\bar{X}_{TA_c \setminus i^*} - x_{i^*}$. Der bedingte Bias hängt somit nicht von m_c, der Anzahl Antwortenden ab. Die Berechnung benötigt also keinerlei Annahmen zu m_c. Einzigartige Unternehmen, im Sinne von gross, werden nun wie folgt identifiziert:

$$U_{c,1} = \left\{ i \in TA_c; \frac{x_i - \bar{X}_{TA_c \setminus i}}{X_c} \geq k \right\} \tag{12}$$

Dabei ist X_c das Total Beschäftigte im Wirtschaftszweig c. Im Rahmen der WS2009 wurde die Konstante k für alle Wirtschaftszweige c gleich 0.025 (= 2.5%) gesetzt und die Wirtschaftszweige entsprechen der NOGA 2008 Stufe 2. Interpretion: Ein Unternehmen wird also als einzigartig identifiziert, falls sein Einfluss auf die Schätzung des Branchentotals, gemessen mittels dem bedingten Bias, mindestens 2.5 % des Totals beträgt.

Bemerkung: Mit einer branchenspezifische Wahl von k könnte die Wichtigkeit des Wirtschaftszweiges berücksichtigt werden.

Iteration

Wir gehen davon aus, dass am Ende der Datenaufbereitungsphase für 100% der Unternehmen in $U_{c,1}$ eine Antwort vorliegen wird. Die Unternehmen in $U_{c,1}$ erhalten dann im Rahmen der Extrapolation ein Gewicht von 1: Die Identifikation beeinflusst also die Schätzung, was zu einer Iteration führt:

- Setze $TA_c^{(2)} := TA_c \setminus U_{c,1}$ und berechne

$$U_{c,2} = \left\{ i \in TA_c^{(2)}; \frac{x_i - \bar{X}_{TA_c^{(2)} \setminus i}}{X_c} \geq k \right\}$$

- Iteration bis $U_{c,\ell} = \emptyset$, d.h. keine Unternehmen mehr identifiziert werden.
- Zusammenfassung der Identifizierten in

$$U_c = \bigcup_{j=1}^{\ell-1} U_{c,j} \tag{13}$$

Die Anzahl der in den verschiedenen Iterationsschritten identifizierten Unternehmen für die WS2009 ist in Tabelle 4 gegeben.

Tabelle 4 Anzahl als einzigartig identifizierte Unternehmen in der WS2009.

Menge	Anzahl
Population	149 011
Take-All	7 778
U_1	141
U_2	32
U_3	5
U_4	1
U	179

2.4.2 Illustration - Einfluss auf Varianz

Annahme: Ohne besonderen Effort ist die Antwortrate in TA_c gleich $\alpha < 1$. Es bezeichnet

- $\mathbf{Var}_1(\hat{X}_{TA_c})$ = Varianz des HT-Schätzer in TA_c
- $\mathbf{Var}_2(\hat{X}_{TA_c})$: analog, aber Behandlung der N_{U_c} Unternehmen in U_c als separate Schicht mit Antwortrate 100%

Somit erhalten wir

$$\mathbf{Var}_1(\hat{X}_{TA_c}) = \frac{N_{TA_c}^2}{\alpha N_{TA_c}}(1 - \frac{\alpha N_{TA_c}}{N_{TA_c}})S_{x,TA_c}^2 \tag{14}$$

$$\mathbf{Var}_2(\hat{X}_{TA_c}) = \frac{(N_{TA_c} - N_{U_c})^2}{\alpha(N_{TA_c} - N_{U_c})}(1 - \frac{\alpha(N_{TA_c} - N_{U_c})}{(N_{TA_c} - N_{U_c})})S_{x,TA_c \setminus U_c}^2 \tag{15}$$

S_{x,TA_c}^2 bezeichnet dabei die Varianz von x in TA_c. Die relative Varianzreduktion in TA_c ist:

$$1 - \frac{\mathbf{Var}_2(\hat{X}_{TA_c})}{\mathbf{Var}_1(\hat{X}_{TA_c})}$$

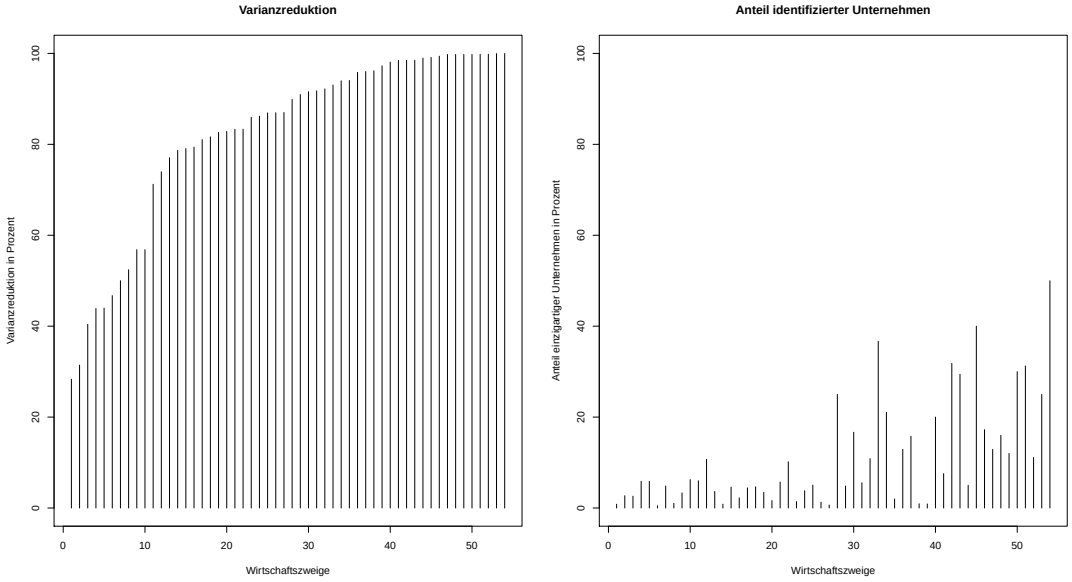
wobei aufgrund von (14) und (15) gilt:

$$\frac{\mathbf{Var}_2(\hat{X}_{TA_c})}{\mathbf{Var}_1(\hat{X}_{TA_c})} = \frac{(N_{TA_c} - N_{U_c})}{N_{TA_c}} \frac{S_{x,TA_c \setminus U_c}^2}{S_{x,TA_c}^2} \tag{16}$$

Die relative Varianzreduktion hängt also nicht von der Antwortrate α ab. In Abbildung 2 sind die in verschiedenen Wirtschaftszweigen erzielten Varianzreduktionen den jeweiligen Anteilen als einzigartig identifizierter Unternehmen gegenübergestellt. Die Rechnungen beruhen auf dem Stichprobenrahmen der WS2009.

Interpretation: In Abbildung 2 erkennt man, dass in vielen NOGAs mit einem kleinen Anteil identifizierter Unternehmen beträchtliche Varianzreduktionen erzielt werden können. Auffallend ist auch, dass es Branchen gibt, in welchen der Anteil identifizierter Unternehmen gross ist. Untersuchungen zeigen, dass es sich hierbei um kleine Wirtschaftszweige handelt, in welchen eben ein Unternehmen schnell einen Einfluss von 2.5% auf die Schätzung haben kann. Möchte man verhindern, dass in kleinen Wirtschaftszweigen relativ viele Unternehmen identifiziert werden, so empfiehlt es sich, die uniforme Handhabung der 2.5% Grenze etwas aufzuweichen und sie bei kleinen Wirtschaftszweigen heraufzusetzen.

Abbildung 2 Anteil identifizierter Unternehmen vs. Varianzreduktion in der Take-All Schicht



2.4.3 Einsetzungen bei den identifizierten Unternehmen

In wenigen Fällen kann es sein, dass trotz grossem Effort ein als einzigartig identifiziertes Unternehmen keine Antwort liefert. In solchen Fällen wird durch MON basierend auf gezielt gesammelter Hilfsinformation (Sekundärdatenquellen, Vorjahreswerte, Mehrwertsteuerangaben), eine fallspezifische Einsetzung vorgenommen. Im Rahmen der WS09 mussten 6 solche unternehmensspezifische Einsetzungen vorgenommen werden.

3 Die Hochrechnung der Resultate

Sämtliche verwendeten Schätzungen sind gewichtete Summen der beobachteten Zielvariablen. Gewichtet wird dabei mit dem Extrapolationsgewicht. Dieses wird ausgehend vom Stichprobengewicht in drei Etappen konstruiert:

1. Korrektur für Antwortausfälle
2. Robustifizierung
3. Kalibrierung

Im Folgenden werden die Konstruktionsschritte beschrieben, welche in den Grundzügen mit der Methodik der Vorjahre [7] übereinstimmt, die ihrerseits auf dem One-Step Ratio Estimator gemäss [2] beruht. Dabei wird die in Tabelle 5 aufgeführte Notation verwendet.

3.1 Korrektur für Antwortausfälle

Ausgehend von der Schichtung der Unternehmen für die Stichprobenziehung (vgl. [9]) wurden Antworthomogenitätsklassen gebildet, für welche man annimmt, dass die darin enthaltenen

Tabelle 5 Notation für die Konstruktion der Hochrechnungsgewichte.

Term	Bedeutung
n_h	Bruttostichprobengrösse in Antworthomogenitätsklasse h
m_h	Nettostichprobengrösse in Antworthomogenitätsklasse h (Anzahl Antwortende)
N_h	Populationsgrösse in Antworthomogenitätsklasse h
$\theta_h = \frac{m_h}{n_h}$	Antwortrate in h

Unternehmen eine einheitliche Antwortwahrscheinlichkeit aufweisen. Die Homogenitätsklassen wurden mittels $NOGA_{NUM} \times clt2$ gebildet. Dabei basiert $NOGA_{NUM}$ auf dem NOGA2 Code des Rahmens, wobei die NOGAs $8, 9$ sowie $38, 39$ aufgrund ihrer geringen Grösse zusammengelegt worden sind ($8, 9 \rightarrow 8$, $38, 39 \rightarrow 38$). Die Grössenkategorisierung clt2 wurde basierend auf den in Tabelle 6 dargestellten und bei der Ausarbeitung des Stichprobenplanes verwendeten Grössenklassen clt wie folgt konstruiert:

Tabelle 6 Definition von Grössenklassen aufgrund der Anzahl Beschäftigten.

betot	clt
3-7	1
8-19	2
20-49	3
50-249	4
250-	5

```
clt2=clt;
if noga_num in (38,61,94) and clt in (4,5)    then clt2 =4;
if noga_num  = 12         and clt in (1,2,3)  then clt2 =1;
if noga_num  = 36         and clt in (2,3,4)  then clt2 =2;
if noga_num  = 91         and clt in (3,4)    then clt2 =3;
if noga_num  = 92         and clt in (2,3)    then clt2 =2;
if noga_num  = 19         and clt in (1,2)    then clt2 =1;
```

Wie man in [9], Seite 12, verifiziert, betreffen diese Zusammenlegungen nur Unternehmen mit einem Stichprobengewicht von 1. Im Rahmen der WS09 wurden die clt-Klassen nur zusammengelegt, falls weniger als zwei antwortende Unternehmen vorhanden wären und somit die Varianz in der Klasse nicht mehr geschätzt werden könnte. Dem Entscheid für dieses zurückhaltende Zusammenlegen lagen Überlegungen zu Grunde, die im Folgenden kurz erklärt werden:

- Die Fälle mit geringer Zellbelegung (Anzahl antwortende Unternehmen m_h klein) betreffen vorallem die Grössenklassen clt=4/5. Diese Klassen sind vollerhoben und besitzen in den Fällen m_h klein auch einen kleinen Bruttostichprobenumfang n_h. Die aus den Antwortraten $\theta_h = \frac{m_h}{n_h}$ hervorgehenden Gewichtskorrekturen sowie die resultierenden Gewichte bewegen sich darum meist in einem moderaten Rahmen. In verschiedenen Fällen (Bsp $noga_{num}$ 93, clt2=5) ist θ sogar 1, was im Rahmen der Hochrechnung einer vollerhobenen Schicht mit Antwortrate 100% entspricht, die keinen Beitrag zur Varianz der Schätzungen liefert.

- Die Klasse clt=5 enthält Unternehmen, die bezüglich betot wesentlich grösser sein können als die Unternehmen in clt=4. Die vorgenommene Korrektur für Antwortausfälle entspricht einer Einsetzung mittels dem Zellmittelwert der Antwortenden. Es darf davon aus-

Tabelle 7 Nettostichprobengrössen m_h für $h \in noga_{num}$ x clt2. (Abk. $noga_{num}$ =nog)

	clt2								clt2						
nog	1	2	3	4	5	6	Total	nog	1	2	3	4	5	6	Total
8	27	22	41	12	.	1	103	50	7	4	16	4	.	4	35
10	24	13	89	109	36	4	275	51	5	5	10	11	.	2	33
11	7	8	24	14	.	4	57	52	13	17	36	69	22	5	162
12	3	.	.	.	.	5	8	53	10	11	8	8	9	3	49
13	9	7	34	37	.	2	89	55	16	13	19	196	13	.	257
14	35	8	5	6	.	5	59	56	111	17	17	115	23	.	283
15	16	10	7	4	.	3	40	58	10	9	22	33	9	3	86
16	22	21	28	54	4	.	129	59	12	14	34	10	.	.	70
17	11	14	22	31	11	3	92	60	5	7	22	9	.	3	46
18	14	13	25	67	11	1	131	61	9	12	18	19	.	4	62
19	3	.	1	.	.	3	7	62	12	41	23	122	22	.	220
20	6	14	61	79	17	6	183	63	8	10	16	13	.	4	51
21	9	15	21	34	19	5	103	68	262	52	77	36	4	.	431
22	15	9	29	89	12	1	155	69	48	19	19	44	4	3	137
23	20	7	22	61	6	2	118	70	10	12	28	79	15	2	146
24	8	14	28	35	17	3	105	71	39	32	30	133	10	.	244
25	23	9	39	236	27	.	334	72	13	12	31	25	5	5	91
26	6	28	37	226	67	5	369	73	13	9	76	37	9	6	150
27	10	10	49	57	32	4	162	74	52	22	19	11	.	.	104
28	16	10	29	243	53	2	353	75	130	40	7	.	.	.	177
29	14	18	16	13	.	7	68	77	14	11	15	12	.	3	55
30	9	4	8	9	2	5	37	78	13	7	16	18	2	2	58
31	15	10	20	44	5	.	94	79	16	10	16	24	2	2	70
32	24	17	23	34	9	2	109	80	14	5	17	20	4	2	62
33	20	18	18	27	.	4	87	81	36	11	32	94	22	3	198
35	10	29	80	55	17	4	195	82	12	12	21	36	5	1	87
36	12	15	.	.	.	2	29	85	12	11	39	185	21	.	268
37	20	18	16	5	.	2	61	87	25	22	59	402	22	.	530
38	10	17	31	23	.	1	82	88	70	45	45	32	1	.	193
41	13	6	17	226	22	.	284	90	16	13	18	15	6	3	71
42	13	10	36	78	11	1	149	91	12	9	4	.	.	3	28
43	66	34	16	220	21	.	357	92	3	3	.	6	.	11	23
45	61	17	16	84	12	1	191	93	22	32	45	37	2	.	138
46	45	25	756	477	72	.	1375	94	26	34	22	28	.	1	111
47	146	38	10	239	83	3	519	95	29	13	8	3	.	5	58
49	19	14	19	162	20	2	236	96	116	14	13	29	2	.	174
								All	2002	1132	2571	5005	820	173	11703

gegangen werden, dass dieser Zellmittelwert zumindest in Bezug auf betot gerade zwischen den Klassen clt=4 und clt=5 stark unterschiedlich sein kann.

Für einzigartige Unternehmen im Sinne von (13), wurde im Rahmen der Datenaufbereitung / Datenerhebung gezielt eine Antwortrate von 100% angestrebt (bis auf wenige Problemfälle in zwei kleinen Wirtschaftszweigen waren diese Bestrebungen im Rahmen der WS09 erfolgreich). Für diese Unternehmen wird im Rahmen der Gewichtung $clt2 = 6$ gesetzt und sie werden als vollerhobene Schicht mit Antwortrate 100% behandelt und von sämtlichen weiteren Varianzberechnungen / Kalibrierungsschritten ausgeschlossen. Tabelle 7 enthält eine Übersicht der Nettostichprobengrössen der Wertschöpfungsstatistik in den Klassen $noga_{num} \times clt2$.
Tabelle 8 enthält eine Übersicht der berechneten Antwortraten θ_h in den für die Gewichtung relevanten (ohne $clt2 = 6$) Klassen. θ_h^{-1} wird im Folgenden mit dem Ausgangsgewicht

$$w_{1,i} = \frac{N_h}{n_h} \tag{17}$$

multipliziert, was für sämtliche Unternehmen i in h auf das an Antwortausfälle angepasste Gewicht $w_{2,i}$ führt:

$$w_{2,i} = \frac{w_{1,i}}{\theta_h} = \frac{N_h}{m_h} \tag{18}$$

Tabelle 8 Antwortraten θ_h nach $noga_{num}$ x clt2. (Abk. $noga_{num}$ =nog)

	clt2						clt2				
nog	1	2	3	4	5	nog	1	2	3	4	5
8	0.49	0.67	0.77	0.75	.	50	0.19	0.19	0.7	1	.
10	0.48	0.25	0.58	0.8	0.92	51	0.14	0.24	0.48	0.69	.
11	0.18	0.31	0.75	0.88	.	52	0.3	0.35	0.71	0.9	0.88
12	0.6	.	.	.	.	53	0.22	0.29	0.38	0.67	0.9
13	0.29	0.24	0.72	0.86	.	55	0.3	0.37	0.59	0.76	0.72
14	0.28	0.21	0.33	0.55	.	56	0.31	0.24	0.57	0.71	0.88
15	0.34	0.4	0.7	0.67	.	58	0.23	0.38	0.79	0.87	1
16	0.44	0.34	0.56	0.78	1	59	0.22	0.47	0.65	0.56	.
17	0.27	0.36	0.73	0.86	0.92	60	0.31	0.44	0.85	1	.
18	0.38	0.38	0.71	0.8	0.85	61	0.24	0.41	0.72	0.9	.
19	0.5	.	1	.	.	62	0.25	0.26	0.77	0.78	0.92
20	0.2	0.5	0.73	0.85	0.85	63	0.2	0.32	0.64	0.87	.
21	0.31	0.48	0.62	0.81	1	68	0.3	0.38	0.77	0.82	0.67
22	0.38	0.39	0.63	0.86	0.8	69	0.42	0.42	0.66	0.83	0.67
23	0.42	0.37	0.69	0.91	0.75	70	0.26	0.4	0.8	0.77	0.83
24	0.21	0.44	0.7	0.83	1	71	0.32	0.51	0.88	0.88	0.91
25	0.4	0.38	0.68	0.78	1	72	0.26	0.3	0.67	0.71	0.83
26	0.22	0.31	0.74	0.74	0.87	73	0.29	0.28	0.75	0.86	0.9
27	0.29	0.48	0.64	0.83	1	74	0.34	0.35	0.7	0.85	.
28	0.42	0.33	0.63	0.76	0.84	75	0.56	0.57	0.7	.	.
29	0.33	0.47	0.64	0.87	.	77	0.33	0.31	0.75	0.8	.
30	0.24	0.2	0.5	0.9	1	78	0.33	0.25	0.52	0.58	1
31	0.35	0.3	0.54	0.85	1	79	0.3	0.33	0.55	0.77	0.67
32	0.43	0.44	0.72	0.72	0.82	80	0.31	0.22	0.5	0.59	1
33	0.44	0.47	0.72	1	.	81	0.49	0.38	0.54	0.53	0.69
35	0.5	0.73	0.88	0.83	1	82	0.29	0.4	0.62	0.75	0.83
36	0.44	0.56	.	.	.	85	0.29	0.34	0.75	0.75	0.91
37	0.48	0.46	0.64	1	.	87	0.49	0.51	0.71	0.78	0.85
38	0.32	0.38	0.66	0.82	.	88	0.59	0.68	0.71	0.71	1
41	0.34	0.22	0.4	0.66	0.92	90	0.37	0.42	0.67	0.79	1
42	0.33	0.3	0.69	0.83	0.92	91	0.43	0.47	0.57	.	.
43	0.27	0.33	0.42	0.64	0.91	92	0.38	0.6	.	1	.
45	0.42	0.4	0.64	0.8	0.92	93	0.29	0.39	0.56	0.8	1
46	0.3	0.48	0.78	0.83	0.93	94	0.57	0.74	0.67	0.78	.
47	0.39	0.44	0.42	0.67	0.91	95	0.39	0.45	0.73	0.75	.
49	0.43	0.36	0.73	0.81	0.91	96	0.51	0.27	0.57	0.83	0.67

Das Gewicht $w_{2,i}$ entspricht dem Ziehungsgewicht einer einfachen Zufallsstichprobe von m_h aus N_h Unternehmen und genauso werden die antwortenden Unternehmen in den weiteren Berechnungen (insbesondere im Rahmen der Varianzschätzung) auch behandelt.

3.2 Robustifizierung

Im Folgenden wird für jede Beobachtung i ein Faktor $u_i \in (0;1]$ hergeleitet, mit dem das Gewicht $w_{2,i}$ im Sinne einer Robustifizierung multipliziert wird. Um zu beurteilen, ob ein Unternehmen extreme Werte aufweist und eine entsprechende Robustifizierungskonstante u_i zu berechnen, wurde es mit Unternehmen desselben Wirtschaftszweiges verglichen. Grundsätzlich beruht die Definition des Wirtschaftszweiges dabei auf der NOGA Stufe 2 gemäss Stichprobenrahmen (Variable noga2_08). Die Ausreisseridentifikation basiert auf den Residuen einer linearen Regression ohne Intercept und ist nur sinnvoll, falls eine gewisse Anzahl an Beobachtungen vorhanden sind. NOGAs mit weniger als 15 Beobachtungen wurden darum entsprechend folgendem SAS-Code mit benachbarten NOGAs zusammengefasst. Die resultierenden Aggregate werden in der Variable $noga_{prov}$ codiert. Die Grenze von 15 beruht auf der Überlegung, dass mindestens 15 Beobachtungen zu Verfügung stehen sollten, um den Steigungsparameter sowie die Standardabweichung der Residuen zu schätzen.

```
noga_prov = noga2_08;
  if noga2_08 in (8,9)   then noga_prov = 0809;
  if noga2_08 in (11,12) then noga_prov = 1112;
```

```
if noga2_08 in (19,20) then noga_prov = 1920;
if noga2_08 in (38,39) then noga_prov = 3839;
```

Die Berechnung des Robustifizierungsgewichtes wird im Folgenden für die Unternehmen innerhalb eines Wirtschaftszweiges $c \in noga_{prov}$ vorgestellt. Wir betrachten für Unternehmen $i \in c$ folgende Regressionsgleichung:

$$y_i = \beta_c x_i + \epsilon_i \tag{19}$$

mit

- $y_i =$ Zielvariable
- $x_i =$ Vollzeitstellenäquivalent gemäss Stichprobenrahmen
- ϵ_i=Residuum mit $\mathbf{E}(\epsilon_i) = 0$ und $\mathbf{Var}(\epsilon_i) = \sigma_c^2 x_i$. Die Varianz ist somit proportional zum Vollzeitstellenäquivalent. Eine ausführliche Besprechung / Diskussion zu dieser Hypothese ist in [7] gegeben.

Nun werden anhand der antwortenden Unternehmen in c, r_c, robuste Schätzungen der Parameter β_c und σ_c berechnet[10]:

$$\hat{\beta}_c := \frac{\mathbf{Med}_{i \in r_c}(y_i)}{\mathbf{Med}_{i \in r_c}(x_i)} \tag{20}$$

$$e_i^{(s)} := \frac{(y_i - \hat{\beta}_c x_i)}{\sqrt{x_i}} \tag{21}$$

$$\hat{\sigma}_c := \mathbf{Med}_{i \in r_c}(|e_i^{(s)}|) \frac{1}{0.67} \tag{22}$$

Der Robustifizierungsfaktor für Unternehmen i wird nun wie folgt berechnet:

$$u_i = \begin{cases} \frac{k\hat{\sigma}_c}{|e_i^{(s)}|}; & \text{falls } |e_i^{(s)}| > k\hat{\sigma}_c \\ 1; & \text{falls } |e_i^{(s)}| \leq k\hat{\sigma}_c \end{cases} \tag{23}$$

k legt die Stärke der Robustifizierung fest und wurde in Anlehnung an die Vorjahre mit $k = 10$ festgelegt. Es ist nun so, dass die Methode je nach betrachteter Zielvariable y ein anderes u liefert. Das Problem wurde so gehandhabt, dass die vorgestellte Berechnung von u für die vier Erhebungsvariablen

- t017: Bruttoproduktion
- t022: Vorleistungen
- t053: Personalkosten
- s015: Vollzeitstellenäquivalent gemäss Erhebung[11]

[10] Analog zu den Vorgängererhebungen werden ungewichtete Parameterschätzungen verwendet. Eine Begründung zu dieser Abweichung zum Vorgehen in [2] wird in [7] gegeben.

[11] Es kann vorkommen, dass ein Unternehmen im Stichprobenrahmen kleine Beschäftigtenzahlen aufweist und sich bei der Erhebung dann herausstellt, dass die effektive Beschäftigtenzahl für das Erhebungsjahr viel höher ist. Dies kann dazu führen, dass ein grosses Unternehmen ein grosses Hochrechnungsgewicht erhält. Um dies zu verhindern, wird das Vollzeitstellenäquivalent der Erhebung in die Robustifizierung einbezogen. Extreme Veränderungen der Beschäftigtenzahlen führen somit zu einer Reduktion des Hochrechnungsgewichtes.

vorgenommen wurde. Dies liefert dann für jedes Unternehmen i vier Werte $u_{1,i}, ..., u_{4,i}$. Analog dem Verfahren für die Vorjahre wird mit dem Minimalwert gearbeitet:

$$u_i^* := \min(u_{1,i}, ..., u_{4,i}) \tag{24}$$

Mittels dem u_i^*-Faktor wird für jedes Unternehmen das Gewicht nach Robustifizierung berechnet:

$$w_{3,i} = \max(u_i^* w_{2,i}, 1) \tag{25}$$

Aus der Definition wird klar, dass der Einfluss der Robustifizierung in jenem Sinn eingeschränkt wird, dass die Gewichte nicht unter Eins fallen. Diese Beschränkung wurde für die WS09 eingeführt. Sie bewirkt, dass

- die korrigierten Gewichte w_3 näher bei w_2 liegen
- die mit w_3 gewichtete Summe der Unternehmen für sämtliche positiven Zielvariablen sicherlich über der ungewichteten Summe liegt: $\sum_{i \in r_c} w_{3,i} y_i \geq \sum_{i \in r_c} y_i$. (Dies ist mit einer unbeschränkten Robustifizierung nicht der Fall, was in der Vergangenheit teilweise zu unbefriedigenden Schätzungen geführt hatte.)

Ein weiterer Grund für die Beschränkung der Robustifizierung liegt im Bereich der Datenaufbereitung: Werte von Unternehmen die auf $u_i^* w_{2,i} < 1$ führen, werden durch die Sektion nochmals prioritär auf ihre Richtigkeit geprüft. Die für die definitive Hochrechnung verwendeten Daten der betreffenden Unternehmen werden somit als geprüft / korrekt betrachtet und es ist nicht abwegig zu Verlangen, dass sich Unternehmen im Rahmen der Hochrechnung zumindest selbst repräsentieren (d.h. ein Gewicht Eins haben). Diese Handhabung gewährt auch eine gewisse Konsistenz mit der Handhabung jener Unternehmen, die aufgrund ihres betot - Wertes im Rahmen der Datenaufbereitung als einzigartig identifiziert und in der Hochrechnung mit Eins gewichtet werden.

Auswirkungen auf die Schätzung: Tests liessen darauf schliessen, dass diese scheinbar kleine Änderung die Hochrechnungsresultate relativ stark beeinflussen kann. Dabei zeigte sich eine tendenzielle Erhöhung der Resultate. Diese kann damit erklärt werden, dass die extremen Residuen, welche zu einer Robustifizierung führen, mehrheitlich positiv sind (etwas rechtsschiefe Verteilung der Residuen) und somit führt die Robustifizierung tendenziell eher zu einer Unterschätzung der Resultate. Durch die Bedingung, dass die Gewichte nicht unter Eins fallen, wird dieser Effekt reduziert, was heisst, dass sich die Resultate gegenüber der ursprünglichen Methode tendenziell erhöhen. Andererseits ist es so, dass aufgrund der Beschränkung der Robustifizierung eine Erhöhung der Varianz beobachtet wird. Eine gleichzeitige Reduktion von Varianz und Verfälschung könnte allenfalls mit sorgfältig gewählten asymmetrischen Robustifizierungslimiten erreicht werden, welche die Schiefe der Residuen einbeziehen. Dies Bedarf aber eines weiteren Studiums der Problematik.

3.3 Kalibrierung

Nun erfolgt eine Kalibrierung mittels der Hilfsvariablen Vollzeitstellenäquivalent im Sinne eines Quotientenschätzers auf Niveau der im Zuge der Robustifizierung präsentierten NOGA2-Aggregate $noga_{prov}$. Für das Unternehmen $i \in c$, mit $c \in noga_{prov}$, ergibt sich durch die Kalibrierung das definitive Hochrechnungsgewicht:

$$w_{4,i} = g_i w_{3,i} \tag{26}$$

Tabelle 9 Illustration zum Einfluss der Robustifizierung: Anzahl Unternehmen deren Gewicht durch die Robustifizierung beeinflusst wird (rob=1).

	rob	
clt2	0	1
1	1964	38
2	1100	32
3	2434	137
4	4652	353
5	680	140
Total	10830	700

Dabei ist

$$g_i = \frac{X_{c\setminus U_c}}{\hat{X}_{c\setminus U_c}} = \frac{X_{c\setminus U_c}}{\sum_{j\in r_c\setminus U_c} w_{3,j}x_j}, \tag{27}$$

wobei $r_c \setminus U_c$ die Nettostichprobe (ohne die einzigartigen Unternehmen U_c) und $X_{c\setminus U_c}$ das entsprechende Total der Variablen Vollzeitstellenäquivalent in Branche c bezeichnet.

3.4 Die Totalschätzung

Die Schätzung für das Gesamttotal der Variablen y entspricht der gewichteten Summe aller beobachteten y-Werte:

$$\hat{Y} = \sum_{i\in r} w_{4,i}y_i, \tag{28}$$

Dieser Schätzer entspricht einer etwas vereinfachten Variante des in [3] beschriebenen robusten Regressionsschätzers und kann als robustifizierter Quotientenschätzer betrachtet werden. Die Schätzung eines Totals für einen Untersuchungsbereich[12] (Domain) d erfolgt mittels

$$\hat{Y}_d = \sum_{i\in r} 1(i \in d)w_{4,i}y_i, \tag{29}$$

wobei r die Menge der antwortenden Unternehmen und $1(i \in d)$ die Indikatorfunktion für die Zugehörigkeit zum Domain d bezeichnet:

$$1(i \in d) = \begin{cases} 1 & \text{falls } i \in d \\ 0 & \text{falls } i \notin d \end{cases} \tag{30}$$

3.4.1 Die Varianzschätzung

Die Kalibrierung erfolgte auf Niveau der Branchen gemäss $noga_{prov}$, welche im Allgemeinen der NOGA2 gemäss Stichprobenrahmen entsprechen. Darum wird das dem Quotientenschätzer zugrunde liegende Regressionsmodell für jede Branche $c \in noga_{prov}$ separat betrachtet. Das heisst für $i \in c$ wird angenommen, dass

$$y_i = \beta_c x_i + \epsilon_i;\ \mathbf{E}(\epsilon_i) = 0,\ \mathbf{Var}(\epsilon_i) = x_i\sigma^2 \tag{31}$$

gilt. Der erste Schritt der verwendeten Varianzschätzung besteht darin, für sämtliche Unternehmen in $r_c \setminus U_c$ die Residuen ϵ_i in (31) zu schätzen. Dies geschieht mittels:

$$\hat{e}_i := y_i - \hat{\beta}_c x_i \tag{32}$$

[12] Branchen gemäss NOGA2 oder NOGA3 sind Beispiele solcher Untersuchungsbereiche.

wobei

$$\hat{\beta}_c = \frac{\sum_{r_c \setminus U_c} w_{3,i} y_i}{\sum_{r_c \setminus U_c} w_{3,i} x_i} \tag{33}$$

Die so berechneten Residuen verwenden wir nun, um die Varianz von $\hat{Y}$ aus (28) zu schätzen:

$$\widehat{\mathbf{Var}}(\hat{Y}) = \sum_{h \in H} \frac{N_h^2}{m_h}(1 - \frac{m_h}{N_h}) \frac{1}{m_h - 1} \sum_{i \in r_h} (u_i g_i \hat{e}_i - \frac{\sum_{i \in r_h} u_i g_i \hat{e}_i}{m_h})^2 \tag{34}$$

wobei H die Antworthomogenitätsklassen $noga_{num} \times clt2$ und r_h die Menge der antwortenden Unternehmen in $h \in H$ bezeichnet. Die Varianz (34) beruht auf der Varianzformel für eine geschichtete einfache Zufallsstichprobe. Eine Begründung hierzu findet man in [4], Seite 138. Um dem Effekt der Kalibrierung Rechnung zu tragen, werden in der Varianzschätzung die Residuen gemäss (32) verwendet (vgl. [1], Seite 246). Im Fall des Schätzers für das Total eines Untersuchungsbereiches (29) wird die Varianz mit

$$\widehat{\mathbf{Var}}(\hat{Y}_d) = \sum_{h \in H} \frac{N_h^2}{m_h}(1 - \frac{m_h}{N_h}) \frac{1}{m_h - 1} \sum_{i \in r_h} (u_i g_i 1(i \in d) \hat{e}_i - \frac{\sum_{i \in r_h} u_i g_i 1(i \in d) \hat{e}_i}{m_h})^2 \tag{35}$$

geschätzt.

Diskussion zu (35): In Fällen, in welchen der Untersuchungsbereich d nicht einer oder einer Vereinigung mehrerer Branchen $c \in noga_{prov}$ entspricht, liefert (35) eine Schätzung basierend auf folgenden Hypothesen:

- der Schätzer in (29) ist identisch zum Schätzer, der aus einer Kalibrierung auf Niveau $noga_{prov} \times d$ (Branche gekreuzt mit dem Untersuchungsbereich) hervorgehen würde[13].
- die Regressionsgleichung (31) gilt unabhängig davon, ob ein Unternehmen im Untersuchungsbereich d ist oder nicht. Das heisst, die Zugehörigkeit zum Untersuchungsbereich hat keinen Einfluss auf den Regressionskoeffizienten.

Die Hypothesen sind kaum immer erfüllt und die resultierenden Varianzschätzungen Approximationen. Es ist aber zu beachten, dass viele wichtige Untersuchungsbereiche effektiv (z.B. Gesamttotal, NOGA2 gemäss Rahmen) oder nahezu (NOGA2 gemäss Erhebung) einer oder mehrerer Branchen gemäss $noga_{prov}$ entsprechen und in diesen Fällen sind die formulierten Bedingungen ganz oder nahezu erfüllt.

Ein alternativer Ansatz, der ungeachtet des betrachteten Untersuchungsbereiches ohne die gemachten Hypothesen auskäme, könnte wie folgt aussehen: Für jedes Domain wird eine separate Zielvariable $y_{d,i}$ mit folgenden Werten geschaffen:

$$y_{d,i} = \begin{cases} y_i \text{ falls } i \in d \\ 0 \text{ falls } i \notin d \end{cases} \tag{36}$$

Die Berechnung für die zur Varianzschätzung benötigten Residuen führt nun für $i \in c$ auf

$$\hat{e}_{d,i} := y_{d,i} - \hat{\beta}_{c,d} x_i,$$

[13] Aufgrund von (27) entspricht dies der Annahme

$$\frac{X_{(c \cap d) \setminus U_c}}{\hat{X}_{(c \cap d) \setminus U_c}} = \frac{X_{c \setminus U_c}}{\hat{X}_{c \setminus U_c}}$$

mit $\hat{X}_{(c \cap d) \setminus U_c} = \sum_{i \in r} w_{3,i} x_i 1(i \in (c \cap d) \setminus U_c)$.

mit

$$\hat{\beta}_{c,d} = \frac{\sum_{r_c \setminus U_c} w_{3,i} y_{d,i}}{\sum_{r_c \setminus U_c} w_{3,i} x_i},$$

und diese Residuen würden direkt in (34) eingesetzt. Das Vorgehen hat also den Preis, dass für jede Kombination Untersuchungsbereich x Variable separate Residuen berechnet werden müssten, womit sich der Rechenaufwand stark vergrössert. Für die Hochrechnung der WS2009 war diese Variante noch nicht implementiert, aber entsprechende Bemühungen und Tests sind im Gange.

3.4.2 Einige Hochrechnungsbeispiele

Die Präzision der Hochrechnungsresultate hängt wie erwartet werden durfte von der betrachteten Zielvariablen und dem Untersuchungsbereich ab. Als Illustration sind in Tabelle 10 die Hochrechnungsresultate für Totale der Variablen

- t017: Bruttoproduktion
- t022: Vorleistungen
- t053: Personalkosten
- t017-t022: Wertschöpfung

auf Niveau NOGA 2 dargestellt. Nebst den Totalen sind auch die Nettostichprobengrösse (n) und der Variationskoeffizient der Schätzung (CV) aufgelistet. Es ist zu beachten, dass es sich bei diesen Ergebnissen um Testhochrechnungen zum Zwecke der Methodenillustration handelt, welche von den effektiven Hochrechnungsresultaten gemäss Publikation abweichen können. Betrachten wir die Variationskoeffizienten, so fällt Folgendes auf:

- Die Hochrechnung der Personalkosten (t053) liefert auf Niveau NOGA2 generell Variationskoeffizienten, welche um die gemäss Stichprobenplan angestrebten 2.5% herum streuen. (Median = 2.5%). Gewisse NOGAs führen also auf einen CV von über 2.5%, aber generell befinden sich die Genauigkeiten im angestrebten Bereich. Die zufriedenstellenden Resultate können damit erklärt werden, dass die Personalkosten relativ gut mit den Beschäftigungsvariablen betot und Vollzeitstellenäquivalent korrelieren, die als Hilfsvariablen zur Konstruktion von Stichprobenplan und Schätzung verwendet wurden.
- Die Hochrechnung der Vorleistungen t022 liefert tendenziell die höchsten Variationskoeffizienten (Median 6.7%). Dies ist dahingehend zu interpretieren, dass die Variable Vorleistungen tendenziell stark streut und weniger gut mit den verwendeten Hilfsvariablen zur Beschäftigtenzahl korreliert. Die Situation für die Bruttoproduktion t017 ist tendenziell ein bisschen besser, aber auch hier resultiert mit 4.6% ein Median, der über den angepeilten 2.5% liegt.
- Die Variationskoeffizienten, welche für die Bruttowertschöpfung resultieren, sind gegenüber denjenigen für t017 und t022 tendenziell tiefer. Dies ist plausibel, weil t017 und t022 positiv korrelieren, was einen varianzreduzierenden Effekt auf t017-t022 hat.
- Die Präzisionsbetrachtungen im Rahmen der Ausarbeitung des Stichprobenplanes [9] wiesen darauf hin, dass die erzielten Variationskoeffizienten sehr sensibel auf Antwortausfälle bei den grossen, vollerhobenen Unternehmen reagieren. Aus Tabelle 8 entnimmt man, dass die Antwortraten in den Klassen $clt2 = 4, 5$ zwar hoch, aber in den meisten Fällen nicht 100% sind. Die beobachteten Präzisionen liegen also durchaus in den erwarteten Bereichen.

Tabelle 10 Hochrechnungsbeispiele auf Niveau NOGA2

		t017		t022		t053		t017-t022	
noga2	N	Sum	cv %	Sum	cv %	Sum	cv %	Sum	cv %
8	102	2 020 619	3,9	1 171 559	4,9	420 171	1,9	849 060	4,0
10	270	25 687 335	2,2	19 448 783	3,2	4 175 180	0,8	6 238 552	2,0
11	58	3 500 986	4,6	2 301 077	5,7	678 618	1,7	1 199 910	3,6
12	8	5 553 890	0,1	4 042 441	0,1	527 899	0,2	1 511 449	0,1
13	88	1 847 192	3,6	1 038 426	6,0	646 469	1,4	808 766	2,8
14	58	1 060 661	5,8	772 402	6,8	239 650	3,4	288 258	3,7
15	39	517 643	2,9	376 797	3,4	129 029	2,3	140 846	2,8
16	128	8 297 625	4,8	4 992 616	6,5	2 507 636	3,8	3 305 009	3,6
17	89	3 988 938	2,7	2 614 303	3,6	974 517	0,7	1 374 635	1,5
18	130	5 147 882	2,5	2 587 017	3,2	2 072 816	2,3	2 560 865	2,5
19	7	1 945 237	1,4	1 827 919	1,2	79 288	1,8	117 318	4,1
20	176	19 981 646	4,3	13 909 721	4,9	3 811 800	0,8	6 071 926	3,6
21	103	64 569 132	3,8	47 003 997	3,8	5 609 317	0,8	17 565 134	5,7
22	160	7 811 730	2,6	4 796 128	3,7	2 068 373	1,3	3 015 603	2,0
23	117	6 525 113	4,2	3 824 377	5,8	1 685 713	1,3	2 700 736	2,7
24	104	4 654 026	1,6	3 084 615	2,1	1 259 732	0,7	1 569 411	1,6
25	334	18 791 231	4,4	10 012 848	6,4	6 804 192	2,0	8 778 382	2,9
26	371	49 216 702	6,9	31 262 785	9,3	10 898 834	1,0	17 953 917	3,0
27	155	21 325 647	1,3	15 926 369	1,3	3 600 229	0,5	5 399 277	1,3
28	352	25 947 592	2,8	16 117 213	2,8	8 015 565	1,6	9 830 379	3,6
29	67	1 931 912	3,6	1 356 550	4,9	448 456	0,9	575 362	1,5
30	39	3 805 278	2,2	2 695 346	1,8	1 168 919	1,3	1 109 932	3,3
31	94	3 374 019	5,8	2 069 441	8,3	1 016 994	2,3	1 304 578	3,0
32	106	5 504 620	7,4	2 979 790	8,6	1 521 457	3,4	2 524 831	6,3
33	87	4 660 475	5,6	2 993 239	7,3	1 278 054	2,6	1 667 237	4,2
35	197	44 599 497	7,2	33 285 175	8,8	3 245 045	0,9	11 314 322	3,9
36	29	608 493	5,4	337 147	5,9	101 954	4,7	271 346	6,6
37	61	714 597	6,5	371 522	7,9	239 021	2,0	343 076	9,5
38	82	3 962 399	7,2	2 444 868	10,6	829 298	3,7	1 517 531	8,2
39	2	311 759	78,7	238 235	84,4	32 859	39,6	73 524	60,2
41	281	21 206 324	7,4	12 949 097	11,0	6 983 856	2,2	8 257 227	2,7
42	148	5 961 731	2,7	3 125 342	4,0	2 397 049	1,2	2 836 389	1,5
43	354	34 371 957	3,7	17 352 671	5,5	13 683 333	2,9	17 019 286	3,1
45	192	12 068 756	6,7	4 952 392	11,5	4 870 684	3,5	7 116 364	4,9
46	1400	105 722 741	4,3	53 504 747	5,6	21 610 782	1,9	52 217 994	4,7
47	518	40 345 794	2,7	13 852 640	4,7	18 251 190	1,7	26 493 154	2,1
49	246	21 014 836	4,7	12 349 278	8,1	8 966 846	2,0	8 665 557	5,7
50	35	12 814 289	0,4	12 327 220	0,4	244 557	3,6	487 069	4,0
51	34	6 427 458	2,2	4 438 199	2,7	1 108 744	2,0	1 989 260	1,7
52	155	12 823 213	4,8	7 905 877	7,5	3 038 358	1,8	4 917 336	3,4
53	49	7 939 832	1,1	3 584 329	1,9	3 831 180	0,5	4 355 503	0,6
55	251	8 904 964	2,6	4 243 654	4,1	3 517 994	2,5	4 661 309	3,2
56	281	14 572 287	3,3	7 235 842	4,1	5 658 449	3,7	7 336 445	3,6
58	87	4 325 390	4,2	2 629 712	5,8	1 453 299	2,9	1 695 678	3,3
59	72	1 156 201	13,5	717 416	16,5	326 962	8,0	438 786	9,7
60	45	2 085 545	1,3	1 107 637	1,8	855 371	0,3	977 909	1,5
61	62	16 693 207	2,5	9 083 718	4,5	2 920 009	0,6	7 609 488	0,6
62	221	17 762 909	5,2	8 506 762	9,6	7 257 140	2,8	9 256 147	3,3
63	51	4 146 205	34,2	2 411 071	33,2	602 030	5,3	1 735 135	35,7
68	435	9 271 821	6,4	3 116 871	8,6	2 131 375	2,7	6 154 950	6,9
69	137	10 403 394	3,5	2 959 960	7,3	5 776 823	4,0	7 443 434	3,3
70	146	20 507 983	8,1	8 845 280	8,9	7 345 361	6,2	11 662 703	9,9
71	244	18 922 141	6,0	9 176 595	10,5	7 519 354	3,4	9 745 546	3,6
72	95	15 487 334	5,9	10 824 609	6,7	2 236 701	2,8	4 662 725	9,0
73	148	6 300 997	11,1	4 295 001	15,7	1 694 293	9,0	2 005 996	7,0
74	102	2 051 217	8,9	963 776	11,5	745 586	5,8	1 087 442	9,7
75	177	517 156	2,2	229 699	2,1	151 330	3,2	287 457	2,7
77	55	2 656 599	10,9	1 657 667	14,0	460 000	3,6	998 932	10,3
78	61	6 103 835	11,5	4 906 958	13,5	1 091 513	6,3	1 196 877	9,8
79	70	6 394 713	7,4	5 411 501	8,4	1 027 916	4,2	983 212	7,4
80	62	1 277 552	4,3	349 130	10,4	832 506	2,3	928 422	2,5
81	200	6 052 869	6,3	2 175 347	10,4	3 235 599	4,2	3 877 523	4,5
82	84	3 332 870	7,1	1 627 664	11,0	1 170 847	3,9	1 705 206	4,8
85	271	5 167 473	10,7	1 920 279	13,8	3 814 899	3,9	3 247 195	9,4
87	531	5 774 570	1,2	1 509 476	2,3	4 168 244	0,8	4 265 093	1,2
88	196	671 539	7,5	231 196	11,2	756 302	3,6	440 343	7,1
90	69	516 356	6,1	373 738	8,9	606 717	3,5	142 618	20,9
91	27	95 874	9,2	62 757	14,4	56 160	4,1	33 117	26,9
92	22	1 781 870	0,4	469 759	0,6	234 124	0,4	1 312 111	0,4
93	136	5 113 211	12,5	3 498 212	16,8	1 140 375	4,8	1 614 999	6,0
94	113	2 243 965	5,3	1 332 105	7,4	1 020 022	4,6	911 860	6,8
95	55	361 986	4,2	167 535	6,9	153 882	3,5	194 451	3,4
96	172	2 119 507	5,2	778 381	8,2	911 240	4,2	1 341 126	4,7
Median			4,6		6,7		2,5		3.6

Fazit: Im Grossen und Ganzen liegen die beobachteten Präzisionen in jenem Bereich, in welchem sie aufgrund des Arbeiten rund um den Stichprobenplan erwartet werden durften. Eine Ausnahme bildet die Variable t022, für welche die Resultate etwas weniger präzis waren als erwartet.

4 Schlussfolgerungen

Im Folgenden werden die wichtigsten Erkenntnisse, die sich aus den in diesem Bericht beschriebenen Arbeiten ergeben haben, kurz zusammengefasst.

Einsetzungen im Rahmen der statistischen Datenaufbereitung

Einsetzungenmethoden konnten im Rahmen der WS2009 bei der Behandlung folgender Situationen eingesetzt werden:

- Berechnung von Positionswerten, die im Rahmen des reduzierten Fragebogens für kleinere Unternehmen nicht erfragt werden. Als Varianten wurden verschiedene Arten von Einsetzungen (Nearest Neighbour, Quotienteneinsetzungen) betrachtet, mit denen die Aggregatswerte aus dem reduzierten Fragebogen auf die Detailpositionen gemäss umfassendem Fragebogen verteilt werden können. Im Rahmen einer Simulationsstudie wurde der durch die Einsetzungen verursachte Schätzfehler bei der Schätzung von Totalen auf Niveau NOGA Stufe 2 untersucht. Dies lieferte folgende Anhaltspunkte:
  - Bei betragsmässig wichtigen Positionen schneiden die betrachteten Einsetzungsvarianten gegenüber der Variante, welche dem Datenzustand nach der elektronischen Datenerfassung entspricht, generell besser ab.
  - Ungeachtet der verwendeten Einsetzungsmethode können die Einsetzungen zu hohen relativen Schätzfehlern bei der Hochrechnung einzelner Positionen führen. Dabei ist dies vorallem bei betragsmässig weniger wichtigen Positionen der Fall, wo aufgrund der kleinen Totale ein relativer Fehler recht schnell gross werden kann. Sind die Hochrechnungsresultate einer solche Einzelposition von Interesse, so empfiehlt es sich, diese im reduzierten Fragebogen separat zu erheben. Zu beachten ist, dass die Hauptvariablen wie die Bruttoproduktion oder die Vorleistungen jeweils betragsmässig wichtigen Positionsaggregaten entsprechen, für welche die gemachten Analysen darauf hinweisen, dass der Effekt der Einsetzungen auf die Totalschätzung im Allgemeinen relativ klein ist.
- Behandlung von Sammelantworten für die Unternehmen in einer Unternehmensgruppe: Sammelantworten werden proportional zur Beschäftigtenzahl auf die einzelnen Unternehmen verteilt. Damit wird gewährleistet, dass die antwortenden Einheiten konsistent sind zu den Einheiten, welche in der Stichprobe gezogen worden sind.
- Unternehmensspezifische Einsetzungen bei sehr grossen Unternehmen, welche trotz prioritärer Behandlung im Rahmen der Datensammlungsphase nicht antworten. Diese Einsetzung wird fallweise von der Sektion MON unter der Ausschöpfung der jeweils zur Verfügung stehenden Hilfsinformation vorgenommen und sollte nur ganz wenige Unternehmen betreffen. Sie dient dazu, den Antwortausfall eines als einzigartig identifizierten Unternehmens im Rahmen der Schätzung möglichst gut zu kompensieren.
- Eine Regressionseinsetzung zur Behandlung von Merkmalausfällen bezüglich der Position Bilanzsumme.

Die statistischen Einsetzungen haben einerseits das Ziel, die Qualität der Schätzungen zu verbessern, andererseits haben sie aber auch andere wünschenswerte Eigenschaften wie die Reproduzierbarkeit und die Tatsache, dass dank entsprechender Indikatoren verschiedene Masse rund um die Einsetzungen berechnet werden können (z.B. gewichtete / ungewichtete Anzahl Einsetzungen etc.).

Identifikation und Behandlung einzigartiger Unternehmen

Aus Erfahrung ist bekannt, dass branchenweise geschätzte Werte der Wertschöpfungsstatistik stark von einzelnen sehr grossen Unternehmen beeinflusst werden können. Falls so ein einzigartiges Unternehmen nicht antwortet, ist es fraglich, inwieweit es durch die Gewichtung der restlichen Unternehmen kompensiert werden kann. Darum wurde im Hinblick auf die WS2009 eine Strategie für die Identifikation und Behandlung solcher Unternehmen eingeführt:

- Verhinderung Schätzfehler mittels
  - Stichprobenplan (Take-All Schichten)
  - gezielter Behandlung im Rahmen der Datenaufbereitung (Selective Editing: 100% Antwortrate)
- Identifikation mittels bedingtem Bias des Horvitz-Thompson Schätzers

Diese Unternehmen werden also vollerhoben und es wird eine Antwortrate von 100% angestrebt. Entsprechende Berechnungen weisen darauf hin, dass dieser Ansatz schon bei einer recht moderaten Anzahl identfizierter Unternehmen zu einer beachtlichen Varianzreduktion der Schätzung führen kann, ohne das dabei ein Bias in Kauf genommen werden muss.

Die Hochrechnung

Die Hochrechnung beruht wie bereits die Vorgängererhebungen auf einer Variante eines robustifizierten Quotientenschätzers. Im Folgenden werden Anpassungen beschrieben, welche in Bezug auf die Hochrechnungsmethodik vorgenommen worden sind:

- Für die Hochrechnung wird mittels der Variable Vollzeitstellenäquivalent des Stichprobenrahmens eine Kalibrierung auf Niveau NOGA2 vorgenommen. In den Vorjahren wurde das Vollzeitstellenäquivalent der Erhebung auf Totale, beruhend auf der jeweils aktuellen Beschäftigtenstatistik und der Betriebszählung, kalibriert. Damit wurde der Tatsache Rechnung getragen, dass die Erhebung seit 2002 in Form eines Panels durchgeführt wurde: Die Kalibrierung hatte auch zum Zweck, die Schätzungen an Populationsveränderungen anzupassen. Mit der Revision 2009 wird die Wertschöpfungsstatistik im Sinne eines rotativen Panels jährlich erneuert. Der zugrunde liegende Stichprobenrahmen reflektiert also bereits die aktuelle Population, was uns erlaubt, direkt mittels dem Rahmen zu kalibrieren. Dies ist eine transparente Lösung, welche die Extrapolationsgewichte im Vergleich zu früher nur recht schwach verändert.
- Im Sinne einer Robustifizierung werden die Gewichte von Unternehmen, welche im dem Quotientenschätzer zu Grunde liegenden Regressionsmodell für gewisse Hauptvariablen grosse Residuen aufweisen, reduziert. Dabei wurde im Hinblick auf die WS2009 diese Gewichtsreduktion so beschränkt, dass die Gewichte nach Robustifizierung nicht kleiner als Eins sein dürfen. Dies wird dadurch gerechtfertigt, dass die verschiedenen Werte nach der Datenaufbereitungsphase als geprüft und richtig angeschaut werden und somit jedes Unternehmen zumindest sich selbst repräsentieren soll (d.h. mindestens ein Gewicht von

Eins hat). Untersuchungen zeigten, dass diese Massnahme das Niveau der Schätzungen tendenziell hebt, was auf eine Verringerung eines allfälligen Bias hinweist. Andererseits muss aber eine Erhöhung der Varianz in Kauf genommen werden, weil die Schätzungen weniger robust sind.

- Im Rahmen der WS2009 werden Schätzungen für verschiedene Untersuchungsbereiche (z.B. Wirtschaftszweige) benötigt. Hierzu werden spezielle Variablen definiert, welche für Unternehmen innerhalb des Untersuchungsbereiches der Zielvariablen entsprechen und ausserhalb des Untersuchungsbereiches den Wert Null annehmen. Mittels diesen Variablen und den aus dem robustifizierten Quotientenschätzer resultierenden Gewichten ergeben sich die für unterschiedliche Untersuchungsbereiche verwendeten Schätzungen aus gewichteten Summen über die ganze Stichprobe.

Nebst den Schätzungen selbst ist auch deren Qualität wichtig. Die im Bericht beschriebenen Methoden liefern auch Varianzschätzungen, mittels denen beispielsweise Variationskoeffizienten konstruiert werden können. Testhochrechnungen der Hauptvariablen Bruttoproduktion, Vorleistungen, Personalkosten und Wertschöpfung auf Ebene NOGA2 zeigten, dass die erzielten Präzisionen, beurteilt mittels Variationskoeffizienten, im Grossen und Ganzen im Bereich liegen, in welchem sie aufgrund des Arbeiten rund um den Stichprobenplan erwartet werden durften. Eine Ausnahme bildet die Variable Vorleistungen, für welche die erzielten Präzisionen etwas unter den Erwartungen lagen.

Anhang

A Wirtschaftszweige gemäss zweiter Stufe NOGA 2008

Tabelle 11 NOGA 2008 - Stufe 2.

Codes	Titel
01	Landwirtschaft, Jagd und damit verbundene Tätigkeiten
02	Forstwirtschaft und Holzeinschlag
03	Fischerei und Aquakultur
05	Kohlenbergbau
06	Gewinnung von Erdöl und Erdgas
07	Erzbergbau
08	Gewinnung von Steinen und Erden, sonstiger Bergbau
09	Erbringung von Dienstleistungen für den Bergbau und für die Gewinnung von Steinen und Erden
10	Herstellung von Nahrungs- und Futtermitteln
11	Getränkeherstellung
12	Tabakverarbeitung
13	Herstellung von Textilien
14	Herstellung von Bekleidung
15	Herstellung von Leder, Lederwaren und Schuhen
16	Herstellung von Holz-, Flecht-, Korb- und Korkwaren (ohne Möbel)
17	Herstellung von Papier, Pappe und Waren daraus
18	Herstellung von Druckerzeugnissen; Vervielfältigung von bespielten Ton-, Bild- und Datenträgern
19	Kokerei und Mineralölverarbeitung
20	Herstellung von chemischen Erzeugnissen
21	Herstellung von pharmazeutischen Erzeugnissen
22	Herstellung von Gummi- und Kunststoffwaren
23	Herstellung von Glas und Glaswaren, Keramik, Verarbeitung von Steinen und Erden
24	Metallerzeugung und -bearbeitung
25	Herstellung von Metallerzeugnissen
26	Herstellung von Datenverarbeitungsgeräten, elektronischen und optischen Erzeugnissen
27	Herstellung von elektrischen Ausrüstungen
28	Maschinenbau
29	Herstellung von Automobilen und Automobilteilen
30	Sonstiger Fahrzeugbau
31	Herstellung von Möbeln
32	Herstellung von sonstigen Waren
33	Reparatur und Installation von Maschinen und Ausrüstungen
35	Energieversorgung
36	Wasserversorgung
37	Abwasserentsorgung
38	Sammlung, Behandlung und Beseitigung von Abfällen; Rückgewinnung
39	Beseitigung von Umweltverschmutzungen und sonstige Entsorgung
41	Hochbau
42	Tiefbau
43	Vorbereitende Baustellenarbeiten, Bauinstallation und sonstiges Ausbaugewerbe
45	Handel mit Motorfahrzeugen; Instandhaltung und Reparatur von Motorfahrzeugen
46	Grosshandel (ohne Handel mit Motorfahrzeugen)
47	Detailhandel (ohne Handel mit Motorfahrzeugen)
49	Landverkehr und Transport in Rohrfernleitungen
50	Schifffahrt
51	Luftfahrt
52	Lagerei sowie Erbringung von sonstigen Dienstleistungen für den Verkehr
53	Post-, Kurier- und Expressdienste
55	Beherbergung
56	Gastronomie
58	Verlagswesen
59	Herstellung, Verleih und Vertrieb von Filmen und Fernsehprogrammen; Kinos; Tonstudios und Verlegen von Musik
60	Rundfunkveranstalter
61	Telekommunikation
62	Erbringung von Dienstleistungen der Informationstechnologie
63	Informationsdienstleistungen
64	Erbringung von Finanzdienstleistungen
65	Versicherungen, Rückversicherungen und Pensionskassen (ohne Sozialversicherung)
66	Mit Finanz- und Versicherungsdienstleistungen verbundene Tätigkeiten
68	Grundstücks- und Wohnungswesen
69	Rechts- und Steuerberatung, Wirtschaftsprüfung
70	Verwaltung und Führung von Unternehmen und Betrieben; Unternehmensberatung
71	Architektur- und Ingenieurbüros; technische, physikalische und chemische Untersuchung
72	Forschung und Entwicklung
73	Werbung und Marktforschung
74	Sonstige freiberufliche, wissenschaftliche und technische Tätigkeiten
75	Veterinärwesen
77	Vermietung von beweglichen Sachen
78	Vermittlung und Überlassung von Arbeitskräften
79	Reisebüros, Reiseveranstalter und Erbringung sonstiger Reservierungsdienstleistungen
80	Wach- und Sicherheitsdienste sowie Detekteien
81	Gebäudebetreuung; Garten- und Landschaftsbau
82	Erbringung von wirtschaftlichen Dienstleistungen für Unternehmen und Privatpersonen a. n. g.
84	Öffentliche Verwaltung, Verteidigung; Sozialversicherung
85	Erziehung und Unterricht
86	Gesundheitswesen
87	Heime (ohne Erholungs- und Ferienheime)
88	Sozialwesen (ohne Heime)
90	Kreative, künstlerische und unterhaltende Tätigkeiten
91	Bibliotheken, Archive, Museen, botanische und zoologische Gärten
92	Spiel-, Wett- und Lotteriewesen
93	Erbringung von Dienstleistungen des Sports, der Unterhaltung und der Erholung
94	Interessenvertretungen sowie kirchliche und sonstige religiöse Vereinigungen (ohne Sozialwesen und Sport)

Weiter auf der nächsten Seite

Codes	Titel
95	Reparatur von Datenverarbeitungsgeräten und Gebrauchsgütern
96	Erbringung von sonstigen überwiegend persönlichen Dienstleistungen
97	Private Haushalte mit Hauspersonal
98	Herstellung von Waren und Erbringung von Dienstleistungen durch private Haushalte für den Eigenbedarf ohne ausgeprägten Schwerpunkt
99	Exterritoriale Organisationen und Körperschaften

Literatur

[1] Särndal, C-E., Svensson, B., Wretman, W. (1992). *Model Assisted Survey Sampling*, Springer Series in Statistics, Springer-Verlag

[2] Hulliger, B. (1999). *Simple and Robust Estimators for Sampling*, 1999 Proceedings of the section on Survey Research Methods, 54-63.

[3] Beaumont, J.-F., Alavi, A. (2004). *Estimation robuste par la regression généralisée*, Techniques d'enquête, 30:217-231.

[4] Särndal, C-E. and Lundström, S. (2005). *Estimation in surveys with nonresponse.* John Wiley & Sons.

[5] Moreno-Rebollo, J.L., Munoz-Reyes, A., Munoz-Pichardo, J. (1999). *Influence diagnostics in survey sampling: Conditional bias.* Biometrika, 86, 923-928.

[6] Luzi, O. et al. (2007). *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys.*
http://edimbus.istat.it/EDIMBUS1/document/RPM_EDIMBUS/RPM_EDIMBUS.pdf

[7] Peters, R., Renfer, J.-P. et Hulliger, B. (2001). *Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données.* Bundesamt für Statistik, Bestellnummer: 338-0005

[8] Miller, D.M. (1984). *Reducing Transformation Bias in Curve Fitting.* The American Statistician, 38(2), 124-126.

[9] Assoulin, D. (2013). *Wertschöpfungsstatistik. Revision 2009: Stichprobenrahmen und Stichprobenplan.* Bundesamt für Statistik, Bestellnummer: 338-0070

Methodenberichte der Sektion Statistische Methoden des BFS
Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodology reports published by the FSO's Statistical Methods Section

Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009: Statistische Datenaufbereitung und Hochrechnung. Bestellnummer: 338-0072

Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro de commande: 338-0071

Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009: Stichprobenrahmen und Stichprobenplan. Bestellnummer: 338-0070

Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007: Stichprobenplan und Hochrechnung. Bestellnummer: 338-0069

Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010: Stichprobenplan und Hochrechnung. Bestellnummer: 338-0068

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. Order number: 338-0067-05

Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0066

Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. Bestellnummer: 338-0065

Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. Bestellnummer: 338-0064

Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0063

Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. Bestellnummer: 338-0062

Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro de commande: 338-0061

Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. Order number: 338-0060

Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro de commande: 338-0059

Graf, E. (2009). Weightings of the Swiss Household Panel: SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. Order number: 338-0058

Graf, E. (2009). Pondérations du Panel Suisse de Ménages: PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro de commande: 338-0057-05

Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro de commande: 338-0056

Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro de commande: 338-0055

Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro de commande: 338-0054

Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires SSalarium". Numéro de commande: 338-0053

Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonage. Numéro de commande: 338-0052

Graf, E. (2008). Pondérations du SILC pilote. SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro de commande: 338-0051

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. Bestellnummer: 338-0050

Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro de commande: 338-0049

Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. Bestellnummer: 338-0048

Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. Bestellnummer: 338-0047

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. Bestellnummer: 338-0046

Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro de commande: 338-0045

Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro de commande: 338-0044

Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0043

Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. Bestellnummer: 338-0042

Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0041

Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro de commande: 338-0040

Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro de commande: 338-0039

Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037

Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036

Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035

Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020.

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande: 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des é lèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d' échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln:

Diffusionsmittel	*Kontakt*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch
Das BFS im Internet	www.statistik.admin.ch
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch
Publikationen zur vertieften Information	032 713 60 60 order@bfs.admin.ch
Online-Datenrecherche (Datenbanken)	www.stattab.bfs.admin.ch

Nähere Angaben zu den verschiedenen Diffusionsmitteln im Internet unter der Adresse www.statistik.admin.ch→Dienstleistungen→Publikationen Statistik Schweiz

Methodenberichte der Sektion Statistische Methoden

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zugrunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicherzustellen, und andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

Die Wertschöpfungsstatistik (WS) basiert auf einer Stichprobenerhebung in welcher eine Vielzahl betriebswirtschaftlicher Variablen erfasst werden. Nach der Datensammlung durchlaufen die Daten eine Aufbereitungsphase, in welcher sie kontrolliert und nötigenfalls bearbeitet werden, so dass sie dann in der Hochrechnung möglichst unverzerrte Resultate liefern können. Im Zusammenhang mit der Revision 2009 wurden verschiedene statistische Einsetzungsmethoden eingeführt, welche im Rahmen der Datenaufbereitung eingesetzt werden, um verschiedene Situationen wie den Einsatz eines reduzierten Fragebogens bei kleinen Unternehmen, Sammelantworten, fehlende Angaben zu gewissen Positionen oder auch Antwortausfälle bei sehr grossen Unternehmen zu behandeln. Nebst statistischen Einsetzungen wurden auch Methoden für die Identifikation und Behandlung von Unternehmen eingeführt, welche aufgrund ihrer Grösse in ihrer Branche als einzigartig betrachtet werden. Im Rahmen der Schätzung von Totalen werden diese Unternehmen mit Eins gewichtet. Für die übrigen Unternehmen besteht die Berechnung der Hochrechnungsgewichte ausgehend von den Stichprobengewichten aus drei Schritten: Korrektur für Antwortausfälle, Robustifizierung und Kalibrierung (Kalibrierungsvariable: Vollzeitstellenäquivalent des Stichprobenrahmens). Nebst Schätzungen für Totale werden im Bericht auch Methoden dargestellt, mit denen entsprechende Varianzen und Variationskoeffizienten geschätzt werden können. Testhochrechnungen zeigten, dass die erzielten Präzisionen im Grossen und Ganzen in jenem Bereich liegen, in welchem sie aufgrund der Arbeiten rund um den Stichprobenplan erwartet werden durften.

Bestellnummer
338-0072

Bestellungen
Tel.: 032 713 60 60
Fax: 032 713 60 61
E-Mail: order@bfs.admin.ch

Preis
Gratis

ISBN 978-3-303-00516-3