Methodenbericht

Wertschöpfungsstatistik

Einsetzungsversuche für fehlende Antworten grosser Unternehmen



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Neuchâtel, 2007

Statistik der Schweiz

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 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, 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

Statistique de la Suisse

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

0 Bases statistiques et produits généraux
1 Population
2 Espace et environnement
3 Vie active et rémunération du travail
4 Economie nationale
5 Prix
6 Industrie et services
7 Agriculture et sylviculture
8 Energie
9 Construction et logement
10 Tourisme
11 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, société de l'information, sport
17 Politique
18 Administration et finances publiques
19 Criminalité et droit pénal
20 Situation économique et sociale de la population
21 Développement durable et disparités régionales et internationales

Statistik der Schweiz | **Statistique de la Suisse**

Methodenbericht | **Rapport de méthodes**

Wertschöpfungsstatistik

Einsetzungsversuche für fehlende Antworten grosser Unternehmen

Autor **Daniel Assoulin**
Bundesamt für Statistik

Herausgeber **Bundesamt für Statistik**

Neuchâtel, 2007

Vorwort

Die Wertschöpfungsstatistik basiert auf einer jährlich durchgeführten Stichprobenerhebung bestehend aus über die ganze Schweiz verteilten Unternehmen. Einzelnen sehr grossen Unternehmen kann bei der Schätzung von Branchentotalen eine grosse Bedeutung zukommen. In solchen Fällen stellt sich die Frage, wie mit fehlenden Werten umgegangen werden soll. In dieser Arbeit werden anhand der Wertschöpfungsstatistiken 2002–2004 verschiedene Einsetzungsvarianten einerseits untereinander und andererseits mit der momentan verwendeten Gewichtungslösung verglichen.

Die Arbeiten wurden durch Daniel Assoulin (METH) ausgeführt. Der Autor dankt Frau Veronica Roth und Herrn Markus Däppen (Sektion MON) sowie Herrn Paul-André Salamin (METH) für die fachliche Unterstützung und die gute Zusammenarbeit. Ein weiterer Dank geht an Herrn Jean-Pierre Renfer (METH) für die aufmerksame Durchsicht des Berichts.

Zusammenfassung

Für die Jahre 2002–2004 wurden fehlende Werte grosser Unternehmen mit vier verschiedenen Einsetzungsmethoden behandelt und die daraus resultierenden Auswirkungen auf die Extrapolationstotale untersucht. Tendenziell erhöhten die Einsetzungen die Schätzungen, wobei die Unterschiede im Allgemeinen moderat ausfielen. In einem Vergleich der Einsetzungsmethoden anhand einer Simulationsstudie schnitt die Einsetzung mittels trendangepassten Vorjahreswerten am besten ab.

Schlüsselwörter

Methodenbericht; Einsetzungen; EUREDIT; fehlende Werte; Quotienteneinsetzung; Quotientenschätzer; Wertschöpfungsstatistik.

Auskunft: Daniel Assoulin, Tel. 032 713 67 51
Daniel.Assoulin@bfs.admin.ch

Bearbeitung: Dienst Statistische Methoden, BFS

Vertrieb: Bundesamt für Statistik
CH-2010 Neuchâtel
Tel. 032 713 60 60 / Fax 032 713 60 61
Order@bfs.admin.ch

Internet: http://www.statistik.admin.ch

Bestellnummer: 338-0048

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Graphik/Layout: BFS

Copyright: BFS, Neuchâtel 2007
Abdruck – ausser für kommerzielle Nutzung –
unter Angabe der Quelle gestattet.

ISBN: 978-3-303-00391-6

Inhaltsverzeichnis

1 Einführung **5**

2 Der aktuelle Ansatz für fehlende Werte bei sehr grossen Unternehmen **5**

3 Die Definition des sehr grossen Unternehmens **7**
3.1 Die NOGA-Zugehörigkeit 7
3.2 Die Behandlung nicht mehr existierender Unternehmen 7

4 Die Verteilung fehlender Werte **8**

5 Die betrachteten Einsetzungsvarianten **8**
5.1 Einsetzung der Untersuchungsvariablen 11
5.2 Einsetzung der Hilfsvariablen 13

6 Der Einfluss der Einsetzungen auf die geschätzten Totale **13**
6.1 Die Veränderungen der geschätzten Totale gegenüber der aktuellen Methode . . 13
6.2 Der Einfluss von Einsetzungen auf die relativen Veränderungen zwischen zwei Jahren 18

7 Qualitätsvergleich der Einsetzungen **18**
7.1 Rahmenbedingungen der Simulation 19
7.2 Die Distanz zwischen echten und eingesetzten Werten 20
7.3 Der Vergleich der ersten Momente 22
7.4 Der lineare Zusammenhang zwischen eingesetzten und tatsächlichen Werten . . 23
7.4.1 Der Steigungsparameter 23
7.4.2 Der mittlere quadratische Fehler 24
7.5 Interpretation der Resultate 25

8 Schlussfolgerungen **27**
8.1 Weitere Arbeitsschritte 28

Literatur **30**

1 Einführung

Im Rahmen der Wertschöpfungsstatistik (WS) [2] werden aufgrund einer Stichprobe Buchhaltungszahlen von Schweizer Unternehmen erhoben. Anhand der Stichprobenwerte werden Populationstotale geschätzt. Um dem teilweise starken Einfluss sehr grosser Unternehmen auf die Populationstotale Rechnung zu tragen, werden diese vollständig erhoben. In der Realität gibt es aber sehr grosse Unternehmen, welche ihre Zahlen nicht einreichen. In der aktuellen Schätzmethode werden fehlende Angaben solcher Unternehmen mittels Extrapolation durch einen Quotientenschätzer behandelt. Als Alternative zu diesem Verfahren könnten fehlende Werte mittels einer Einsetzung (Imputation) ersetzt werden. In diesem Bericht wird untersucht, ob und unter welchen Bedingungen eine Einsetzung fehlender Werte bei sehr grossen Unternehmen im Vergleich zur aktuellen Methode die Qualtät der Resultate verbessern könnte. Zu diesem Zweck wurden die Extrapolationswerte aus der aktuellen Methode für drei Zielvariablen (Wert der Bruttoproduktion: $t017$, Vorleistungen: $t022$ und Personalkosten: $t053$) über die drei Jahre 2002-4 mit den aus vier Einsetzungsvarianten resultierenden Extrapolationswerten verglichen. Anhand der Extrapolationsresultate und einer Simulationsstudie wurde zudem untersucht, welche der betrachteten Einsetzungsmethoden die besten Ergebnisse liefern würde.

2 Der aktuelle Ansatz für fehlende Werte bei sehr grossen Unternehmen

Um zu verstehen, was im aktuellen Ansatz mit fehlenden Werten sehr grosser Unternehmen geschieht, muss man das für die Schätzungen verwendete Extrapolationsverfahren betrachten. Im Folgenden wird zu diesem Zwecke eine grobe Skizze des Verfahrens, welches in [1] detailliert beschrieben wird, gegeben. Die Schätzung der Populationstotale erfolgt für die Wertschöpfungsstatistik auf Ebene der NOGA2-Wirtschaftsklassen (allgemeine Systematik der Wirtschaftszweige). Die Menge c bezeichnet sämtliche in der NOGA2-Klasse c enthaltenen Unternehmen. Die verwendete Schätzmethodik unterscheidet grundsätzlich zwischen sehr grossen und den restlichen Unternehmen. Darum wird die Menge der nicht als sehr gross eingestuften Unternehmen der NOGA2-Klasse c mit $c*$ bezeichnet. Als Grössenkriterium wird ein Vollzeitstellenäquivalent (kurz: s_{14}) verwendet. Die Extrapolationsmethode für die nicht sehr grossen Unternehmen beruht auf einem Quotientenschätzer, der als Hilfsvariable das Vollzeitstellenäquivalent s_{14} verwendet. Für die Population $c*$ kann das geschätzte Populationstotal für eine Variable Y, $\hat{Y}_{Tot,c*}$, als

$$\hat{Y}_{Tot,c*} = \hat{\beta}_{c*} X_{Tot,c*} \tag{1}$$

dargestellt werden. Dabei ist $X_{Tot,c*}$ das Populationstotal des Vollzeitstellenäquivalentes zum Zeitpunkt der Erhebung und $\hat{\beta}_{c*}$ der geschätzte Koeffizient gemäss Herleitung in [1]. Die Summe $\sum_{c\setminus c*} y_i$ ergibt das Total der Werte aller sehr grossen Unternehmen des Wirtschaftssektors c. Addiert man diese Summe zur Schätzung in (1), erhält man den verwendeten Schätzer für das Total der Variable Y im Wirtschaftssektor c:

$$\hat{Y}_{Tot,c} = \hat{\beta}_{c*} X_{Tot,c*} + \sum_{c\setminus c*} y_i \tag{2}$$

Würden die sehr grossen Unternehmen immer antworten, könnten die Terme $X_{Tot,c*}$ und $\sum_{c\setminus c*} y_i$ berechnet und in (2) eingesetzt werden. In der Realität existieren aber sehr grosse Unternehmen, deren Zahlen nicht zur Verfügung stehen. Somit lassen sich auch die beiden Terme nicht genau berechnen und es macht Sinn, die Menge der antwortenden Unternehmen

Tabelle 1 Beispielwerte für (4) in 1000 Franken.

$\hat{\beta}_{c*}X_{Tot,c*}$	$\hat{\beta}_{c*}X_{Tot,L_c\cap M}$	$\sum_{R_{L_c}} y_i$	$\hat{Y}^{(W)}_{Tot,c}$
9500	500	1500 $(= y_1)$	11500

mit R und jene der nicht antwortenden mit M zu bezeichnen. Die sehr grossen Unternehmen in der NOGA2-Klasse c bezeichnen wir mit $L_c = c \setminus c*$ und mit $R_{L_c} = (L_c \cap R)$ die Menge der sehr grossen Unternehmen, die auch geantwortet haben. Hat nur ein Teil der sehr grossen Unternehmen geantwortet, so ist $R_{L_c} \subset L_c$ und (2) wird zu

$$\hat{Y}^{(W)}_{Tot,c} = \hat{\beta}_{c*}X_{Tot,c\setminus R_{L_c}} + \sum_{R_{L_c}} y_i, \tag{3}$$

wobei $X_{Tot,c\setminus R_{L_c}}$ das Populationstotal der Unternehmen aus $c*$ zuzüglich der Vollzeitstellenäquivalente der nicht antwortenden grossen Unternehmen aus der Klasse c bezeichnet. Mit dem Term $X_{Tot,L_c\cap M} = X_{Tot,c\setminus R_{L_c}} - X_{Tot,c*}$ (das Vollzeitstellenäquivalent der nicht antwortenden sehr grossen Unternehmen) lässt sich der Schätzer auch als

$$\hat{Y}^{(W)}_{Tot,c} = \underbrace{\hat{\beta}_{c*}X_{Tot,c*}}_{T_1} + \underbrace{\hat{\beta}_{c*}X_{Tot,L_c\cap M} + \sum_{R_{L_c}} y_i}_{T_2} \tag{4}$$

schreiben. T_1 kann als Beitrag der Unternehmen in $c*$ und T_2 als Beitrag der sehr grossen Unternehmen in der NOGA2-Klasse c betrachtet werden, wobei deutlich wird, dass das Total der nicht antwortenden Unternehmen (egal ob sehr gross oder nicht) mit dem Quotientenschätzer geschätzt wird. Die antwortenden sehr grossen Unternehmen erhalten nach wie vor das Gewicht 1. Im Folgenden wird (4) anhand eines fiktiven Beispiels illustriert. Situation:

- Es wird davon ausgegangen, dass sich in der NOGA2-Kategorie c zwei grosse Unternehmen U1 und U2 befinden.
- Unternehmen U1 hat bei der Erhebung wie folgt geantwortet: Vollzeitstellenäquivalent 1000 und $y_1 = 1.5$ Mio.
- U2 hat nicht geantwortet und es wird angenommen, dass sein Vollzeitstellenäquivalent 500 beträgt.
- $\hat{\beta}_{c*} = 1000$ und $X_{Tot,c*} = 9500$.

Die Werte der in (4) auftretenden Terme sind in Tabelle 1 aufgelistet. Dem nicht antwortenden Unternehmen U2 wird mit einem Wert von 0.5 Mio. Rechnung getragen.

Bemerkung zu Gleichung (4): Ersetzt (imputiert) man die fehlenden Werte bei den sehr grossen Unternehmen mit $\frac{1}{|L_c\cap M|}\hat{\beta}_{c*}X_{Tot,L_c\cap M}$, wobei $|L_c \cap M|$ die Anzahl nicht antwortender sehr grosser Unternehmen in c bezeichnet und berechnet mit diesen eingesetzten Werten (2), führt dies auf das gleiche Resultat wie (4). Das Resultat der aktuellen Methode könnte somit theoretisch, falls der Term $X_{Tot,L_c\cap M}$ bekannt wäre, auch mit einer entsprechend gewählten Einsetzung erziehlt werden. Da in der aktuellen Methode aber nie eine explizite Ersetzung fehlender Werte stattfindet, wird diese im weiteren Verlauf der Arbeit nicht als Einsetzungsvariante bezeichnet.

3 Die Definition des sehr grossen Unternehmens

Es gibt verschiedene Möglichkeiten, um anhand des Vollzeitstellenäquivalentes sehr grosse Unternehmen zu definieren. Für die Extrapolation in der WS04 wurde ein Unternehmen unabhängig von der NOGA2-Klasse als sehr gross definiert, falls es in der Erhebung ein Vollzeitstellenäquivalent von mindestens 300 angegeben hat. Falls $s_{14}(i, 04)$ das Vollzeitstellenäquivalent des Unternehmens i im Jahre 2004 bezeichnete, würde dieses also in der WS04 als sehr gross behandelt, falls $s_{14}(i, 04) \geq 300$ erfüllt wäre. Da Werte für jene Unternehmen eingesetzt werden sollen, die in der Berechnung als sehr gross behandelt werden, wurde für die Einsetzung eine ähnliche Definition verwendet. Damit aber über die drei untersuchten Jahre hinweg mit derselben Menge von Unternehmen gearbeitet werden konnte, wurde ein Unternehmen als sehr gross definiert, falls sein Vollzeitstellenäquivalent entweder zum Zeitpunkt der Ziehung oder in einer der Erhebungen 02-04 mindestens 300 betrug [1]. Insgesamt sind nach dieser Definition 736 Unternehmen aus dem Stichprobenrahmen sehr gross. Ob Unternehmen i sehr gross ist oder nicht, entscheidet sich somit anhand der Variablen $s_{14}(i, t),\ t = Z, 2, 3, 4$, wobei $s_{14}(i, Z)$ das Vollzeitstellenäquivalent zum Zeitpunkt der Ziehung bezeichnet. Bei einem Unternehmen, dass während sämtlichen Jahren geantwortet hat, ist dies kein Problem, da die Vollzeitstellenäquivalente sämtlicher Jahre vorhanden sind. Hat ein Unternehmen in einem gewissen Jahr nicht geantwortet, wurde das Vollzeitstellenäquivalent mit dem letzten beobachteten Wert („last carry forward") eingesetzt. Diese Einsetzungsart für die Hilfsvariable Vollzeitstellenäquivalent hat den praktischen Vorteil, dass die Menge der sehr grossen Unternehmen nicht durch Einsetzungswerte verändert wird.

3.1 Die NOGA-Zugehörigkeit

Im Rahmen der Erhebung geben die Unternehmen auch ihre aktuellen NOGA2-Klassen an. Die entsprechende Variable wird für die Erhebung des Jahres $0t$ mit $noga2_{ws0t}$ bezeichnet. Ihr Wert ist in der Regel mit der NOGA2-Klassifizierung zum Zeitpunkt der Ziehung identisch, dies muss aber nicht zwingend der Fall sein. Bei nicht antwortenden Unternehmen wurde die Variable $noga2_{ws0t}$ analog zum Vollzeitstellenäquivalent mittels dem letzten beobachteten Wert eingesetzt. Die für die Einsetzungsberechnungen verwendeten Schichten beruhen auf der $noga2$-Einteilung des Jahres 2004, dass heisst auf $noga2_{ws04}$.

3.2 Die Behandlung nicht mehr existierender Unternehmen

Gemäss einer durch die Sektion Monetäre Unternehmensstatistik erstellten Liste figurieren unter den im Stichprobenrahmen enthaltenen sehr grossen Unternehmen 48, die beispielsweise in Folge von Liquidation oder Fusionierung nicht mehr existieren. Der aktuelle Extrapolationsansatz behandelt diese Unternehmen methodisch gleich wie fehlende Unternehmen. Werden die fehlenden Werte bei den sehr grossen Unternehmen durch Einsetzungen ersetzt, sind für nicht mehr existierende Unternehmen keine Einsetzungen vorzunehmen. Deshalb müssten die nicht mehr existierenden sehr grossen Unternehmen jeweils vor der Schätzung identifziert werden. In dieser Studie wurden für die Berechnung der Einsetzungswerte sämtliche 2004 nicht mehr existierenden sehr grossen Unternehmen ausgeschlossen. Zudem wurden für solche Unternehmen auch keine Werte eingesetzt. In den gemachten Analysen werden darum 688 (statt 736) sehr grosse Unternehmen betrachtet. Bei den restlichen Unternehmen wurden fehlende Werte mittels der aktuellen Methode (Quotientenschätzer) behandelt, wobei davon ausgegangen wird, dass sich Auflösungen und Neugründungen von Unternehmen kompensieren.

[1] Im Jahr 2002 gab es unter den antwortenden 575, im Jahr 2003 567 und im Jahr 2004 551 Unternehmen mit einem Vollzeitstellenäquivalent von über 300.

4 Die Verteilung fehlender Werte

Wie aus Tabelle 2 entnommen werden kann, beträgt der Anteil fehlender Werte [2] im Rahmen der WS04 bei den sehr grossen Unternehmen 8.3 %. In den anderen Grössenkategorien liegt der Anteil der nicht antwortenden Unternehmen wesentlich höher. Die Anteile fehlender Werte in Tabelle 2 sind je nach NOGA2-Klasse unterschiedlich. Andererseits gilt zu beachten, dass die Antwortquoten zwischen verschiedenen Jahren auch innerhalb einer NOGA2-Klasse stark variieren können. Dies hängt natürlich damit zusammen, dass die Anzahl sehr grosser Unternehmen in den verschiedenen NOGA2-Klassen nicht sehr gross ist. Da die NOGA2-Klasse und die Unternehmensgrösse auch einen Einfluss auf die Werte der Zielvariablen haben, wurde für die Einsetzungsmethoden die Schichtung (NOGA2, Grösse=sehr gross) verwendet.

Innerhalb jeder Schicht (NOGA2, Gösse=sehr gross) wurde die Annahme getroffen, dass die Verteilung der fehlenden Werte unabhängig von den Zielvariablen sei (Missing at Random (MAR)).

Um die Verteilung der fehlenden Werte darzustellen wurden für $t = 2, 3, 4$ die Variablen

$$miss_{ws0t}(i) = \begin{cases} m & \text{falls Unternehmen } i \text{ in ws0t nicht antwortete,} \\ r & \text{sonst.} \end{cases}$$

verwendet. Das Buchstabentripel, welches entsteht, wenn man die drei Variablen miteinander verbindet, wird als Variable Struktur bezeichnet. Falls beispielsweise Unternehmen i in den WS02 und 03 antwortete, in der WS04 jedoch nicht, so nimmt die Variable Struktur den Wert Struktur$(i) = rrm$ an. Die Ausprägungen der Variable Struktur werden in Tabelle 3 aufgelistet. In Tabellen 2 und 3 wurden die betrachteten NOGA2-Klassen mit k1,...,k43 codiert. Diese Codierung wird auch im weiteren Verlauf der Arbeit beibehalten.
In Bezug auf eine Einsetzung könnten NOGA2-Klassen mit einem grossen Anteil fehlender Antworten heikel sein. In Tabelle 2 entdeckt man, dass für die WS04 in den Klassen k21 und k40 die Anteile fehlender Werte über 25% liegen. Den Resultaten in diesen Klassen ist bei der Auswertung besonderes Augenmerk zu schenken.

5 Die betrachteten Einsetzungsvarianten

Die Auswahl der betrachteten Einsetzungsmethoden beschränkt sich vorderhand auf deterministische Methoden, was bedeutet, dass keine Zufallsmodelle verwendet wurden, um die Einsetzungen zu erstellen. Gründe für diesen Entscheid sind:

- Deterministische Einsetzungsvarianten liefern einfach reproduzierbare Resultate, was deren Diskussion und Analyse vereinfacht.
- Untersuchungen zur Varianz wurden vorerst ausgeklammert. Um die aus den Einsetzungen herrührende Unsicherheit zu berücksichtigen, müsste den Einsetzungen ein Zufallsmodell zu Grunde gelegt werden.

[2] Grundsätzlich ist es so, dass für ein Unternehmen in einem bestimmten Jahr entweder sämtliche oder gar keine Variablen vorhanden sind. Das heisst innerhalb der Wertschöpfungsstatistik eines Jahres gibt es keine sogenannten Merkmalsausfälle (item nonresponse).

Tabelle 2 Die Häufigkeit fehlender Werte nach NOGA2. Dabei bezeichnet m (missing) die Anzahl Unternehmen, die nicht geantwortet haben und Total die Anzahl sehr grosser Unternehmen in der betrachteten NOGA2-Klasse.

	2002			2003			2004		
NOGA2	m	Total	fehlend in %	m	Total	fehlend %	m	Total	fehlend in %
k1	0	1	0.0	0	1	0.0	0	1	0.0
k2	1	31	3.2	2	31	6.5	3	31	9.7
k3	0	2	0.0	0	2	0.0	0	2	0.0
k4	0	1	0.0	0	1	0.0	0	1	0.0
k5	0	2	0.0	0	2	0.0	0	2	0.0
k6	0	1	0.0	0	1	0.0	0	1	0.0
k7	0	3	0.0	0	3	0.0	0	3	0.0
k8	1	13	7.7	2	13	15.4	2	13	15.4
k9	1	19	5.3	2	19	10.5	1	19	5.3
k10	0	1	0.0	0	1	0.0	0	1	0.0
k11	4	39	10.3	1	39	2.6	4	39	10.3
k12	0	11	0.0	0	11	0.0	0	11	0.0
k13	2	8	25.0	2	8	25.0	0	7	0.0
k14	2	13	15.4	0	13	0.0	2	13	15.4
k15	1	15	6.7	1	15	6.7	1	15	6.7
k16	2	67	3.0	3	66	4.5	1	66	1.5
k17	0	1	0.0	0	1	0.0	0	1	0.0
k18	2	25	8.0	2	26	7.7	3	26	11.5
k19	3	12	25.0	2	12	16.7	2	12	16.7
k20	7	44	15.9	6	44	13.6	5	45	11.1
k21	1	3	33.3	0	3	0.0	1	3	33.3
k22	0	6	0.0	0	6	0.0	0	6	0.0
k23	1	7	14.3	0	7	0.0	1	7	14.3
k24	0	1	0.0	0	1	0.0	0	1	0.0
k25	2	15	13.3	3	15	20.0	1	15	6.7
k26	4	45	8.9	2	45	4.4	6	45	13.3
k27	2	12	16.7	3	12	25.0	2	12	16.7
k28	5	49	10.2	6	50	12.0	1	50	2.0
k29	4	62	6.5	2	62	3.2	2	62	3.2
k30	4	14	28.6	3	14	21.4	1	14	7.1
k31	0	8	0.0	0	8	0.0	0	8	0.0
k32	2	17	11.8	1	17	5.9	1	17	5.9
k33	0	2	0.0	0	2	0.0	0	2	0.0
k34	2	31	6.5	4	31	12.9	7	31	22.6
k35	2	18	11.1	2	18	11.1	3	18	16.7
k36	0	6	0.0	0	6	0.0	0	6	0.0
k37	1	18	5.6	2	18	11.1	0	18	0.0
k38	1	8	12.5	0	8	0.0	0	8	0.0
k39	5	40	12.5	4	39	10.3	4	39	10.3
k40	4	6	66.7	3	6	50.0	3	6	50.0
k41	0	5	0.0	0	5	0.0	0	5	0.0
k42	0	1	0.0	0	1	0.0	0	1	0.0
k43	0	5	0.0	0	5	0.0	0	5	0.0
Total	66	688	9.6	58	688	8.4	57	688	8.3

Tabelle 3 Die Struktur der fehlenden Werte bei den sehr grossen Unternehmen (Ausprägungen der Variablen Struktur).

NOGA2	mmm	mmr	mrm	mrr	rmm	rmr	rrm	rrr	Total
k1	0	0	0	0	0	0	0	1	1
k2	0	0	0	1	2	0	1	27	31
k3	0	0	0	0	0	0	0	2	2
k4	0	0	0	0	0	0	0	1	1
k5	0	0	0	0	0	0	0	2	2
k6	0	0	0	0	0	0	0	1	1
k7	0	0	0	0	0	0	0	3	3
k8	1	0	0	0	1	0	0	11	13
k9	1	0	0	0	0	1	0	17	19
k10	0	0	0	0	0	0	0	1	1
k11	0	0	2	2	1	0	1	33	39
k12	0	0	0	0	0	0	0	11	11
k13	0	1	0	0	0	0	0	6	7
k14	0	0	0	2	0	0	2	9	13
k15	1	0	0	0	0	0	0	14	15
k16	0	0	0	2	0	3	1	60	66
k17	0	0	0	0	0	0	0	1	1
k18	2	0	0	0	0	0	1	23	26
k19	1	1	0	1	0	0	1	8	12
k20	3	2	0	3	1	1	1	34	45
k21	0	0	1	0	0	0	0	2	3
k22	0	0	0	0	0	0	0	6	6
k23	0	0	0	1	0	0	1	5	7
k24	0	0	0	0	0	0	0	1	1
k25	1	1	0	0	0	1	0	12	15
k26	1	1	0	2	0	0	5	36	45
k27	2	0	0	0	0	1	0	9	12
k28	1	1	0	4	0	4	0	40	50
k29	0	1	0	3	0	1	2	55	62
k30	1	1	0	2	0	1	0	9	14
k31	0	0	0	0	0	0	0	8	8
k32	1	0	0	1	0	0	0	15	17
k33	0	0	0	0	0	0	0	2	2
k34	1	1	0	0	1	1	5	22	31
k35	1	0	0	1	0	1	2	13	18
k36	0	0	0	0	0	0	0	6	6
k37	0	0	0	1	0	2	0	15	18
k38	0	0	0	1	0	0	0	7	8
k39	3	1	0	0	0	0	1	34	39
k40	3	0	0	1	0	0	0	2	6
k41	0	0	0	0	0	0	0	5	5
k42	0	0	0	0	0	0	0	1	1
k43	0	0	0	0	0	0	0	5	5
Total	24	11	3	28	6	17	24	575	688

5.1 Einsetzung der Untersuchungsvariablen

Einerseits war das Ziel, Einfluss und Zweckmässigkeit einer möglichst vielseitigen Auswahl an Einsetzungsvarianten zu vergleichen; andererseits erhöht sich der Arbeitsaufwand mit der Anzahl betrachteter Methoden beachtlich. Die schliesslich verwendete Auswahl stellt somit einen Kompromiss zwischen diesen beiden Kriterien dar. Für die formale Definition der Einsetzungsvarianten wird die Notation in Tabelle 4 verwendet:

Tabelle 4 Die im formalen Beschrieb verwendete Notation.

Term	Erklärung
$R_{L_c}(t)$	Die Menge der antwortenden sehr grossen Unternehmen der NOGA2-Klasse c in der WS200t.
$M_{L_c}(t)$	Die Menge der nicht antwortenden sehr grossen Unternehmen der NOGA2-Klasse c in der WS200t.
$s_{14}(i,t)$	Vollzeitstellenäquivalent des Unternehmens i in der WS200t.
$s_{14}(i,Z)$	Vollzeitstellenäquivalent des Unternehmens i zum Zeitpunkt der Ziehung.
$y_v(i,t)$	Wert von Unternehmen i für die Variable $v = t017, t022, t053$ in der WS200t.
$ts_{14}(i,t),\ ty_v(i,t)$	Diese Variablen gehen durch Standardisierung (Abzug vom Mittelwert und Division durch Standardabweichung, berechnet anhand der Unternehmen in $R_{L_c}(t)$) aus $y_v(i,t)$ und $s_{14}(i,t)$ hervor.

Im Folgenden werden die betrachteten Einsetzungsvarianten vorgestellt.

Methode $Mittel$

Beschreibung: Es wird der Stichprobenmittelwert der Zielvariablen in der jeweiligen Schicht eingesetzt.

Formal: Falls Unternehmen k in der WS2004 nicht geantwortet hat, kann die Berechnung der eingesetzten Werte, $y_v^*(k,4)$, als

$$y_v^*(k,4) = \frac{1}{|R_{L_c}(4)|} \sum_{i \in R_{L_c}(4)} y_v(i,4), \tag{5}$$

dargestellt werden, wobei $|R_{L_c}(4)|$ die Anzahl antwortende sehr grosse Unternehmen in der WS2004 bezeichnet.

Methode $Trend$

Beschreibung: Es wurden die an die Schichttendenzen (berechnet aufgrund jener Unterneh-

men, die sowohl im betrachteten Jahr wie auch im Vorjahr geantwortet haben) angepassten Vorjahreswerte der Zielvariablen eingesetzt.

Formal: Falls Unternehmen k in der WS2004 nicht geantwortet hat, ergeben sich die eingesetzten Werte aus

$$y_v^*(k,4) = \frac{\sum_{i \in R_{Lc}(3) \cap R_{Lc}(4)} y_v(i,4)}{\sum_{i \in R_{Lc}(3) \cap R_{Lc}(4)} y_v(i,3)} y_v(k,3), \tag{6}$$

wobei im Falle, dass Unternehmen k im Vorjahr nicht geantwortet hat, $y_v(k,3)$ durch $y_v^*(k,3)$ ersetzt wird.

Bemerkung: Der für die Trendschätzung verwendete Koeffizient ist im Grunde der Regressionskoeffizient eines Quotientenschätzers. Die Einsetzung Trend gehört somit zu den Regressions- oder genauer den Quotienteneinsetzungen ([4], Seite 305). Für die WS02 (erstes Jahr) wurde auf eine Regressionseinsetzung anhand des Vollzeitstellenäquivalentes ausgewichen.

Methode $nnbr$

Beschreibung: Es werden die Werte des nächsten Nachbarn (nearest neighbour) eingesetzt, wobei dieser mittels der L1-Distanz auf den zehn Dimensionen, welche sich aus dem Vollzeitstellenäquivalent sowie den untersuchten drei Zielvariablen während der Jahre 2002-04 ergeben, berechnet wurde. Um eine sinnvolle Distanz zu erhalten, wurden sämtliche Variablen vorgängig standardisiert (Abzug Mittelwert, Division mit Standardabweichung).

Formal: Mit der Methode $nnbr$ erhält Unternehmen k in der WS04 die Werte des nächsten Nachbarn d, mit

$$d = argmin_{i \in \cap_{t=2}^{4} R_{Lc(t)}} Dist_{L1}(\vec{v}(i), \vec{v}(k)). \tag{7}$$

Dabei bezeichnet $\vec{v}(k)$ den Vektor mit jenen Komponenten aus $ts_{14}(k,4), ty_v(k,t); t = 2,3,4; v = t017, t022, t053$ für die ein Wert vorhanden ist und $Dist_{L1}$ die $L1$-Distanz.

Bemerkung: Diese Variante wurde wegen ihrer multivariaten Eigenschaften ausgewählt. Bei der Beurteilung der Resultate muss beachtet werden, dass die Anwendung auf den zehn Dimensionen in Realität nur für die Einsetzung des Jahres 2004 möglich gewesen wäre, da beispielsweise im Jahr 2003 die Werte des Jahres 2004 für eine Distanzberechnung noch nicht zur Verfügung gestanden hätten. Falls Unternehmen i in der WS04 nicht geantwortet hat, wurde in der Distanzberechnung statt $s_{14}(i,4)$ der eingesetzte Wert $s_{14}^*(i,4)$ gemäss Abschnitt 5.2 verwendet. Die Einsetzung $nnbr$ wurde mit einem, von Herrn Daniel Kilchmann, Dienst Statistische Methoden, entwickelten SAS-Macro vorgenommen. Die Methode verwendet nur jene Unternehmen als Spender, die in sämtlichen untersuchten Jahren (2002-04) ihre Daten eingereicht haben. Diese Einsetzung gehört wie auch die unten vorgestellte Methode $nnbr_{s14}$ zu den Hot-Deck-Einsetzungen ([3], Seite 592).

Methode $nnbr_{s14}$

Beschreibung: Einsetzen der Werte des nächsten Nachbarn, ermittelt mittels dem Vollzeitstellenäquivalent.

Formal: Die Methode $nnbr_{s14}$ funktioniert analog zu $nnbr$, in die Distanzberechnung fliesst aber nur $s_{14}(i,t)$ ein.

Bemerkung: Die Variante mit dieser univariaten Distanz hat gegenüber $nnbr$ den Vorteil, dass sich die Anzahl der Spender erhöht. Zudem geht man möglichen Inkonsistenzproblemen aus dem Weg, die sich beim multivariaten Ansatz über die Jahre hinweg ergeben könnten (bei Spenderwechsel würden auf einmal auch weiter zurückliegende Einsetzungswerte geändert).

5.2 Einsetzung der Hilfsvariablen

Genau genommen lässt sich das Vollzeitstellenäquivalent des Unternehmens i im Jahr t als

$$s_{14}(i,t) = s_{14,Besta}(i,t) + s_{0016}(i,t) \tag{8}$$

schreiben, wobei $s_{14,Besta}$ das Vollzeitstellenäquivalent der Angestellten und s_{0016} jenes der Geschäftsinhaber von Einzel- und Personengesellschaften bezeichnet. Wie in Abschnitt 3 erwähnt, werden fehlende Werte für die NOGA2-Klassifizierung und das Vollzeitstellenäquivalent mittels dem letzten beobachteten Wert eingesetzt:

$$s^*_{14}(i,t) = s_{14}(i,t-1),$$

wobei $s_{14}(i,t-1)$ im Fall, dass i im Jahr $t-1$ nicht geantwortet hat, durch $s^*_{14}(i,t-1)$ ersetzt wird. Der Einfluss von s_{0016} ist bei sehr grossen Unternehmen vernachlässigbar [3] und es wird $s^*_{14,Besta}(i,t) = s^*_{14}(i,t)$ gesetzt. Die Variable $s_{14,Besta}$ fliesst in die Berechnung des Populationstotales des Vollzeitstellenäquivalentes ($X_{Tot,c*}$) ein. Der letzte beobachtete Wert ist kaum eine optimale Einsetzungsmethode, bietet aber den Vorteil, dass nicht durch die Einsetzung selbst grosse Unternehmen kreiert werden [4]. Um die Einsetzungen in die zu publizierenden Schätzungen einfliessen zu lassen, sollten die Hilfsvariablen für die relevanten Unternehmen beispielsweise mittels anderen Datenquellen oder Rückfragen (und nicht mittels dem letzten beobachteten Wert) eingesetzt werden.

6 Der Einfluss der Einsetzungen auf die geschätzten Totale

Mittels jeder der vorgestellten Einsetzungsmethoden wurden für die Jahre 2002-04 die Extrapolationswerte der Variablen Bruttoproduktion ($t017$), Vorleistungen ($t022$) sowie Personalkosten ($t053$) neu berechnet. Die Resultate wurden den mittels der aktuellen Methode berechneten Schätzungen gegenübergestellt. Bei diesem Vergleich stehen folgende zwei Kriterien im Vordergrund:

- Wie gross wären die zu erwartenden Differenzen gegenüber der aktuellen Methode?
- Wie würde eine Einsetzung die Veränderungen zwischen den Schätzungen zwei aufeinanderfolgender Jahre beeinflussen?

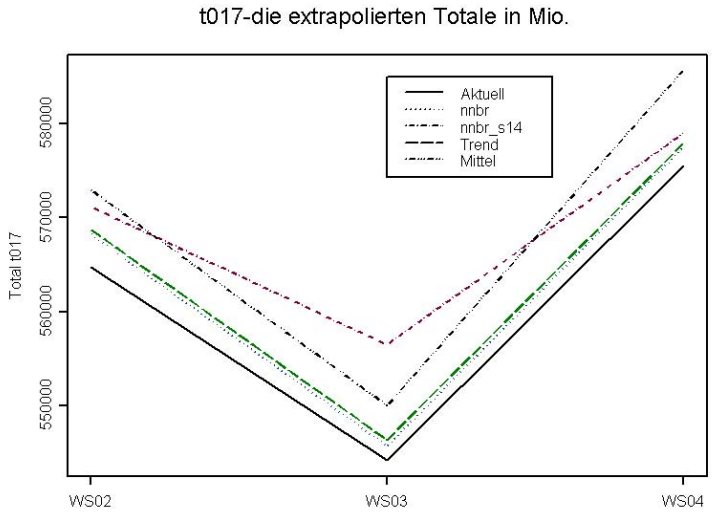
6.1 Die Veränderungen der geschätzten Totale gegenüber der aktuellen Methode

In Abbildung 1 werden die global (Aggregation sämtlicher NOGAs) geschätzten Totale der Bruttoproduktion (t017) für die verschiedenen Einsetzungsmethoden jenen der aktuellen Methode ($Aktuell$) gegenübergestellt.

[3] Beispielsweise ergaben sich bei den sehr grossen Unternehmen, welche in der WS04 geantwortet haben, folgende Totale: $s_{0016} = 12$, $s_{14} = 658043$.

[4] Wie in Abschnitt 3 beschrieben, beruht die Definition der sehr grossen Unternehmen auf dem Wert der Variablen s_{14}.

Abbildung 1 Vergleich der mittels verschiedenen Varianten berechneten Extrapolationstotale.



Es fällt auf, dass die aktuelle Methode jeweils die kleinsten Werte aufweist. Diese Beobachtung wird durch Tabelle 5, welche die relativen Differenzen der mittels den Einsetzungsmethoden erhaltenen (globalen) Schätzungen gegenüber den aktuellen enthält [5], auch für die Variablen t022 und t053 bestätigt. Aus Tabelle 5 lässt sich ebenfalls ablesen, dass die relativen Veränderungen für die Methode

- $Mittel$ im Bereich $[0.71\%, 2.00\%]$
- $nnbr$ im Bereich $[0.04\%, 0.63\%]$
- $nnbr_{s14}$ im Bereich $[0.08\%, 2.69\%]$
- $Trend$ im Bereich $[0.08\%, 0.85\%]$

liegen. Um die Grössenordnungen der beobachteten relativen Differenzen besser einschätzen zu können, wurde Tabelle 5 mit der Grösse $1.96 \cdot CV$ ergänzt. Dabei bezeichnet CV den Variationskoeffizienten des mit der aktuellen Methode geschätzten Totales, $\hat{Y}_{Tot,Aktuell}$. Falls eine relative Differenz kleiner als $1.96 \cdot CV$ ist, kann mit Sicherheit gesagt werden, dass sich der Extrapolationswert der entsprechenden Methode nicht signifikant (zu einem 5 % Niveau) von jenem der aktuellen Methode unterscheidet. Das Umgekehrte kann nicht behauptet werden, da die Aussage nur auf einem Vertrauensintervall der aktuellen Methode beruht[6]. In Tabelle 5 überschreiten nur zwei Werte der Methode $Mittel$ die Grenze $1.96 \cdot CV$. Weiter fällt auf, dass insbesondere $Trend$ und $nnbr$ Werte lieferten, die nie auch nur in der Nähe von $1.96 \cdot CV$ liegen.

[5] *Die relative Differenz zwischen den Schätzungen der Methode $Mittel$ in bezug zu $Aktuell$ wäre beispielsweise* $\frac{\hat{Y}_{Tot,Mittel} - \hat{Y}_{Tot,Aktuell}}{\hat{Y}_{Tot,Aktuell}}$

[6] Die Normalverteilungsapproximation liefert $P(Y_{Tot} \in [\hat{Y}_{Tot,Aktuell} - 1.96 \cdot \hat{\sigma}(\hat{Y}_{Tot,Aktuell}), \hat{Y}_{Tot,Aktuell} + 1.96 \cdot \hat{\sigma}(\hat{Y}_{Tot,Aktuell})]) \approx 95\%$. Aus einer Division der Intervallgrenzen mit $\hat{Y}_{Tot,Aktuell}$ folgt $P(\frac{Y_{Tot}}{\hat{Y}_{Tot,Aktuell}} \in [1 - 1.96 \cdot CV, 1 + 1.96 \cdot CV]) \approx 95\%$. Nun wird noch Eins subtrahiert, was auf das gewünschte Vertrauensintervall mit der oberen Grenze $1.96 \cdot CV$ führt: $P(\frac{Y_{Tot} - \hat{Y}_{Tot,Aktuell}}{\hat{Y}_{Tot,Aktuell}} \in [-1.96 \cdot CV, +1.96 \cdot CV]) \approx 95\%$.

Tabelle 5 Relative Differenzen in Prozent.

WS	Methode	t017	t022	t053
2002	$Mittel$	1.45	1.65	1.03
	$nnbr$	0.61	0.59	0.49
	$nnbr_{s14}$	1.13	1.49	0.12
	$Trend$	0.70	0.85	0.32
	$1.96 \cdot CV$	1.92	2.80	1.12
2003	$Mittel$	1.07	1.06	0.71
	$nnbr$	0.28	0.37	0.16
	$nnbr_{s14}$	2.26	2.69	0.29
	$Trend$	0.40	0.42	0.22
	$1.96 \cdot CV$	3.08	3.02	1.16
2004	$Mittel$	**1.76**	2.00	**1.17**
	$nnbr$	0.34	0.63	0.04
	$nnbr_{s14}$	0.60	1.04	0.08
	$Trend$	0.41	0.60	0.08
	$1.96 \cdot CV$	1.35	2.08	1.02

Die relativen Differenzen in Bezug zur aktuellen Methode lassen sich auch auf Stufe NOGA2 berechnen. In Tabelle 6 sind die Mittelwerte und Standardabweichungen dieser Differenzen dargestellt. Dabei wurden nur jene NOGA2-Klassen berücksichtigt, in welchen tatsächlich fehlende Werte auftraten [7], so wurden die Mittelwerte für die WS2002 anhand von 26, für die WS2003 anhand von 22 und für die WS2004 anhand von 23 NOGA2-Klassen berechnet. Bei den Resultaten fällt auf, dass sämtliche Mittelwerte positiv sind.

Die beobachteten Differenzen weisen darauf hin, dass Einsetzungen die Jahresschätzungen tendenziell erhöhen würden.

Auch hier lassen sich die Bereiche ablesen, in welchen sich die durchschnittlichen relativen Veränderungen bewegten:

- $Mittel$ im Bereich $[2.14\%, 2.68\%]$
- $nnbr$ im Bereich $[0.56\%, 2.90\%]$
- $nnbr_{s14}$ im Bereich $[0.64\%, 2.87\%]$
- $Trend$ im Bereich $[0.59\%, 1.84\%]$

Die relativ (zu den beobachteten Mittelwerten) hohen Standardabweichungen zeigen, dass die beobachteten relativen Effekte in den verschiedenen NOGA2-Klassen stark variierten. Dies bestätigt sich in Abbildung 2, wo die relativen Differenzen für die Variable $t017$ in Form von Boxplots dargestellt sind. Für die Variablen t022 und t053 ergäbe sich ein ähnliches Bild. In den Boxplots zeigt sich, dass die Mediane, ausser für die Methode $Mittel$, nahe bei 0% liegen. Bei der Einsetzung mit dem Mittelwert waren drei Viertel aller betrachteten Differenzen positiv.

[7] *Formal: Falls $\hat{Y}_{Tot,c,Aktuell}$ den Schätzwert für die NOGA2-Klasse c der aktuellen Methode und $\hat{Y}_{Tot,c,Mittel}$ jenen der Methode Mittel bezeichnet, ergibt sich die relative Differenz $d_{1,c} = \frac{\hat{Y}_{Total,c,Mittel} - \hat{Y}_{Total,c,Aktuell}}{\hat{Y}_{Total,c,Aktuell}}$. Mittelwert und Standardabweichung werden anhand der Werte $d_{1,c}, c \in I$, berechnet, wobei I die Menge jener NOGAs bezeichnet, in welchen für das betrachtete Jahr Einsetzungen vorgenommen worden sind.*

Tabelle 6 Mittelwerte und Standardabweichungen der in NOGAs mit fehlenden Werten berechneten relativen Differenzen in Prozent.

WS	Methode	Variable	Mittelwert (%)	Standardabweichung (%)
2002	$Mittel$	t017	2.55	2.43
	$Mittel$	t022	2.65	2.30
	$Mittel$	t053	2.17	2.60
	$nnbr$	t017	2.66	6.93
	$nnbr$	t022	2.90	6.84
	$nnbr$	t053	1.93	4.44
	$nnbr_{s14}$	t017	1.90	2.90
	$nnbr_{s14}$	t022	2.20	2.71
	$nnbr_{s14}$	t053	0.64	1.11
	$Trend$	t017	1.45	2.01
	$Trend$	t022	1.78	2.41
	$Trend$	t053	0.91	1.68
2003	$Mittel$	t017	2.64	2.98
	$Mittel$	t022	2.68	2.74
	$Mittel$	t053	2.14	3.34
	$nnbr$	t017	1.01	1.32
	$nnbr$	t022	1.30	1.72
	$nnbr$	t053	0.77	1.01
	$nnbr_{s14}$	t017	2.32	3.83
	$nnbr_{s14}$	t022	2.87	5.76
	$nnbr_{s14}$	t053	0.65	0.73
	$Trend$	t017	1.23	1.71
	$Trend$	t022	1.47	1.86
	$Trend$	t053	0.71	1.04
2004	$Mittel$	t017	2.50	2.53
	$Mittel$	t022	2.58	2.38
	$Mittel$	t053	2.20	2.38
	$nnbr$	t017	1.16	1.80
	$nnbr$	t022	1.68	2.71
	$nnbr$	t053	0.56	0.72
	$nnbr_{s14}$	t017	1.31	1.74
	$nnbr_{s14}$	t022	1.78	2.50
	$nnbr_{s14}$	t053	0.56	0.76
	$Trend$	t017	1.40	2.12
	$Trend$	t022	1.84	2.91
	$Trend$	t053	0.59	0.94

Abbildung 2 Boxplot-Darstellung der über die drei Jahre in den verschiedenen NOGAs mit fehlenden Werten beobachteten relativen Differenzen.

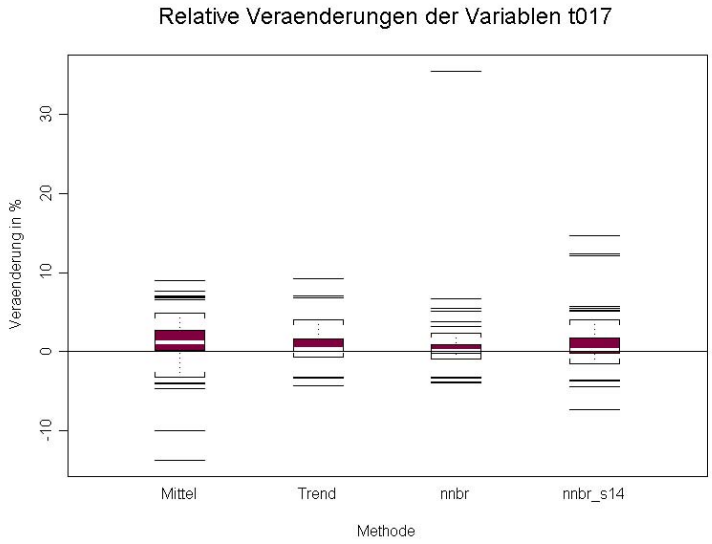


Tabelle 7 Grosse relative Differenzen in Bezug auf die aktuelle Methode.

Fall	Methode	WS	noga2	t017	t022	t053
1	$Mittel$	ws2002	k19	-9.97%	-8.78%	-10.70%
2	$Trend$	ws2002	k19	9.26%	11.30%	8.06%
3	$Mittel$	ws2003	k19	-13.80%	-13.50%	-15.20%
4	$nnbr_{s14}$	ws2003	k28	14.60%	27.50%	1.83%
5	$nnbr$	ws2004	k34	6.69%	10.60%	-0.58%
6	$Trend$	ws2004	k34	7.08%	10.50%	-0.24%
7	$nnbr_{s14}$	ws2002	k35	12.10%	9.55%	0.61%
8	$nnbr_{s14}$	ws2003	k35	12.40%	3.32%	0.58%
9	$nnbr$	ws2002	k38	35.50%	35.00%	22.70%

Die Verteilungen der relativen Differenzen sind für die Methoden $nnbr$, $nnbr_{s14}$ sowie $Trend$ stark rechtsschief, was die durchwegs positiven Mittelwerte erklärt. Die Einsetzungsmethoden führen in einzelnen NOGA2-Klassen zu grossen relativen Abweichungen in Bezug auf die aktuelle Methode. In Tabelle 7 werden Differenzen aufgelistet, welche ausserhalb des Bereiches $\pm 10\%$ liegen. Es wurde versucht, für jede dieser grossen Differenzen eine Erklärung zu geben. Dabei zeigte sich, dass die extremen Veränderungen durch einzelne Einsetzungen verursacht werden können. Die untersuchten Beispiele lassen zudem vermuten, dass von den vier betrachteten Einsetzungsvarianten $nnbr$ und $Trend$ Werte liefern, die auf Ebene Unternehmen plausibler sind als die Einsetzungen von $Mittel$ und $nnbr_{s14}$.

Einzelne Einsetzungen für sehr grosse Unternehmen können die Resultate auf Ebene NOGA2 stark beeinflussen. In diesen Fällen sind Massnahmen wie unternehmensspezifische Überprüfung der eingesetzten Werte bis hin zu einem Rückruf nötig.

6.2 Der Einfluss von Einsetzungen auf die relativen Veränderungen zwischen zwei Jahren

In Diskussionen zu den Extrapolationsresultaten spielen Veränderungen gegenüber Vorjahreswerten eine wichtige Rolle, wobei für grössere Sprünge eine Erklärung gesucht werden muss. In diesem Abschnitt wird untersucht, inwieweit eine allfällige Einsetzung diese Veränderungen beeinflussen würde. Dazu wurden für sämtliche NOGAs mit fehlenden Werten die Absolutbeträge der relativen Veränderungen zwischen den Jahren 2002/03, respektive 2003/04, berechnet. Ausgehend von diesen Werten wurden dann die in Tabelle 8 aufgelisteten Mittelwerte und Standardabweichungen berechnet[8]. Um den Vergleich der Methoden zu vereinfachen, wurde in Tabelle 8 eine Rangvariable eingefügt, welche die Methoden für jede Zielvariable (t017, t022, t053) nach den berechneten Mittelwerten rangiert.

Es zeigt sich, dass

1. es keine Methode gibt, die in allen Fällen die kleinsten oder in allen Fällen die grössten Werte liefert,
2. man nicht davon ausgehen darf, dass eine Einsetzung zu einer Reduktion der betrachteten relativen Veränderungen führen würde.

Bildet man für jede Methode das Total der Rangvariablen, erhält man für $Trend$ und $Aktuell$ 13, für $nnbr$ 14, für $nnbr_{s14}$ 23 und für $Mittel$ 27. $Trend$, $Aktuell$ und $nnbr$ lieferten also im Vergleich zu $Mittel$ und $nnbr_{s14}$ eher kleinere Veränderungen. Zu beachten ist, dass die berechneten Standardabweichungen relativ (im Vergleich zu den errechneten Mittelwerten) hoch sind.

7 Qualitätsvergleich der Einsetzungen

Während im vorangehenden Abschnitt der Einfluss der Einsetzungsmethode auf die Extrapolationsresultate im Vordergrund stand, wird in diesem Abschnitt die Qualität der Einsetzungsvarianten aufgrund von Vergleichen der eingesetzten mit den tatsächlichen Werten für die Variable t017, Bruttoproduktion, untersucht. Im aus dem Euredit Projekt resultierenden Handbuch, *Towards Effective Statistical Editing and Imputation Strategies - Findings of the Euredit Project*, werden verschiedene Masszahlen vorgeschlagen, welche für solche Vergleiche verwendet werden können. Diese hatten auch einen massgebenden Einfluss auf die Auswahl an Masszahlen, welche im Rahmen dieser Arbeit verwendet wurde. Bei den Einsetzungsmethoden wurde bewusst auf das Einfügen von Zufallskomponenten verzichtet, da das Problem der Varianzschätzungen im Rahmen dieser Arbeit nicht behandelt wird. Konsequenterweise beschränkt sich der Vergleich der verschiedenen Einsetzungsmethoden auf Masszahlen, welche Angaben dazu liefern, ob die Einsetzungen die tatsächlichen Werte unverfälscht und genau schätzen. Masszahlen, welche die höheren Momente und allenfalls gar die Verteilung der eingesetzten Werte mit jenen der tatsächlichen Werte vergleichen, werden nicht betrachtet. Da die echten

[8] Formal: Falls beispielsweise $\hat{Y}_{Tot,c,Aktuell}(t)$ den Schätzwert für die NOGA2-Klasse c mit der aktuellen Methode für Jahr t und $\hat{Y}_{Tot,c,Aktuell}(t-1)$ jenen für Jahr $t-1$ bezeichnet, ergibt sich die relative Veränderung $d_{2,c} = \frac{\hat{Y}_{Tot,c,Aktuell}(t) - \hat{Y}_{Tot,c,Aktuell}(t-1)}{\hat{Y}_{Tot,c,Aktuell}(t-1)}$. Mittelwert und Standardabweichung werden anhand der Absolutbeträge, $|d_{2,c}|, c \in I_{t-1,t}$ berechnet, wobei $I_{t-1,t}$ die Menge jener NOGAs bezeichnet, in welchen in mindestens einem der Jahre $t-1$ und t Einsetzungen vorgenommen worden sind.

Tabelle 8 Absolute prozentuale Veränderungen zwischen zwei aufeinanderfolgenden Jahren.

WS	Methode	Variable	Mittelw. (%)	Standardabw. (%)	Rang	Rangsumme
2002 versus 2003	$Aktuell$	t017	8.44	12.47	3	
	$Aktuell$	t022	11.21	18.31	3	
	$Aktuell$	t053	3.98	2.88	1	7
	$Mittel$	t017	8.58	12.74	4	
	$Mittel$	t022	11.41	18.58	4	
	$Mittel$	t053	4.31	3.68	4	12
	$nnbr$	t017	6.98	7.20	1	
	$nnbr$	t022	9.38	10.63	1	
	$nnbr$	t053	4.64	3.41	5	7
	$nnbr_{s14}$	t017	9.14	14.65	5	
	$nnbr_{s14}$	t022	12.42	20.53	5	
	$nnbr_{s14}$	t053	4.11	3.37	3	13
	$Trend$	t017	8.13	11.85	2	
	$Trend$	t022	11.08	17.31	2	
	$Trend$	t053	4.09	3.06	2	6
2003 versus 2004	$Aktuell$	t017	6.73	6.90	1	
	$Aktuell$	t022	7.50	6.69	1	
	$Aktuell$	t053	3.97	4.49	4	6
	$Mittel$	t017	8.44	10.23	5	
	$Mittel$	t022	9.02	9.31	5	
	$Mittel$	t053	5.71	9.94	5	15
	$nnbr$	t017	6.79	7.29	2	
	$nnbr$	t022	7.82	7.50	3	
	$nnbr$	t053	3.78	4.53	2	7
	$nnbr_{s14}$	t017	6.91	7.01	3	
	$nnbr_{s14}$	t022	8.09	7.44	4	
	$nnbr_{s14}$	t053	3.87	4.35	3	10
	$Trend$	t017	7.05	6.72	4	
	$Trend$	t022	7.60	7.23	2	
	$Trend$	t053	3.73	4.17	1	7

Werte nicht antwortender Unternehmen nicht bekannt waren, wurden die Voraussetzungen für die gewünschten Vergleiche, wie im nächsten Abschnitt beschrieben, mittels einer Simulation geschaffen.

7.1 Rahmenbedingungen der Simulation

Für eine zufällige Auswahl von Unternehmen sollen die Antworten von 2004 als fehlend definiert und mit den verschiedenen Einsetzungsmethoden ersetzt werden. Danach können die eingesetzten mit den echten Werten verglichen werden. Die Simulation basiert auf 610 sehr grossen Unternehmen, von denen die Daten der WS 2004 vorhanden waren und die sich in einer NOGA2-Klasse befanden, für welche mindestens fünf Unternehmen geantwortet haben. Mit dieser Einschränkung, welche 21 Unternehmen aus der Simulation ausschliesst, wird vermieden, dass Problemfälle auftauchen, in denen durch die Simulation beispielsweise sämtliche Beobachtungen einer NOGA2-Klasse gelöscht und die betrachteten Einsetzungsmethoden somit nicht mehr funktionieren würden. Die Antwortquote lag bei den sehr grossen Unternehmen für die WS 2004 bei 91.7 %. Um die Einsetzungen miteinander zu vergleichen, wurden darum

mittels der Prozedur Surveyselect (SAS) aus den Unternehmen, die in der WS04 geantwortet haben, fünf Unterstichproben zu je 51 Unternehmen (8.3 %) zufällig ausgewählt. In jeder dieser Unterstichprobe wurden anschliessend die Bruttoproduktionszahlen (t017) für sämtliche Unternehmen mittels den vier betrachteten Einsetzungsmethoden berechnet.

Begründung des Vorgehens: Eine Alternative wäre gewesen, die Unterstichproben stratifiziert nach NOGA2-Kategorien auszuwählen. Aus Tabelle 2 kann man entnehmen, dass die Antwortraten in den verschiedenen NOGA2-Kategorien aufgrund der teilweise kleinen Anzahl Unternehmen zwischen 2003 und 2004 recht stark variieren und die Gesamtantwortquote hingegen relativ stabil bleibt. Darum wurde jene Lösung vorgezogen, in welcher die Anzahl fehlender Werte in einer NOGA2-Kategorie nicht fixiert ist. Es ist möglich, dass ein Verfahren je nach Beschaffenheit der gezogenen Unterstichprobe besser oder schlechter funktioniert. Um dieser Tatsache Rechnung zu tragen, wurden fünf Unterstichproben gezogen. Zudem werden durch diesen Ansatz auch Vergleiche der Methoden auf Ebene NOGA2 möglich. (Sonst ist die Anzahl ersetzter Werte in den einzelnen NOGAs hierzu meist zu klein).

Notation: Die Menge der Unternehmen aus der NOGA2-Klasse c die sich in der k-ten (k=1,...,5) Unterstichprobe befindet, wird mit $M_{c,k}$ bezeichnet. Im Zusammenhang mit den Simulationen wird weiter folgende Terminologie verwendet:

- $M_k = \cup_c M_{c,k}$
- die 51 Unternehmen in M_k werden mit den Indices $I_k = \{k_1, ..., k_{51}\}$ identifiziert
- $m_{c,k}$ bezeichnet die Anzahl Unternehmen in $M_{c,k}$
- $m_c = \sum_{k=1}^{5} m_{c,k}$
- Für jedes Unternehmen k_i bezeichnet im Folgenden y_{k_i} den tatsächlichen und $y_{k_i,meth}$, $meth = Trend, Mittel, nnbr, nnbr_{s14}$, den mit der jeweiligen Methode eingesetzten Wert für die Bruttoproduktion.

Die Werte der Variablen $m_{c,k}$ sowie m_c, welche sich aus der Simulation ergeben haben, sind in Tabelle 9 dargestellt.

7.2 Die Distanz zwischen echten und eingesetzten Werten

Die Differenz

$$y_{k_i,meth} - y_{k_i} = \gamma_{k_i,meth} \tag{9}$$

gibt den Fehler an, den man begeht, falls für den Wert y_{k_i} der Wert $y_{k_i,meth}$ eingesetzt wird. Hierzu werden in Tabelle 10 die mittleren L1-Distanzen[9],

$$D^{(k)}_{meth} = \frac{1}{51} \sum_{i=1}^{51} |\gamma_{k_i,meth}| \tag{10}$$

angegeben, welche sich ungeachtet der NOGA2-Kategorie in den fünf gezogenen Stichproben ergeben haben. Da das Interesse eher dem Vergleich zwischen den verschieden Methoden und weniger den absoluten Werten gilt, wurden die Werte für jede Unterstichprobe (k) rangiert (Variable Rg). Dabei erhielt die Methode mit dem grössten Wert Rang 4 und jene mit dem kleinsten Wert Rang 1. Die Methode Trend schneidet in allen fünf Unterstichproben am besten

[9] Die L1-Distanz liefert einfach interpretierbare Resultate, darum wurde sie anderen möglichen Distanzen vorgezogen.

Tabelle 9 Häufigkeitstabelle der fehlenden Werte aus der Simulation.

c (noga2)	$m_{c,1}$	$m_{c,2}$	$m_{c,3}$	$m_{c,4}$	$m_{c,5}$	m_c	Schichtgrösse
k2	2	5	1	2	4	14	28
k8	2	1	0	1	2	6	11
k9	1	0	2	0	1	4	18
k11	2	7	5	2	4	20	35
k12	1	3	0	0	1	5	11
k13	0	0	1	0	1	2	7
k14	0	1	0	2	2	5	11
k15	0	2	1	2	1	6	14
k16	6	4	11	6	7	34	65
k18	2	2	3	1	0	8	23
k19	0	0	1	0	1	2	10
k20	4	1	3	3	2	13	40
k22	1	0	0	0	1	2	6
k23	0	0	0	0	0	0	6
k25	1	0	1	1	0	3	14
k26	5	3	5	1	3	17	39
k27	1	1	0	1	2	5	10
k28	5	2	2	4	4	17	49
k29	6	5	3	9	4	27	60
k30	1	1	0	1	2	5	13
k31	0	0	0	3	0	3	8
k32	2	1	1	2	0	6	16
k34	4	1	3	2	2	12	24
k35	0	3	1	1	2	7	15
k36	0	0	1	1	0	2	6
k37	2	3	1	3	0	9	18
k38	0	0	0	1	0	1	8
k39	2	5	3	1	5	16	35
k41	1	0	2	1	0	4	5
k43	0	0	0	0	0	0	5
Total	51	51	51	51	51	255	610

ab. Es darf also davon ausgegangen werden, dass die Einsetzungen aus der Variante Trend im Durchschnitt über alle NOGAs hinweg am nächsten bei den effektiven Werten liegen. Ebenso klar liegt die Methode $nnbr$ auf dem zweiten und Mittel (wie vermutet werden durfte) auf dem letzten Rang. Nun wird untersucht, wie es sich mit den Abständen zwischen den eingesetzten und den effektiven Werten in einzelnen NOGAs verhält. Dazu werden in Tabelle 11 die mittleren L1-Distanzen nach NOGA2-Klassen verglichen. Die Werte in der Tabelle ergeben sich aus

$$D_{c,meth} = \frac{1}{m_c} \sum_{k=1}^{5} \sum_{i=1}^{51} |\gamma_{k_i}| \cdot 1(k_i \in c), \tag{11}$$

wobei die Indikatorfunktion $1(k_i \in c)$ den Wert 1 annimmt, falls Unternehmen k_i in der NOGA2-Klasse c liegt und sonst 0 ist. Die aufgelisteten Vergleiche beschränken sich auf NOGA2-Klassen mit $m_c \geq 10$. Es bestätigt sich, dass die mittels Trend eingesetzten Werte im Durchschnitt am nächsten bei den tatsächlichen Werten liegen. Dabei ist beeindruckend, dass dies in jeder der betrachteten NOGA2-Kategorien der Fall ist.

Tabelle 10 Die mittlere L1-Distanz zwischen den eingesetzten und den tatsächlichen Werten (RS=Rangsumme, Rg = Rangierung der Methode nach den Distanzwerten).

k=	$D^{(k)}_{Mittel}$	Rg	$D^{(k)}_{Trend}$	Rg	$D^{(k)}_{nnbr_{s14}}$	Rg	$D^{(k)}_{nnbr}$	Rg
1	205217	4	30914	1	129349	3	60408	2
2	534118	4	90250	1	398630	3	262980	2
3	389398	4	57750	1	150281	3	100574	2
4	454991	4	42549	1	336179	3	299376	2
5	295720	4	41228	1	192312	3	68401	2
RS		20		5		15		10

Tabelle 11 Die mittlere L1-Distanz zwischen den eingesetzten und den tatsächlichen Werten nach NOGA2 (c = NOGA2-Klasse, RS=Rangsumme, Rg = Rangierung der Methode anhand der Distanzenwerte).

c	$D_{c,Mittel}$	Rg	$D_{c,Trend}$	Rg	$D_{c,nnbr_{s14}}$	Rg	$D_{c,nnbr}$	Rg
k16	90227	4	20548	1	54567	3	32041	2
k29	329989	4	37939	1	181763	3	167663	2
k11	1680711	4	141505	1	834353	3	570660	2
k26	110588	4	28269	1	80706	3	44164	2
k28	656067	4	85316	1	555500	3	406802	2
k39	228935	4	44189	1	199095	3	109835	2
k2	223949	4	20284	1	182616	3	78068	2
k20	437572	3	204351	1	489381	4	283904	2
k34	187455	3	30385	1	198172	4	56148	2
RS		34		9		29		18

7.3 Der Vergleich der ersten Momente

Kleine Werte von $|\gamma_{k_i,meth}|$ garantieren noch nicht, dass mit den Einsetzungen keine Verfälschung einfliesst. Eine diesbezüglich gute Methode zeichnet sich dadurch aus, dass der Durchschnitt der Terme $\gamma_{k_i,meth}$ nahe bei Null liegt. Hierzu werden in Tabelle 12 die Beträge der Mittelwertsdifferenzen (MD),

$$MD^{(k)}_{meth} = |\frac{1}{51}\sum_{i}\gamma_{k_i,meth}| \tag{12}$$

angegeben, welche sich ungeachtet der NOGA2-Kategorie in den fünf gezogenen Stichproben ergeben haben. Die Methode Trend schneidet auch hier mit Abstand am besten ab. Bei den anderen drei Methoden sind die Unterschiede nicht mehr so eindeutig wie bei der mittleren L1-Distanz. Nun wird auch für dieses Kriterium untersucht, wie sich die Methoden in einzelnen NOGAs bewähren. Dazu werden in Tabelle 13 die Beträge der Mittelwertsdifferenzen nach NOGA2-Klassen verglichen. Die Werte in der Tabelle ergeben sich aus

$$MD_{c,meth} = |\frac{1}{m_c}\sum_{k=1}^{5}\sum_{i=1}^{51}\gamma_{k_i}\cdot 1(k_i \in c)|. \tag{13}$$

Die aufgelisteten Vergleiche beschränken sich auf NOGA2-Klassen mit $m_c \geq 10$. Auch unter dieser Betrachtung liefert Trend im Allgemeinen die besten Resultate, wobei die Methode in der NOGA k20 nur auf dem dritten Rang liegt. Ansonsten belegte Trend in sämtlichen Betrachtungen mindestens den zweiten Rang.

Tabelle 12 Die absolute Differenz der Mittelwerte (MD) zwischen den eingesetzten und den tatsächlichen Werten (RS=Rangsumme, Rg = Rangierung der Methoden anhand der MD-Werte).

k=	$MD^{(k)}_{Mittel}$	Rg	$MD^{(k)}_{Trend}$	Rg	$MD^{(k)}_{nnbr_{s14}}$	Rg	$MD^{(k)}_{nnbr}$	Rg
1	82644	4	12879	2	1950	1	15923	3
2	238660	4	1541	1	176886	2	235833	3
3	24993	2	2129	1	79290	4	78489	3
4	181567	2	5479	1	285198	4	239415	3
5	74101	4	26266	2	33041	3	16684	1
RS		16		7		14		13

Tabelle 13 Die absolute Differenz der Mittelwerte (MD) zwischen den eingesetzten und den tatsächlichen Werten nach NOGA2 (RS=Rangsumme, Rg = Rangierung der Methode anhand der MD-Werte).

	Mittel		Trend		$nnbr_{s14}$		$nnbr$	
c	$MD_{c,Mittel}$	Rg	$MD_{c,Trend}$	Rg	$MD_{c,nnbr_{s14}}$	Rg	$MD_{c,nnbr}$	Rg
k16	16705	3	7714	1	23069	4	10863	2
k29	144360	3	14056	1	158667	4	140820	2
k11	193515	2	59324	1	388560	3	525389	4
k26	631	1	12425	2	48339	4	26645	3
k28	228913	2	22666	1	403557	4	344193	3
k39	88490	4	20740	2	55746	3	9665	1
k2	40278	3	5486	1	133862	4	24663	2
k20	210054	4	193326	3	21899	1	156448	2
k34	43460	4	9404	2	91	1	28014	3
RS		26		14		28		22

7.4 Der lineare Zusammenhang zwischen eingesetzten und tatsächlichen Werten

In diesem Abschnitt wird davon ausgegangen, dass der Zusammenhang zwischen eingesetzten $y_{k_i,meth}$ und echten Werten y_{k_i} durch

$$y_{k_i} = \beta_{meth} \cdot y_{k_i,meth} + \epsilon_{k_i,meth} \tag{14}$$

beschrieben wird. In Realität setzt man für y_{k_i} den Wert $y_{k_i,meth}$ (und nicht $\beta_{meth} \cdot y_{k_i,meth} + \epsilon_{k_i,meth}$) ein, daher wird der gemachte Fehler kleiner, falls β_{meth} nahe bei Eins und $\epsilon_{k_i,meth}$ nahe bei Null liegt. Im Folgenden werden diese zwei Kriterien betrachtet.

7.4.1 Der Steigungsparameter

Zum Vergleich der Einsetzungsmethoden werden in Tabelle 14 die geschätzten Parameter

$$\hat{\beta}^{(k)}_{meth} = \frac{\sum_i y_{k_i} \cdot y_{k_i,meth}}{\sum_i y^2_{k_i,meth}} \tag{15}$$

angegeben, welche sich ungeachtet der NOGA2-Kategorie in den fünf gezogenen Stichproben ergeben. Interessanter als die Werte der Steigungsparameter ist deren Abstand zu 1. Darum sind die Werte $|\hat{\beta}^{(k)}_{meth} - 1|$, aufgrund welcher die Methoden schliesslich auch rangiert werden, in Tabelle 14 ebenfalls aufgelistet. Während sich die anderen Methoden bezüglich diesem Kriterium nicht gross unterscheiden, schneidet Trend klar am besten ab.

Tabelle 14 Die geschätzten Steigungsparameter einer linearen Regression ohne Konstante, global. (Rg = Rangierung der Methoden nach $|\hat{\beta}-1|$, RS = Rangsumme)

	Mittel			Trend			$nnbr_{s14}$			$nnbr$		
k=	$\hat{\beta}^{(k)}$	$\lvert\hat{\beta}^{(k)}-1\rvert$	Rg	$\hat{\beta}^{(k)}$	$\lvert\hat{\beta}^{(k)}-1\rvert$	Rg	$\hat{\beta}^{(k)}$	$\lvert\hat{\beta}^{(k)}-1\rvert$	Rg	$\hat{\beta}^{(k)}$	$\lvert\hat{\beta}^{(k)}-1\rvert$	Rg
1	0.47	0.53	4	0.95	0.05	1	0.74	0.26	3	0.87	0.13	2
2	1.44	0.44	2	1.11	0.11	1	1.70	0.70	3	2.07	1.07	4
3	0.92	0.08	2	1.01	0.01	1	1.25	0.25	3	1.27	0.27	4
4	0.86	0.14	2	0.99	0.01	1	2.46	1.46	3	2.93	1.93	4
5	0.66	0.34	4	0.88	0.12	2	0.84	0.16	3	1.06	0.06	1
RS			14			6			15			15

Tabelle 15 Die geschätzten Steigungsparameter einer linearen Regression ohne Konstante, nach NOGA2 (c). (Rg = Rangierung der Methoden nach $|\hat{\beta}-1|$, RS = Rangsumme)

	Mittel			Trend			$nnbr_{s14}$			$nnbr$		
c	$\hat{\beta}$	$\lvert\hat{\beta}-1\rvert$	Rg	$\hat{\beta}$	$\lvert\hat{\beta}-1\rvert$	Rg	$\hat{\beta}$	$\lvert\hat{\beta}-1\rvert$	Rg	$\hat{\beta}$	$\lvert\hat{\beta}-1\rvert$	Rg
k16	0.91	0.09	3	1.02	0.02	1	1.12	0.12	4	1.08	0.08	2
k29	1.40	0.40	2	1.13	0.13	1	2.26	1.26	4	2.10	1.10	3
k11	0.86	0.14	2	1.11	0.11	1	1.73	0.73	3	2.02	1.02	4
k26	0.95	0.05	1	1.25	0.25	2	1.45	0.45	3.5	1.45	0.45	3.5
k28	1.27	0.27	2	0.94	0.06	1	1.66	0.66	3	3.42	2.42	4
k39	1.37	0.37	4	0.96	0.04	2	0.82	0.18	3	0.99	0.01	1
k2	1.09	0.09	4	1.02	0.02	2	1.07	0.07	3	1.00	0.00	1
k20	1.58	0.58	3	0.71	0.29	1	0.65	0.35	2	1.66	0.66	4
k34	1.18	0.18	2	1.08	0.08	1	0.75	0.25	4	1.19	0.19	3
RS			23			12			29.5			25.5

Nun wird auch für dieses Kriterium untersucht, wie sich die Methoden in einzelnen NOGAs bewähren. Dazu werden in Tabelle 15 die Steigungsparameter, berechnet aufgrund der Werte innerhalb einer NOGA2-Klasse, verglichen. Die Werte in der Tabelle ergeben sich aus

$$\hat{\beta}_{c,meth} = \frac{\sum_{i,k} y_{k_i} \cdot y_{k_i,meth} \cdot 1(k_i \in c)}{\sum_{i,k} y^2_{k_i,meth} \cdot 1(k_i \in c)}. \tag{16}$$

Die aufgelisteten Vergleiche beschränken sich auf NOGA2-Klassen mit $m_c \geq 10$. In sechs der neun betrachteten NOGAs erzielt Trend das beste Ergebnis, in den restlichen drei liegt die Methode auf Rang 2. Global bestätigt sich das Resultat, dass Trend am besten abschneidet.

7.4.2 Der mittlere quadratische Fehler

Dieser Abschnitt ist dem zweiten Kriterium, der Grösse der Terme ϵ_{k_i}, gewidmet. Hierzu werden in Tabelle 16 die Wurzeln der mittleren quadrierten Fehler (RMSE=Root Mean Square Error) dargestellt. Die mittleren quadrierten Fehler (MSE) berechnen sich aus

$$MSE^{(k)}_{meth} = \frac{1}{50} \sum_i (y_{k_i} - \hat{\beta}^{(k)}_{meth} \cdot y_{k_i,meth})^2, \tag{17}$$

wobei $\hat{\beta}^{(k)}_{meth}$ in Gleichung (15) definiert ist. Das Kriterium liefert eine klare Unterscheidung: $Trend$ auf Rang 1, $nnbr$ auf Rang 2, $nnbr_{s14}$ auf Rang 3 und $Mittel$ auf Rang 4.

Tabelle 16 Die Quadratwurzel des mittleren quadrierten Fehlers (RMSE) aus der Regression global (Rg=Rangierung der Methoden nach RMSE, RS=Rangsumme).

	Mittel		Trend		$nnbr_{s14}$		$nnbr$	
k=	$RMSE^{(k)}_{Mittel}$	Rg	$RMSE^{(k)}_{Trend}$	Rg	$RMSE^{(k)}_{nnbr_{s14}}$	Rg	$RMSE^{(k)}_{nnbr}$	Rg
1	250241	4	62128	1	177198	3	100955	2
2	2250488	4	206396	1	985512	3	290494	2
3	977821	4	165671	2	233261	3	101602	1
4	1234568	4	120392	1	1073982	3	751178	2
5	490636	4	59418	1	426188	3	126731	2
RS		20		6		15		9

Tabelle 17 Die Quadratwurzel des mittleren quadrierten Fehlers (RMSE) aus der Regression nach NOGA2 (Rg=Rangierung der Methoden nach RMSE, RS=Rangsumme).

c	$RMSE_{c,Mittel}$	Rg	$RMSE_{c,Trend}$	Rg	$RMSE_{c,nnbr_{s14}}$	Rg	$RMSE_{c,nnbr}$	Rg
k16	136180	4	31517	1	77066	3	45338	2
k29	914593	4	66376	1	613892	2	634621	3
k11	3766014	4	204078	1	1602040	3	608844	2
k26	204056	4	55298	1	191981	3	112593	2
k28	1714558	3	106641	1	1717004	4	960816	2
k39	295557	4	73493	1	228844	3	164596	2
k2	319776	4	22083	1	210330	3	97596	2
k20	674359	4	282138	1	671114	3	357891	2
k34	249060	3	27329	1	270386	4	47888	2
RS		34		9		28		19

Schliesslich wird auch in diesem Fall untersucht, wie sich die Methoden in einzelnen NOGAs bewähren. Dazu werden in Tabelle 17 die RMSE-Werte, berechnet nach NOGA2-Klasse, verglichen. Die mittleren quadrierten Fehler, deren Quadratwurzeln in der Tabelle aufgelistet sind, ergeben sich aus

$$MSE_{c,meth} = \frac{1}{m_c - 1} \sum_{i,k} (y_{k_i} - \hat{\beta}_{c,meth} \cdot y_{k_i,meth})^2 \cdot 1(k_i \in c). \tag{18}$$

Der Parameter $\hat{\beta}_{c,meth}$ ist in Gleichung (16) definiert. Trend liegt in sämtlichen betrachteten NOGAs auf Rang 1. Ebenso klar schneidet $nnbr$ auch auf Ebene NOGA2 am zweitbesten ab. Abbildung 3 zeigt als Illustration zu den in Tabelle 17 verwendeten Methoden Scatterplots mit den Residuen und den erwarteten Werten ($\hat{\beta}_{c,meth} \cdot y_{k_i,meth}$) für die Klasse k16. Auf den Grafiken ist gut ersichtlich, dass die Methoden $Trend$ und $nnbr$ kleinere Residuen ergeben als $nnbr_{s14}$ und $Mittel$.

Bemerkung: Analog zu den Tabellen 16 und 17 wurden auch die aus den Regressionen resultierenden R^2-Werte betrachtet. Für die Methode Trend lagen dabei die meisten Werte im Bereich $[0.95, 1]$.

7.5 Interpretation der Resultate

Die Resultate bestätigen, was bereits aufgrund der Extrapolationsresultate vermutet werden durfte:

Abbildung 3 Residuen versus erwartete Werte in Mio. für die NOGA2-Klasse k16.

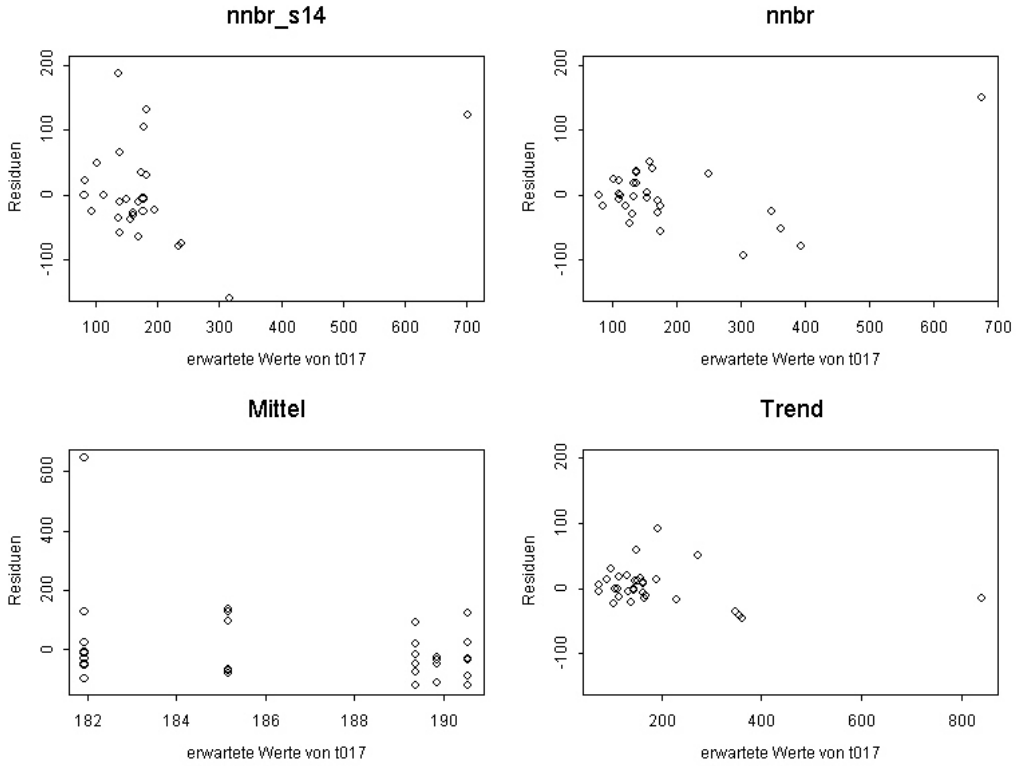


Tabelle 18 Die Rangsummen über sämtliche Kriterien.

noga2	Mittel	Trend	$nnbr_s14$	$nnbr$
k16	14	4	14	8
k29	13	4	13	10
k11	12	4	12	12
k26	10	6	13.5	10.5
k28	11	4	14	11
k39	16	6	12	6
k2	15	5	13	7
k20	14	6	10	10
k34	12	5	13	10

Die betrachteten Kriterien sprechen allesamt dafür, dass Trend auf globaler Ebene die besten Einsetzungen liefert. Über sämtliche Kriterien hinweg, scheint Trend auch in den betrachteten NOGA2-Klassen am besten abzuschneiden.

Hierzu sind in Tabelle 18 die Rangsummen nach NOGA2-Klassen dargestellt, welche sich ergeben, falls man die Ränge über die verschiedenen Kriterien hinweg addiert. Am zweitbesten schneidet sowohl global wie auch in den meisten NOGAs die Variante $nnbr$ ab.

Das gute Abschneiden der Variante Trend lässt sich durch die hohen Korrelationen zwischen den Bruttoproduktionswerten aus der WS04 und der WS03 erklären. In den Tabellen 19 und

Tabelle 19 Die Pearson-Korrelationen zwischen den betrachteten Variablen (gerundet auf zwei Nachkommastellen). Die für die Methode Trend interessanten Bereiche der Korrelationsmatrix sind fett hervorgehoben.

Variable	$t053_{ws02}$	$t053_{ws03}$	$t053_{ws04}$	$t022_{ws02}$	$t022_{ws03}$	$t022_{ws04}$	$t017_{ws02}$	$t017_{ws03}$	$t017_{ws04}$
$t053_{ws02}$	**1**	**1**	**0.99**	0.57	0.57	0.56	0.72	0.71	0.7
$t053_{ws03}$	**1**	**1**	**1**	0.58	0.59	0.58	0.73	0.73	0.72
$t053_{ws04}$	**0.99**	**1**	**1**	0.58	0.59	0.58	0.73	0.73	0.72
$t022_{ws02}$	0.57	0.58	0.58	**1**	**0.96**	**0.95**	0.97	0.93	0.92
$t022_{ws03}$	0.57	0.59	0.59	**0.96**	**1**	**0.99**	0.94	0.96	0.97
$t022_{ws04}$	0.56	0.58	0.58	**0.95**	**0.99**	**1**	0.94	0.96	0.97
$t017_{ws02}$	0.72	0.73	0.73	0.97	0.94	0.94	**1**	**0.97**	**0.97**
$t017_{ws03}$	0.71	0.73	0.73	0.93	0.96	0.96	**0.97**	**1**	**0.99**
$t017_{ws04}$	0.7	0.72	0.72	0.92	0.97	0.97	**0.97**	**0.99**	**1**

Tabelle 20 Die Spearman-Korrelationen zwischen den betrachteten Variablen (gerundet auf zwei Nachkommastellen). Die für die Methode Trend interessanten Bereiche der Korrelationsmatrix sind fett hervorgehoben.

Variable	$t053_{ws02}$	$t053_{ws03}$	$t053_{ws04}$	$t022_{ws02}$	$t022_{ws03}$	$t022_{ws04}$	$t017_{ws02}$	$t017_{ws03}$	$t017_{ws04}$
$t053_{ws02}$	**1**	**0.97**	**0.93**	0.63	0.64	0.62	0.76	0.76	0.73
$t053_{ws03}$	**0.97**	**1**	**0.98**	0.65	0.66	0.65	0.79	0.8	0.78
$t053_{ws04}$	**0.93**	**0.98**	**1**	0.64	0.65	0.65	0.77	0.79	0.79
$t022_{ws02}$	0.63	0.65	0.64	**1**	**0.98**	**0.96**	0.94	0.92	0.91
$t022_{ws03}$	0.64	0.66	0.65	**0.98**	**1**	**0.98**	0.93	0.94	0.93
$t022_{ws04}$	0.62	0.65	0.65	**0.96**	**0.98**	**1**	0.91	0.93	0.94
$t017_{ws02}$	0.76	0.79	0.77	0.94	0.93	0.91	**1**	**0.98**	**0.96**
$t017_{ws03}$	0.76	0.8	0.79	0.92	0.94	0.93	**0.98**	**1**	**0.98**
$t017_{ws04}$	0.73	0.78	0.79	0.91	0.93	0.94	**0.96**	**0.98**	**1**

20 sind hierzu die Korrelationen nach Pearson, respektive Spearman für die Variablen t017, t022 sowie t053 für die Jahre 2002-05 dargestellt. Würde eine Konstante (Intercept) die Qualität der Einsetzungen wesentlich verbessern? Betrachtungen entsprechender Regressionsmodelle liessen nicht darauf schliessen, dass der Einbezug einer Konstanten generell eine wesentliche Verbesserung bringen würde. Zudem ist zu beachten, dass in gewissen NOGAs nur wenige Daten vorhanden sind um die Regressionsparameter zu schätzen. Dieses Problem würde mit einem zusätzlichen Parameter verschärft.
Es zeigt sich, dass die Werte der Variablen t022 und t053 aus aufeinanderfolgenden Jahren ähnlich hohe Korrelationen aufweisen wie t017. Dies spricht für die Tatsache, dass die Variante Trend auch für diese Variablen gute Einsetzungen liefert.

8 Schlussfolgerungen

Die aktuellen Total- und Varianzschätzungen sind darauf ausgelegt, dass sämtliche sehr grossen Unternehmen antworten. In Realität gibt es sehr grosse Unternehmen, die ihre Daten nicht liefern (Grössenordnung 8 %). Die in den vorangehenden Abschnitten dokumentierten Untersuchungen zeigen, dass die Behandlung fehlender Werte sehr grosser Unternehmen die Resultate auf Ebene NOGA2 wesentlich beeinflussen kann. Zudem zeichnete sich ab, dass die Verwendung einer geeigneten expliziten Einsetzung die Schätzungen wohl tendenziell erhöhen würde. Für die Behandlung fehlender Werte bei sehr grossen Unternehmen ergeben sich aus

diesem Methodenbericht folgende Erkenntnisse:

1. Einsetzungen können eine Verbesserung der Qualität der Schätzungen bewirken. Von den in dieser Arbeit betrachteten Einsetzungsvarianten scheint $Trend$ am geeignetsten zu sein.
   Begründung: Die beobachtete tendenzielle Erhöhung der Schätzungen unter der Verwendung von Einsetzungsvarianten weist darauf hin, dass mit einer geeigneten Einsetzung der Bias verkleinert werden könnte. Dabei ist der Effekt der mittels 'last carry forward' eingesetzten Vollzeitstellenäquivalente (s_{14}) noch genauer zu untersuchen. Die Resultate der Variante $Trend$ erscheinen plausibel (ist bei $Mittel$ und $nnbr_{s14}$ teilweise nicht der Fall). Zudem schnitt $Trend$ bei der gemachten Simulationsstudie, in der eingesetzte mit echten Werten verglichen wurden, am besten ab.

2. Einflussreiche Einsetzungswerte sollten identifiziert (beispielsweise durch einen Vergleich der Resultate mit und ohne dem betreffenden Unternehmen) und einzeln plausibilisiert werden.
   Begründung: Es hat sich gezeigt, dass einzelne Einsetzungen einen gewichtigen Einfluss auf die Schätzungen haben können.

3. Aktuelle Werte der Hilfsvariablen s_{14} (Vollzeitstellenäquivalent) und falls möglich auch $noga2_{ws0t}$ sollten für alle angefragten Unternehmen ab einer gewissen Grösse erfasst werden.
   Begründung: Diese Variablen sind nicht konstant und falls ein alter Wert eingesetzt wird, handelt es sich im Grunde um eine Einsetzung („last carry forward"). Die durch die Einsetzung entstehende Variabilität ist schwierig zu schätzen. Zudem hängt vom Vollzeitstellenäquivalent zum Zeitpunkt der Erhebung ab, ob ein Unternehmen überhaupt als sehr gross betrachtet wird und andererseits wird dadurch das für den Quotientenschätzer verwendete Populationstotal des Vollzeitstellenäquivalentes beeinflusst.

8.1 Weitere Arbeitsschritte

Fragestellungen betreffend der Robustheit der Einsetzungsmethode sowie der Varianzschätzung wurden in diesem Bericht noch weitgehend ausgeklammert. Es ist geplant diese beiden Themengebiete aufbauend auf der Methode Trend zu behandeln. Konkret geht es dabei um folgende Probleme:

- In dieser Arbeit wurde der (Regressions-)Koeffizient der Methode Trend jeweils aufgrund der sehr grossen Unternehmen einer bestimmten NOGA2-Klasse geschätzt. Dies kann dazu führen, dass nur sehr wenige Unternehmen in die Schätzung des Koeffizienten einfliessen. Dieses Problem könnte man behandeln, in dem man beispielsweise auch kleinere Unternehmen für die Trendschätzung berücksichtigt. Dabei muss aber ein allfälliger Bias im Auge behalten werden.

- Der Koeffizient der Methode Trend entspricht dem Regressionskoeffizienten des Quotientenschätzers (vgl. beispielsweise [3], Abschnitt 7.3). Dieser kann durch einzelne Beobachtungen stark beeinflusst werden. In [5] wurden robuste Versionen des Regressionskoeffizenten vorgestellt, die im Hinblick auf die Einsetzungen attraktive Alternativen bieten können.

- Die Einsetzungen dürfen in der Varianzschätzung nicht wie echte Beobachtungen behandelt werden, da so die Varianz unterschätzt würde. Es wird nötig sein, die Varianzschätzungen an die Tatsache anzupassen, dass Werte eingesetzt wurden.

Literatur

[1] 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

[2] Däppen, M., Roth, V. (2005). *Buchhaltungsergebnisse schweizerischer Unternehmen. Geschäftsjahre 2002-2003.* Bestellnummer: 029-0300

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

[4] Groves, M., Dillman, D. A., Eltinger, J. L., Little, R. J. A. (2002). *Survey Nonresponse* Wiley Series in Probability and Statistics, John Wiley and Sons, Inc.

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

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodology reports published by the SFSO's Statistical Methods Unit

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

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

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

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

Diffusionsmittel	*Kontakt* *N° à composer*	*Moyen de diffusion*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch	Service de renseignements individuels
Das BFS im Internet	www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch→Aktuell→Publikationen.

La liste des publications, mise à jour régulièrement, donne davantage de détails sur les divers moyens de diffusion. Elle se trouve sur Internet à l'adresse www.statistique.admin.ch →Actualités→Publications.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodology Reports by the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde 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 sicher zu stellen, und sie sollen 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.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site internet de l'OFS.

Für die Jahre 2002–2004 wurden fehlende Werte grosser Unternehmen mit vier verschiedenen Einsetzungsmethoden behandelt und die daraus resultierenden Auswirkungen auf die Extrapolationstotale untersucht. Tendenziell erhöhten die Einsetzungen die Schätzungen, wobei die Unterschiede im Allgemeinen moderat ausfielen. In einem Vergleich der Einsetzungsmethoden anhand einer Simulationsstudie schnitt die Einsetzung mittels trendangepasster Vorjahreswerten am besten ab.

Bestellnummer: 338-0048

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

Preis: gratis

ISBN 978-3-303-00391-6