

Methodenbericht

Umweltschutzausgaben der Unternehmen 2003

Stichprobenplan, Datenaufbereitung und Schätzverfahren



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Neuchâtel, 2006

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

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

0	Statistische Grundlagen und Übersichten	0	Bases statistiques et produits généraux
1	Bevölkerung	1	Population
2	Raum und Umwelt	2	Espace et environnement
3	Arbeit und Erwerb	3	Vie active et rémunération du travail
4	Volkswirtschaft	4	Economie nationale
5	Preise	5	Prix
6	Industrie und Dienstleistungen	6	Industrie et services
7	Land- und Forstwirtschaft	7	Agriculture et sylviculture
8	Energie	8	Energie
9	Bau- und Wohnungswesen	9	Construction et logement
10	Tourismus	10	Tourisme
11	Verkehr und Nachrichtenwesen	11	Transports et communications
12	Geld, Banken, Versicherungen	12	Monnaie, banques, assurances
13	Soziale Sicherheit	13	Protection sociale
14	Gesundheit	14	Santé
15	Bildung und Wissenschaft	15	Education et science
16	Kultur, Informationsgesellschaft, Sport	16	Culture, société de l'information, sport
17	Politik	17	Politique
18	Öffentliche Verwaltung und Finanzen	18	Administration et finances publiques
19	Kriminalität und Strafrecht	19	Criminalité et droit pénal
20	Wirtschaftliche und soziale Situation der Bevölkerung	20	Situation économique et sociale de la population
21	Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene	21	Développement durable et disparités régionales et internationales

Umweltschutzausgaben der Unternehmen 2003

Stichprobenplan, Datenaufbereitung und Schätzverfahren

Autor

Beat Hulliger
Bundesamt für Statistik

Herausgeber

Bundesamt für Statistik

Vorwort

Die Erhebung über Umweltschutzausgaben 2003 wurde durch die Sektion Umwelt, Nachhaltigkeit, Landwirtschaft im Jahr 2004 durchgeführt. Es handelt sich um eine schriftliche Erhebung bei den Unternehmen der Privatwirtschaft. Der Dienst Statistische Methoden wurde beauftragt, den Stichprobenplan zu entwerfen, die Stichprobe zu ziehen, die statistische Datenaufbereitung durchzuführen und die Schätzverfahren zu entwickeln.

Dem Erhebungsleiter Jacques Roduit danke ich herzlich für die hervorragende Zusammenarbeit. Mein Dank geht auch an Dr. Daniel Assoulin vom Dienst Statistische Methoden für die aufmerksame Durchsicht des Berichts.

Zusammenfassung

Der Stichprobenplan für die Erhebung über Umweltschutzausgaben der Unternehmen 2003 wurde mit Hilfe der Daten der entsprechenden Piloterhebung aus dem Jahr 1993 erstellt. Daten für Kleinst-Unternehmen wurden mit einer Kontrollstichprobe erhoben. Bei der Aufbereitung der Daten wurden die Daten von Unternehmensgruppen konsolidiert, Inkonsistenzen kontrolliert und, so weit möglich, bereinigt und fehlende Werte eingesetzt. Die Schätzverfahren benützen eine spezielle Schicht der grössten Unternehmen und berücksichtigen die Antwortausfälle.

Schlüsselwörter

Methodenbericht; Kleinst-Unternehmen; Unternehmensgruppen; Kontrollen; Einsetzungen; Ausreisser; Genauigkeit.

Auskunft:	Beat Hulliger, Tel. 032 713 60 15 Beat.Hulliger@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:	338-0042
Preis:	gratis
Reihe:	Statistik der Schweiz
Fachbereich:	0 Statistische Grundlagen und übersichten
Originaltext:	Deutsch
Grafik/Layout:	BFS
Copyright:	BFS, Neuchâtel 2006 Abdruck – ausser für kommerzielle Nutzung – unter Angabe der Quelle gestattet.
ISBN:	3-303-00342-4

Inhaltsverzeichnis

1	Einführung	5
2	Stichprobenplan	5
2.1	Datenbasis und Annahmen	5
2.2	Grobe Abschätzung der Stichprobengrösse	6
2.3	Modellierung der Variabilität der USchA93	8
2.4	Konstanz der Variationskoeffizienten	9
2.5	Modellierung der Mittelwerte	10
2.6	Zuteilung der Bruttostichprobengrösse	11
2.7	Antwortausfälle und projektierte Nettostichprobe	14
3	Ziehung der Stichprobe	15
3.1	Stichprobenrahmen	16
3.2	Stichprobe	16
4	Erhebung, manuelle Datenaufbereitung und Antwortraten	17
4.1	Erhebung und manuelle Aufarbeitung	17
4.2	Antwortraten	18
5	Automatische Datenaufbereitung und Analyse	20
5.1	Datenübernahme	20
5.2	Strategie für die Aufbereitung	21
5.3	Bilanz-Kontrollen	23
5.4	Bilanz-Einsetzungen	24
5.5	Kontrolle und Einsetzung der monetären Variablen	25
5.6	Anzahl Beschäftigte	26
5.7	Unternehmensgruppen	27
5.8	Untersuchung Kilo-Fehler	29
5.9	Einsetzungen für Gebühren	30
6	Schätzverfahren	32
6.1	Schichtung und selbst-repräsentierende Unternehmen	33
6.2	Wahl des Schätzmodells	33
6.3	Untersuchungsbereiche	34

6.4	Ausreisser	35
6.5	Robustifizierung	39
6.6	Resultate und Genauigkeit	41

Anhang	44
---------------	-----------

A.1	Construction of sampling weights for EPE93	44
A.2	Wirtschaftsgruppen und Populationsgrößen	44
A.3	Programme	45

Literatur	55
------------------	-----------

1 Einführung

Das Bundesamt für Statistik führte im Jahr 2004 nach 10 Jahren Unterbruch wieder eine Erhebung über die Umweltschutzausgaben der Privatwirtschaft im Jahr 2003 durch (UWSA2003). Das Haupt-Ziel der Erhebung 2003 war, nach der Piloterhebung 1993 aktuelle und genauere Daten zu erheben. Ausserdem sollte die Erhebung als neuer Pilot für eine regelmässige Erhebung über Umweltschutzausgaben dienen, wie sie das bilaterale Abkommen über Statistik zwischen der Europäischen Union und der Schweiz vorsieht. Für die Entwicklung des Stichprobenplans wurden die Daten der Piloterhebung aus dem Jahr 1993 herangezogen (Kapitel 2). Dabei war der kleine Stichprobenumfang der Piloterhebung ($n = 1239$), die Datenlage (viele 0-Werte, Ausreisser, unklare Gewichtung, Wirtschaftsnomenklatur ASWZ85) und der lange Zeitraum seit der Piloterhebung zu berücksichtigen.

Die Daten der Piloterhebung über die Umweltausgaben der Privatwirtschaft 1993 (USchA93) wurden im Projekt EUREDIT verwendet und sind daher recht gut dokumentiert (Béguin and Hulliger 2003). Die im Rahmen von EUREDIT rekonstruierte Gewichtung der USchA93 wurde für die Entwicklung des Stichprobenplans übernommen (A.1).

Die manuelle Datenaufbereitung wurde durch die Sektion Umwelt, Nachhaltigkeit und Landwirtschaft durchgeführt. Diese beauftragte das Marktforschungsinstitut LINK mit telefonischen Rückrufen. Nach der Phase der manuellen Aufbereitung wurden die Daten dem Dienst Statistische Methoden übergeben und es folgte eine Phase der automatischen Datenaufbereitung (Kapitel 4). Es wurden automatische Kontrollen durchgeführt und in wichtigen Fällen Daten eingesetzt.

Die letzte Phase war die Entwicklung der Schätzverfahren (Kapitel 6). Dabei wurde auch der Einfluss von extremen Beobachtungen untersucht.

Die ersten offiziellen Ergebnisse wurden im Jahr 2005 veröffentlicht (BFS 2005), die detaillierten Resultate wurden Ende 2006 veröffentlicht (BFS 2006). Resultate in diesem Bericht dienen allein der Illustration der Methoden.

2 Stichprobenplan

2.1 Datenbasis und Annahmen

Die Daten über Umweltschutzausgaben, welche für die Entwicklung des Stichprobenplans zur Verfügung standen, stammen von der Erhebung über die Umweltschutzausgaben der Privatwirtschaft 1993 (USchA93). Es handelt sich um eine relativ kleine Piloterhebung (Nettostichprobengrösse 1239), die durch die Firma INFRAS im Auftrag des Bundesamts für Statistik durchgeführt wurde. Der Stichprobenplan, die Erhebung, die Schätzverfahren und die Ergebnisse sind im Bericht von INFRAS (INFRAS 1996) und in der Publikation des BFS (BFS 1996) dokumentiert. Überlegungen zur Robustifizierung der Schätzer finden sich auch im Methodenbericht (Hulliger and Kassab 1998).

In der Datenbasis der USchA93 mussten zwei Unternehmen, deren Schicht nicht bekannt war, ausgeschlossen werden. Es verbleiben $n = 1237$ Unternehmen. Da die wirtschaftliche Aktivität der Unternehmen der USchA93 noch nach der ASWZ85 codiert war, musste ein

grober Umrechnungsschlüssel auf der Ebene Zweisteller zur NOGA2 erstellt werden (siehe Anhang Tabelle A.2). Dabei wurde der Abbau und die Verarbeitung von Steinen und Erden (ASWZ85=33) gesamthaft der NOGA2=14 zugewiesen (In den späteren Kapitel wurde ein genauerer Schlüssel verwendet, vgl. Abschnitt 2.3). Die Unternehmen, die nach der Umschlüsselung innerhalb des Bereichs $10 \leq \text{NOGA2} \leq 74$ lagen, wurden für die Berechnungen hinzugezogen. Die Unternehmen mit weniger als 20 Beschäftigten wurden *nicht* ausgeschlossen. Es resultierten 1205 Unternehmen im Datensatz. Als interessierende Variablen wurden die totalen Investitionen ($\text{totinvto}=\text{F2FZ_4}$), die totalen laufenden Ausgaben ($\text{totexpto}=\text{F3FZ_4}$), die totalen Einnahmen ($\text{rectot}=\text{F4FZ_1}$) und die totalen Subventionen ($\text{subtot}=\text{F5FZ_1}$) verwendet.

Den Berechnungen liegen also Daten aus dem Jahr 1993 zugrunde. Wir nehmen an, dass die Variationskoeffizienten im Jahr 2003 in der selben Grössenordnung liegen. Der Stichprobenplan 1993 wurde von der Firma INFRAS entwickelt. Er war geschichtet nach Wirtschaftsklassen (ASWZ85) und Grösse der Unternehmen. Ausserdem wurde sowohl bei der Abgrenzung der Population wie auch bei der Schichtung die Intensität der USchA berücksichtigt (INFRAS 1996). Es ist also ein klassischer Stichprobenplan für Unternehmenserhebungen, der in ähnlicher Weise auch bei der UWSA03 Anwendung finden könnte. Die Daten wurden unbehandelt belassen, obwohl klare Inkonsistenzen und Ausreisser vorhanden sind. Während INFRAS für die Hochrechnung robustifizierte Quotientenschätzer (mit der Anzahl Beschäftigten im Quotienten) verwendete, beschränken wir uns für die Entwicklung des Stichprobenplans auf geschichtete Mittel. Die Schätzungen für die Totale und für deren Varianzen dürften mit geschichteten Mitteln höher ausfallen als bei INFRAS. Für die grobe Abschätzung der Variabilität bei einer grösseren Stichprobe wurde wie folgt vorgegangen: Sei die Varianz σ_a^2 eines geschichteten Mittels unter dem ursprünglichen Stichprobenplan ausdrückbar als Summe der Varianzen über die Schichten h :

$$\sigma_a^2 = \sum_h (1 - f_h) \sigma_h^2 / n_{ah}. \quad (1)$$

Dabei bezeichnet $f_h = n_{ah} / N_{ah}$ die Stichprobenrate in den Schichten. Wir nehmen nun an, dass wir die selbe Schichtung benützen und lediglich in den Schichten die Stichprobengrössen um einen Faktor n_b / n_a erhöhen. Dies dürfte nicht optimal sein und ist auch nicht unbedingt möglich, weil die Stichprobengrösse durch die Populationsgrösse begrenzt ist. Die Varianz des geschichteten Mittels wird damit neu geschätzt durch

$$\sigma_b^2 = \sum_h (1 - f_h) \sigma_h^2 / (n_b n_{ah} / n_a) = \sigma_a^2 n_a / n_b. \quad (2)$$

Es werden also die alten Endlichkeitskorrekturen benützt, wodurch die Varianz überschätzt wird. Für die Berechnung der Variationskoeffizienten wurde angenommen, dass die Mittelwerte (bzw. Totale) gleich bleiben wie 1993.

Für die Schätzung der Varianzen in den Schichten wurden die Gewichte übernommen, die im Rahmen des Projekts EUREDIT entwickelt wurden (Charlton 2004).

2.2 Grobe Abschätzung der Stichprobengrösse

Bei den folgenden Berechnungen gehen wir davon aus, dass eine globale Antwortrate von 60% erreicht werden kann. Bei einer Bruttostichprobengrösse von 7500 kann also mit 4500

Antworten gerechnet werden. Die Ergebnisse für die betrachteten vier Kernvariablen sind in Tabelle 1 für neue Stichprobengrößen $n = 3000$ und $n = 4500$ zusammengefasst. Dabei handelt es sich um Schätzungen für die gesamte Population. Die Populationsgröße ist 17096, die Stichprobengröße der USchA93 war $N = 1205$. Die Wirtschaftsaktivitäten umfassten die Wirtschaftsklassen NOGA2=10 bis 74 ausser NOGA2=37 und NOGA5=51.57A. Die Angaben tot-infras und cv-infras beziehen sich auf den Bericht von INFRAS über die USchA93 (INFRAS 1996). Die Variationskoeffizienten mit einfachen geschichteten Mitteln (Zeile cv) weichen beträchtlich von den von INFRAS berechneten Variationskoeffizienten für einen robusten Quotientenschätzer ab.

Wir versuchen, die Variationskoeffizienten anhand der Regeln für das statistische Jahrbuch zu beurteilen: $cv \leq 5\%$ ist in Ordnung, $5\% < cv \leq 15\%$ ist bedingt zuverlässig und wird nur noch in Klammern veröffentlicht, $cv > 15\%$ ist nicht zu veröffentlichen, bzw. es werden nur leere Klammern veröffentlicht (oder allenfalls die Ergebnisse in doppelten Klammern).

Tabelle 1 Variationskoeffizienten für die USchA93 und Abschätzungen für UWSA03.

	totinvto	totexpto	rectot	subtot
mean	107.26	88.55	2.94	7.99
var	137.55	346.63	4.74	9.76
cv	0.11	0.21	0.74	0.39
total	1'833'719	1'513'860	50'234	136'538
cvneu (n=3000)	0.07	0.13	0.47	0.25
cvneu (n=4500)	0.06	0.11	0.38	0.20
tot-infras	1'236'000	1'008'000	39'700	9'600
cv-infras	0.21	0.08	0.34	2.06

Variationskoeffizienten cvneu für UWSA03 mit Stichprobengröße netto 3'000 bzw. 4'500.

Bei einer Netto-Stichprobengröße von 4500 kann man die Schätzung des Totals der Investitionen (totinvto) und der Ausgaben (totexpto) noch als bedingt zuverlässig bezeichnen, während die totalen Erträge (rectot) und Subventionen (subtot) nicht mehr veröffentlicht werden sollten.

Bei einer Netto-Stichprobengröße von 3000 bleibt die Situation im Wesentlichen gleich, weil die Variationskoeffizienten nur leicht zunehmen.

Um die Genauigkeit auf der Ebene von Wirtschaftsgruppen grob abschätzen zu können wurde eine Vergrößerung der 17 von Eurostat vorgeschlagenen Wirtschaftsgruppen auf 7 Wirtschaftsgruppen untersucht (gruppe.bdp, siehe Tabelle 38 im Anhang). Dabei ist zu berücksichtigen, dass beim Übergang von der ASWZ85 zur NOGA einige Unschärfe dazukommt (Umschlüsselung siehe Tabelle A.2 im Anhang). Da die Tabellen von INFRAS nicht in dieser aggregierten Form vorliegen, haben wir keine Vergleichsmöglichkeit bei den Resultaten nach Klassen.

Wir betrachten die Variable Total der Investitionen (totinvto) (Tabelle 2). Die Schätzung des Mittelwerts (und damit des Totals) könnte bei $n = 4500$ für alle Klassen ausser 1 noch in Klammern veröffentlicht werden, sind also noch bedingt zuverlässig.

Bei $n = 3000$ könnte die Schätzung des Mittelwerts nur für die Klassen 2 (knapp), 4 und 5 in Klammern veröffentlicht werden. Bei den Resultaten nach groben (!) Wirtschaftsklassen ist

Tabelle 2 Variationskoeffizienten für die Variable Total der Investitionen (totinvto)

Kl.	Name	<i>N</i>	<i>n</i>	Mittel	SAM	cv	cv(3000)	cv(4500)
1	Extraction	398	179	402	152	0.38	0.24	0.20
2	Production	2'109	300	135	30	0.22	0.14	0.12
3	Chimie	506	141	1027	258	0.25	0.16	0.13
4	Machines	2'828	301	108	20	0.18	0.12	0.09
5	Energie	90	21	551	103	0.19	0.12	0.10
6	Constr.	4'483	97	19	5	0.26	0.16	0.13
7	Services	6'682	166	65	16	0.25	0.16	0.13

SAM: Standardabweichung des Mittelwerts

also ein Rückgang der Qualität merkbar.

Für beide Stichprobengrößen scheint es möglich, mit einem verbesserten Stichprobenplan und besseren Schätzern die Genauigkeit verbessern zu können (siehe auch (Hulliger and Kassab 1998)).

Die Ergebnisse für die Schätzungen des Totals der Ausgaben, der Erträge und der Subventionen sind weniger gut als bei den Investitionen. In Anbetracht der bereits hohen Aggregation in 7 Wirtschaftsklassen wurde empfohlen, die Brutto-Stichprobengröße bei 7500 zu belassen.

Um eine grobe Idee der Verhältnisse der Varianzen in den Grössenklassen zu erhalten, wurden die Variationskoeffizienten in Tabelle 3 berechnet. Obwohl die Schicht der grössten Unternehmen die höchste Standardabweichung des Mittelwerts aufweist, wird der Variationskoeffizient durch den sehr hohen Mittelwert unter den Wert der Schicht der zweitgrössten Unternehmen gedrückt. Es ist klar, dass die grösseren Unternehmen den Hauptanteil am Total und an der Variabilität beisteuern.

Tabelle 3 Variationskoeffizienten für die Schätzung des Mittels von totinvto für die 5 Grössenklassen.

gkl5	anzbe	<i>N</i>	<i>n</i>	Mittel	SAM	cv
1	0-19	917	74	17.5	11.3	0.65
2	20-49	9999	316	28.3	9.8	0.35
3	50-99	3231	266	53.4	11.2	0.21
4	100-249	1855	335	207.5	113.1	0.55
5	250+	1094	214	893.0	305.7	0.34

SAM: Standardabweichung des Mittelwerts

2.3 Modellierung der Variabilität der USchA93

Für die weitere Untersuchung der Variabilität der USchA93 wurde nur noch die Variable totinvto verwendet. Die Varianzen bzw. Variationskoeffizienten sollen mit Hilfe der Wirtschaftsklassen und der Grössenklassen erklärt werden. Um die Resultate besser für den Stichprobenplan verwenden zu können, wurde eine feinere Umschlüsselung auf der untersten Ebene ASWZ85

und NOGA95 vorgenommen (Umschlüsselung nach der NOGA mit dem grössten Gewicht, cf. Datei Wart85-Noga95-Gewichtet1.xls).

Der Ansatz, der im Folgenden verfolgt wird, basiert auf einer optimalen Zuteilung der Bruttostichprobe nach Neyman-Pearson. Die Brutto-Stichprobengrösse $n = 7500$ wird als bekannt und fixiert angenommen.

Die Neyman-Tschuprow Zuteilung ist dann:

$$n_{gk} = n \frac{N_{gk} S_{gk}}{\sum_{g',k'} N_{g'k'} S_{g'k'}}, \quad (3)$$

wobei sich der Doppel-Index gk auf die Wirtschaftsgruppe g und die Grössenklasse k bezieht.

Die Grössen der Schichten, N_{gk} sind bekannt aus dem BUR (Stichprobenrahmen). Unbekannt sind die Standardabweichungen S_{gk} in den Schichten. Es kann immer nur bezüglich einer Variable, bzw. mit Hilfe der Standardabweichung einer bestimmten Variablen optimiert werden. Ein klassischer Ansatz ist, anstatt der Schätzung der Standardabweichung die Schätzung des Schicht-Mittelwerts $\bar{x}_{gk}$ heranzuziehen und zu postulieren, dass $S_{gk} = c\bar{x}_{gk}$ für eine Konstante c . Mit anderen Worten nehmen wir an, dass der Variationskoeffizient in den Schichten konstant ist. Setzt man die entsprechende Schätzung in die Formel (3) ein, ergibt sich

$$n_{gk} = n \frac{N_{gk} c \bar{x}_{gk}}{\sum_{g',k'} N_{g'k'} c \bar{x}_{g'k'}} = n \frac{x_{gk+}}{\sum_{g',k'} x_{g'k'+}}, \quad (4)$$

wobei x_{gk+} das Total der Variablen in der Schicht ist, also $x_{gk+} = N_{gk} \bar{x}_{gk}$. Der Variationskoeffizient verschwindet also aus der Gleichung. Entsprechend kann Optimalität für mehrere Variablen postuliert werden, wenn deren jeweilige Variationskoeffizienten über die Schichten konstant sind.

Das spezielle Problem bei der UWSA03 ist, dass wir die Variationskoeffizienten in den beabsichtigten Schichten aus der USchA93 schätzen müssen, also auf einer alten und kleinen Datenbasis, bei denen die Schichten auch noch mit einer anderen Wirtschaftsamenklatur gebildet worden sind.

Wir gehen so vor, dass wir zuerst die Hypothese der Konstanz der Variationskoeffizienten für die Variable `totinvto` überprüfen.

2.4 Konstanz der Variationskoeffizienten

Um eine genügend gute Schätzung des Variationskoeffizienten zu erhalten, wurde mit der Vergrößerung der Wirtschaftsklassen auf 7 Klassen gearbeitet (`gruppe.bdp`, siehe Anhang A.2).

Damit ergibt sich die Besetzung der Zellen in Tabelle 4. Die Grössenklasse 1 wird nun in eine eigene Schicht versetzt. Für den Rest der Gruppen wird ein Variationskoeffizient berechnet, wie wenn es sich um eine Schicht handeln würde. Die zusätzliche Schichtung innerhalb dieser Zellen wurde also vernachlässigt. Daraus dürfte sich eine konservative Schätzung der Varianz ergeben.

Für die folgende Varianzanalyse wurde die Grössenklasse 1 ausgeschlossen, weil die Zellenbesetzung zu klein war. Als Arbeitshypothese nehmen wir an, dass die Variationskoeffizienten

in den Zellen für die Grössenklassen 2 bis 5 gleich sind. Eine Varianzanalyse der Variationskoeffizienten mit den Faktoren `gruppe.bdp = wgr` und `gkl5 = gkl` soll diese Hypothese untersuchen. Wir nehmen also an, dass die Quadrate der Variationskoeffizienten c verteilt sind wie χ^2/n . Die Wilson-Hilferty Transformation schlägt zusätzlich eine Kubikwurzel vor, um eine bessere Annäherung an normal verteilte Werte zu erreichen. Wir untersuchen also die Werte $c^{2/3}$ mit einer Varianzanalyse.

Obwohl individuelle F-Tests signifikante Unterschiede zwischen den Variationskoeffizienten einzelner Gruppen zeigen, ist weder `wgr` noch `gkl` signifikant. Mit anderen Worten hängt der Variationskoeffizient nicht von diesen zwei Variablen ab. D.h. entweder wir nehmen an, dass die Variationskoeffizient alle gleich sind, oder wir übernehmen die Variationskoeffizienten in den Gruppen direkt für den Stichprobenplan (entsprechend der Vermutung, dass die Wechselwirkungen `gkl : wgr` signifikant ist und entsprechend den signifikanten Unterschieden zwischen einzelnen Variationskoeffizienten).

Wir können die Signifikanz der Wechselwirkungen nicht überprüfen und arbeiten jetzt mit der Annahme, dass die Variationskoeffizienten der UScha93 in allen Schichten gleich waren. Damit können die Standardabweichungen in den Schichten als proportional zu den Mittelwerten der Schichten angenommen werden.

2.5 Modellierung der Mittelwerte

Wir analysieren nun in einem zweiten Schritt, ob die Schichten sich betreffend ihrer Mittelwerte unterscheiden. Hier steigen wir um auf leicht robustifizierte Mittelwerte (beidseitig winsorisiert um je 1%), weil einzelne Ausreisser hier einen grossen Einfluss haben können. Die Standardabweichungen im vorherigen Abschnitt waren natürlich auch beeinflusst, aber dieser Einfluss wurde dank der Division durch den ebenfalls nicht robusten Mittelwert gedämpft. Ausserdem ist die Berechnung von winsorisierten Mittelwerten natürlich einfacher als diejenige von winsorisierten Standardabweichungen.

In Tabelle 6 werden die gewichteten Mittel (`wmean`) und die gewichteten winsorisierten Mittel (`wwmean`) in den Schichten für die 17 Gruppen (`gruppe.ac`) verglichen.

Die winsorisierten Mittel zeigen eine klare Abhängigkeit von der Grössenklasse (p-Wert des

Tabelle 4 Zellenbesetzung bei Kreuzung Wirtschaftsklassen (`gruppe.bdp`) mit Grössenklassen (`gkl5`).

gruppe.bdp	Grössenklasse				
	1	2	3	4	5
1	6	18	12	11	9
2	20	71	48	75	42
3	0	27	45	52	25
4	29	116	115	132	79
6	18	29	11	28	13
7	1	46	29	32	22
8	0	9	6	5	24

Tabelle 5 Varianzanalyse von $cv^{2/3}$: Modell wgr+gkl.

Faktor	DF	SS	MS	F	p
wkl	6	9.74220	1.623701	1.652048	0.1903975
gkl	3	4.04401	1.348002	1.371536	0.2833685
Residuals	18	17.69114	0.982841		

Faktoren, Freiheitsgrade (DF), Quadratsummen (SS) und mittlere Quadratsummen (MS), F-Werte und p-Werte für die Varianzanalysen

F-Tests=0.004), während die Wirtschaftsklasse nicht signifikant zu sein scheint (p-Wert des F-Tests=0.112). Dies ist nun aber im Widerspruch zu den Erkenntnissen aus der genaueren Analyse der Umweltschutzausgaben, die z.B. in der Chemie (Gruppe 3) klar grössere Ausgaben pro Unternehmen feststellt als in den anderen Branchen.

Wir versuchen nun, die winsorisierten Mittelwerte selbst, mit einem einfachen ANOVA Modell zu erklären (siehe Tabelle 7). Dabei gehen wir wieder auf die 7 Wirtschaftsgruppen (gruppe.bdp) zurück. Die Effekte für das Modell mit den beiden Faktoren wgr und gkl zeigen, dass die Wirtschaftsgruppen 1 und 3 wesentlich höhere Mittelwerte ausweisen als die anderen.

Offensichtlich sind auch die Wechselwirkungen nicht vernachlässigbar. Jedenfalls ist eine klare Struktur zu erkennen: Der Anstieg mit der Grössenklasse ist für die Chemie (wgr=3) noch wesentlich stärker als für die anderen Wirtschaftsgruppen.

Wir betrachten nun je innerhalb der Wirtschaftsgruppen (gruppe.bdp) die Abhängigkeit zwischen Anzahl Beschäftigte (anzbe) und Gesamtinvestitionen (totinvto). Die Korrelation ist relativ hoch: $\rho = 0.63$. Die Anzahl Beschäftigte kann also als Annäherung für die Gesamtinvestitionen benützt werden.

2.6 Zuteilung der Bruttostichprobengrösse

2.6.1 Zuteilung auf Wirtschaftsgruppen

Für die Zuteilung der Bruttostichprobengrösse von $n = 7500$ auf die Wirtschaftsgruppen werden zuerst diese neu definiert. Der Vorschlag der Sektion UNL im Anhang (gruppe.ac) wird dahingehend modifiziert, dass die Wirtschaftsgruppe 6 "Cokéfaction" zur Wirtschaftsgruppe 7 "Chimie" geschlagen wird, weil sie sehr klein ist und sowieso aus Datenschutzgründen keine Ergebnisse für "Cokéfaction" veröffentlicht werden könnten.

In einem weiteren Schritt werden die Unternehmen mit weniger als 20 Beschäftigten des zweiten und dritten Sektors je in die speziellen Gruppen 20 und 21 zusammengefasst: Gruppe 20 besteht also aus den Unternehmen mit $betot < 20$ und $gruppe.ac \in \{1, \dots, 12\}$ und Gruppe 21 besteht aus den Unternehmen mit $betot < 20$ und $gruppe.ac \in \{13, \dots, 17\}$. Die Variable *grind* ist der Gruppen-Indikator für diese Wirtschaftsgruppen. Der Indikator *grind* nimmt also die Werte 1, ..., 5, 7, ..., 12, 13, ..., 17, 20, 21 an.

Für diese Gruppen werden nun winsorisierte gewichtete Mittelwerte berechnet. Dabei wurden neu die totalen Ausgaben (also Investitionen und laufende Ausgaben zusammen: $totaus = totinvto + totexpcto$) pro Unternehmen als Variable verwendet, damit noch mehr In-

Tabelle 6 Mittelwerte und winsorisierte Mittelwerte (beidseitig 1%) in den Schichten

Schicht	$\hat{N}_h$	n_h	wmean	VK	wwmean
1	917.21	74	17.50	7.95	8.10
12	68.09	18	230.67	3.13	74.33
13	50.26	12	325.01	1.04	297.42
14	33.40	11	348.83	1.43	277.06
15	24.79	9	1115.59	1.19	934.04
22	1025.72	71	14.39	2.43	12.31
23	352.27	48	52.02	2.12	48.34
24	261.19	75	190.90	2.69	170.70
25	142.67	42	1021.68	4.35	633.13
32	241.50	27	88.52	2.75	63.08
33	138.50	45	116.41	2.35	92.67
34	101.01	52	2217.06	5.62	624.03
35	60.82	25	4708.36	3.53	2503.57
42	1634.38	116	65.33	13.73	17.51
43	767.09	115	137.43	3.62	110.82
44	518.33	132	103.50	5.10	79.88
45	255.87	79	743.16	2.61	736.60
62	2550.56	29	10.80	2.74	7.03
63	781.00	11	9.55	2.00	6.55
64	369.79	28	94.20	2.06	73.75
65	95.12	13	131.55	2.02	45.61
72	3489.75	46	1.57	8.12	1.04
73	880.00	29	4.84	4.43	3.59
74	438.15	32	25.16	2.64	25.00
75	316.33	22	847.46	1.50	726.82
82	988.56	9	92.70	2.35	17.62
83	261.80	6	18.11	1.94	0.00
84	133.61	5	0.87	9.87	0.37
85	198.12	24	232.92	5.88	89.34

wmean: gewichtetes Mittel, VK: Variationskoeffizient des gewichteten Mittels,
 wwmean: winsorisiertes gewichtetes Mittel

formation herangezogen werden konnte. Die Unterschiede zwischen totaus und totinvto sind allerdings klein. Winsorisiert wurde dafür ein bisschen strenger, nämlich je um 5% unten und oben. Da die Besetzung im dritten Sektor relativ schwach war, wurden die Mittelwerte hier auf der Ebene der gruppe.d Gruppierung geschätzt. Mit anderen Worten wurde je für gruppe.ac $\in \{13, 14, 15\}$ und für gruppe.ac $\in \{16, 17\}$ der gemeinsame Mittelwert (winsorisiert und gewichtet) eingesetzt. Sonst hätte die Hotellerie z.B. einen Mittelwert von 0 erhalten.

Den speziellen Gruppen 20 und 21 wurden ohne Rücksicht auf die Variabilität je eine Brutstichprobengrösse von $ns_g=250$ zugeteilt. Dies sollte auch bei einer Antwortquote von 40% noch je etwa 100 Unternehmen in der Netto-Stichprobe ergeben. Bei einer Verteilung der Stichprobe auf den Rest der Wirtschaftsgruppen strikt proportional zum Total der Gesamt-

Tabelle 7 Modellierung winsorisierte Mittel

Tables of effects

wgr							
	1	2	3	4	6	7	8
	121.70	-57.89	546.83	-37.80	-240.77	-84.89	-247.18
gkl							
	2	3	4	5			
	-246.45	-194.09	-95.32	535.86			
wgr:gkl							
Dim 1 : wgr							
Dim 2 : gkl							
	2	3	4	5			
1	-74.9	95.8	-23.3	2.5			
2	42.6	26.3	49.9	-118.9			
3	-511.3	-534.1	-101.5	1146.9			
4	27.8	68.7	-61.0	-35.5			
6	220.2	167.4	135.8	-523.5			
7	58.4	8.6	-68.8	1.8			
8	237.2	167.3	68.9	-473.4			

ausgaben ergibt sich für den Sektor III eine Stichprobe der Grösse 2374. Dies wurde als zu grosse Stichprobe betrachtet, weil die Erhebung prioritär im zweiten Sektor stattfinden sollte. Andererseits wären nur 320 Unternehmen in der Baubranche (NOGA2=45 bzw. grind=12) in die Brutto-Stichprobe gelangt. Dies wurde vor allem auch im Hinblick auf die veränderte Umweltschutzgesetzgebung beim Bau als zu klein betrachtet. Die Schätzung des Totals der Gesamtausgaben wurde darum ad-hoc für die Baubranche verdoppelt und für den Sektor III halbiert.

Damit ergab sich die Aufteilung der Stichprobe auf die Wirtschaftsgruppen in Tabelle 8. Wenn dabei unter strikter Proportionalität in einer Wirtschaftsgruppe $n_{sg} > n_p$ wurde, dann wurde $n_{sg} = n_p$ gesetzt und der Rest der Stichprobe wiederum proportional auf die Wirtschaftsgruppen mit $n_{sg} < n_p$ verteilt. Der Sektor III erhält immer noch eine Stichprobe der Grösse 2119, in der Baubranche werden neu 866 Unternehmen befragt. Die Populationsgrössen n_p und die totale Anzahl Beschäftigte $betot$ stammen dabei aus dem Auszug aus dem BUR vom 22. Februar 2004, das bereinigt wurde (Vorgehen siehe Abschnitt 3).

2.6.2 Zuteilung auf Grössenklassen

Innerhalb jeder Wirtschaftsgruppe (ausser 20 und 21) werden die Unternehmen entsprechend den in Tabelle 3 aufgeführten Klassen und aufgrund der Variable $betot$ aus dem BUR (Anzahl Beschäftigte insgesamt) in 5 Grössenklassen eingeteilt.

Die Schichtgrössen sind in Tabelle 9 aufgeführt. Die Stichprobe der Wirtschaftsgruppe wurde proportional zur Anzahl Beschäftigte $betot$ in einer Grössenklasse auf die Grössenklassen verteilt. Eine erste Verteilung der Stichprobengrösse der Wirtschaftsgruppen auf die Grössenklassen zeigt, dass im Sektor II nur für die Wirtschaftsgruppen 10 und 12 die Grössenklasse

Tabelle 8 Wirtschaftsgruppen: Grösse und Zuteilungskriterium

grind	ns	np	betot	wwmean	total1	total2	nsg
1	29	65	2979	227	14749	14749	65
2	42	413	55843	217	89722	89722	413
3	34	221	17286	186	41098	41098	221
4	51	302	12278	37	11219	11219	100
5	109	557	50282	125	69693	69693	380
7	149	558	84023	561	313266	313266	558
8	121	182	13651	398	72418	72418	182
9	17	94	15232	878	82544	82544	94
10	304	2985	310784	105	314253	314253	1714
11	21	177	22504	677	119895	119895	177
12	81	2762	150620	29	79364	158727	866
13	93	3428	362718	87	298585	149293	814
14	6	1679	98813	87	146244	73122	399
15	30	1031	215274	87	89802	44901	245
16	22	680	176341	80	54177	27089	148
17	22	2363	196418	80	188266	94133	513
20	73	68869	285334	7	452222	452222	250
21	1	189167	667417	0	0	0	250

grind: Gruppenindikator, ns: Stichprobengrösse in USchA93, np: Populationsgrösse N_g im BUR (22.2.04), betot: Total der Beschäftigten aus dem BUR, wwmean: Winsorisiertes Mittel aus USchA93 $\hat{x}_g$, total1: Schätzung für das Total $\hat{x}_{g+} = N_g \hat{x}_g$, total2: Geänderte Schätzung für das Total (siehe Text), nsg: Stichprobengrösse in der Wirtschaftsgruppe g.

4 nicht vollständig erhoben wird (Stichprobenraten 0.765 bzw. 0.444). Um einen einheitlichen Stichprobenplan zu erhalten, wurde hier die Stichprobenrate ad-hoc auf 1 gesetzt. Ausserdem wurde die Stichprobenrate für Wirtschaftsgruppe 1 und Grössenklasse 2 von 0.980 ebenfalls auf 1 gesetzt. Damit ergibt sich die Stichproben-Zuteilung in Tabelle 10.

2.7 Antwortausfälle und projektierte Nettostichprobe

Die Antwortquote in der Piloterhebung im Jahr 2002 war global 0.636. Die Antwortquoten nach Beschäftigtenklasse (definiert nur mit Vollzeitstellen) sind in der ersten Zeile der Tabelle 11 aufgeführt. Wendet man die Antwortquoten in der zweiten Zeile von Tabelle 11 auf die Bruttostichprobengrösse an, dann ergibt sich eine Nettostichprobe der Grösse 4693, welche sich wie in Tabelle 12 auf die Schichten verteilt. Es zeigt sich dabei, dass die Nettostichprobengrösse einiger Schichten bereits an der unteren Grenze ist.

Bei unserem Vorgehen, zuerst die Bruttostichprobe möglichst optimal zu verteilen und dann auf die Nettostichprobe zurückzuschliessen, verlieren wir natürlich wegen den unterschiedlichen Antwortquoten an Optimalität. Andererseits erlaubt die Zuteilung proportional zur Anzahl Beschäftigte innerhalb der Wirtschaftsgruppen eine relativ grosse Brutto-Stichprobe bei den kleinen Unternehmen, was die etwas grösseren Ausfälle bei den kleinen Unternehmen teil-

Tabelle 9 Populationsgrössen in den Schichten

grind	0-19	20-49	50-99	100-249	250+
1	0	51	11	2	1
2	0	207	87	74	45
3	0	124	51	37	9
4	0	243	47	9	3
5	0	322	113	79	43
7	0	252	139	102	65
8	0	106	41	26	9
9	0	33	19	23	19
10	0	1673	651	442	219
11	0	91	41	26	19
12	0	2011	484	213	54
13	0	2329	621	317	161
14	0	1253	273	112	41
15	0	638	201	111	81
16	0	383	108	76	113
17	0	1583	415	251	114
20	68869	0	0	0	0
21	189167	0	0	0	0

weise kompensiert. Im Weiteren sind die Antwortquoten nur aufgrund einer kleinen Pilotstudie geschätzt worden. Die Bruttostichprobe sollte also genügend robust sein, um auch bei einer leichten Abweichung von der Optimalität noch gute Schätzungen zu ermöglichen.

Die Stichprobengewichte, die sich mit dieser erwarteten Nettostichprobe ergeben, variieren von 1 bis 16.8 für die Unternehmen ab 20 Beschäftigte. Dies sind Grössenordnungen, die keine Probleme bereiten sollten bei der Hochrechnung. Für die Schichten 20 und 21 betragen Stichprobengewichte 688.7 und 1891.7. Diese zwei Gewichte sind weit von den anderen entfernt. Die Stichprobe der Unternehmen mit unter 20 Beschäftigten kann also nicht ohne grosse Vorsicht mit der Hauptstichprobe kombiniert werden. Selbstverständlich sollten verbesserte Gewichtungen mit Quotientenschätzern und robusten Schätzern bei der Auswertung geprüft werden.

3 Ziehung der Stichprobe

Für die Ziehung der Stichprobe wurden ein Schicht-Indikator gebildet: $\text{stratind} = \text{grind} * 10 + \text{gk15}$. Die Stichprobenraten des Stichprobenplans wurden je mit dem Schicht-Indikator versehen aus S-Plus exportiert und in SAS eingelesen.

Tabelle 10 Stichprobengrösse in den Schichten

grind	0-19	20-49	50-99	100-249	250+
1	0	51	11	2	1
2	0	207	87	74	45
3	0	124	51	37	9
4	0	56	47	9	3
5	0	144	113	79	43
7	0	252	139	102	65
8	0	106	41	26	9
9	0	33	19	23	19
10	0	555	498	442	219
11	0	91	41	26	19
12	0	390	215	213	54
13	0	283	174	197	161
14	0	181	94	83	41
15	0	63	50	55	81
16	0	50	50	50	113
17	0	162	101	136	114
20	250	0	0	0	0
21	250	0	0	0	0

Tabelle 11 Antwortquoten in der Pilotstudie und Annahmen

	0-19	20-49	50-99	100-249	250+
Pilot (Vollzeitstellen)	-	0.559	0.596	0.603	0.812
Annahme (Beschäftigte)	0.40	0.60	0.60	0.65	0.80

3.1 Stichprobenrahmen

Die Datenbasis für den definitiven Stichprobenrahmen ist ein Auszug aus dem BUR vom 22. März 2004. Die Unternehmen mit Status 1 ("aktiv") und Betriebstyp E13 ("aktives Unternehmen des privaten Sektors") oder E27 ("öffentliche Betriebe") wurden ausgewählt. Die Wirtschaftsklassen NOGA2=37 und NOGA5=5157A wurden eliminiert. Die Struktur der Population ist in Tabelle 39 im Anhang beschrieben.

Die Struktur entsprechend dem Stichprobenplan, d.h. die Schichtgrössen, ist in Tabelle 13 aufgeführt.

3.2 Stichprobe

Die Stichprobe wurde in eine Excel-Datei exportiert. Mit Hilfe der Unternehmens-Identifikators `entid` wurde aus dem BUR-Auszug `week_loc040322` eine Liste von Arbeitsstätten, die je zu den gezogenen Unternehmen gehören, erstellt und in eine separate Datei abgespeichert. Dabei wurden die Arbeitsstätten mit Status 1 ("aktiv") oder 4 ("unbekannt") berücksichtigt.

Tabelle 12 Erwartete Nettostichprobe

grind	0-19	20-49	50-99	100-249	250+
1	0	31	7	1	1
2	0	124	52	48	36
3	0	74	31	24	7
4	0	34	28	6	2
5	0	86	68	51	34
7	0	151	83	66	52
8	0	64	25	17	7
9	0	20	11	15	15
10	0	333	299	287	175
11	0	55	25	17	15
12	0	234	129	138	43
13	0	170	104	128	129
14	0	109	56	54	33
15	0	38	30	36	65
16	0	30	30	33	90
17	0	97	61	88	91
20	100	0	0	0	0
21	100	0	0	0	0

Auf den an die Unternehmen versandten Listen sollten aber nur die aktiven Arbeitsstätten aufgeführt werden.

Die Stichprobe (Unternehmen und Arbeitsstätten) wurde am 30. März 2004 an die Sektion UNL geliefert.

4 Erhebung, manuelle Datenaufbereitung und Antwortraten

4.1 Erhebung und manuelle Aufarbeitung

Die Erhebung wurde mit einem schriftlichen Fragebogen (Abbildungen 7 und 8 im Anhang, Seite 53) bei den Unternehmen der Bruttostichprobe durchgeführt. Zwei schriftlichen Mahnung wurden an die Unternehmen gesandt.

Die Fragebogen wurden durch das Bundesamt für Informatik und Telekommunikation eingescannt und automatisch erfasst. Die Sektion UNL speicherte die Daten auf einer Access-Datenbank. Die Bilder der Fragebogen wurden mit der Datenbank verbunden, um die Datenerfassung überprüfen zu können. Nach einer ersten Kontrolle mit Hilfe von Access-Programmen wurden die Fragebogen wo nötig automatisch und manuell korrigiert. Manuelle Korrekturen wurden einzeln dokumentiert. In Zweifelsfällen wurden die Fragebeogen an das Marktforschungsinstitut LINK weitergeleitet. Die Unternehmen wurden telefonisch kontaktiert, um Inkonsistenzen zu bereinigen und fehlende Werten zu überprüfen bzw. zu vervollständigen. Die manuelle Aufbereitung der Daten war am 9. September 2004 abgeschlossen. Es mussten

Tabelle 13 Schichtgrössen N_{gk} bei der Ziehung

	gkl5					
grind	1	2	3	4	5	Total
1		52	11	2	1	318
2		206	87	74	44	2588
3		123	50	37	9	1838
4		243	47	9	3	6573
5		320	113	77	43	4591
7		249	139	102	64	1748
8		106	41	26	9	1387
9		33	19	22	19	258
10		1667	649	443	216	21104
11		90	41	26	19	402
12		2004	484	214	53	35857
13		2317	620	315	158	74359
14		1251	267	113	40	26504
15		635	202	112	82	11578
16		383	107	77	113	5534
17		1573	416	249	113	78323
20	68378					
21	187155					
Total	255533	11252	3293	1898	986	272962

aber auch später noch einzelne Daten auf der Datenbank abgeändert werden.

Die vermuteten Einheits-Fehler (Angaben in Franken anstatt in tausend Franken, sogenannter Kilo-Fehler) wurden markiert, aber nicht korrigiert (Flagge Ufausses2).

Unternehmensgruppen wurden nach den Angaben der Unternehmen gebildet und dokumentiert.

Der Eingang der Antworten der wichtigsten Unternehmen wurde manuell überprüft.

Nach der Bereinigung der Daten, den Rückfragen und der Kontrolle der Gruppen wurden die Daten auf der Datenbank auf dem Stand vom 9. Dezember 2004 eingefroren. Die Datenbank enthält einen Zustand der Daten nach dem Einscannen, nach den Test und nach den manuellen Bereinigungen und im Endzustand. Alle nachträglichen Änderungen wurden vollständig mit den in diesem Bericht aufgeführten Programmen durchgeführt.

4.2 Antwortraten

Betrachtet man die ursprüngliche Bruttostichprobe der Grösse $n_b = 7505$ und den Rücklauf an Fragebogen $n_n = 4434$, ergibt sich eine globale Antwortrate von $r_1 = 4434/7505 = 0.5908$. Diese Antwortrate berücksichtigt allerdings nicht, dass 38 Unternehmensgruppen gebildet wurden. In diesen Unternehmensgruppen sind 162 Unternehmen aus der Bruttostichprobe

Tabelle 14 Stichprobengrössen n_{gk} bei der Ziehung

	gkl5					
grind	1	2	3	4	5	Total
1	.	52	11	2	1	66
2	.	206	87	74	44	411
3	.	123	50	37	9	219
4	.	56	47	9	3	115
5	.	144	113	77	43	377
7	.	249	139	102	64	554
8	.	106	41	26	9	182
9	.	33	19	22	19	93
10	.	554	497	443	216	1710
11	.	90	41	26	19	176
12	.	389	215	214	53	871
13	.	282	174	196	158	810
14	.	181	92	84	40	397
15	.	63	51	56	82	252
16	.	50	50	51	113	264
17	.	161	102	135	113	511
20	249	.	.	.	.	249
21	248	.	.	.	.	248
Total	497	2739	1729	1554	986	7505

vertreten. Die bereinigte Nettostichprobe enthält 4408 Unternehmen bzw. Unternehmensgruppen. Wir können also eine Antwortrate $r_2 = 4408 / (7505 - 162 + 38) = 0.5972$ berechnen. Im Vergleich mit anderen Unternehmenserhebungen des BFS und in Anbetracht des schwierigen Fragebogens sind diese Antwortraten befriedigend.

Um eine Antwortrate pro Schicht bzw. Wirtschaftsgruppe oder Grössenklasse zu berechnen, ist es einfacher, jeweils die Bruttostichprobengrösse pro Schicht nach unten zu korrigieren, wenn die Grösse der Population im bereinigten Stichprobenrahmen kleiner ist. Führt man diese Operation pro Schicht aus, erhält man eine Bruttostichprobengrösse von 7438 und entsprechend eine Antwortrate $r_3 = 4408 / 7438 = 0.5926$. Wir führen die Antwortraten pro Schicht aus Datenschutzgründen hier nicht auf. Hingegen können die Antwortraten pro Wirtschaftsgruppe und pro Grössenklasse (Tabelle 15) angegeben werden. Die tiefste Antwortrate bei den Wirtschaftsgruppen haben die Unternehmen des Bergbaus. Die höchste Antwortrate haben die Banken und Versicherungen. Bei den Grössenklassen fällt auf, dass die kleinsten Unternehmen eine leicht höhere Antwortrate als die Unternehmen mit 20-49 Beschäftigten haben. Die höchste Antwortrate haben die grössten Unternehmen. Insgesamt variiert aber die Antwortrate relativ wenig mit der Grösse.

Tabelle 15 Antwortraten nach Wirtschaftsgruppen und Grössenklassen

Wirtschaftsgruppen			
grind	nsnetto	nsbrutto	Rate (%)
1	23	66	34.8
2	246	403	61.0
3	124	219	56.6
4	61	115	53.0
5	212	374	56.7
7	356	548	65.0
8	95	182	52.2
9	61	93	65.6
10	1057	1685	62.7
11	111	172	64.5
12	483	870	55.5
13	449	805	55.8
14	190	396	48.0
15	146	245	59.6
16	175	263	66.5
17	334	505	66.1
20	132	249	53.0
21	153	248	61.7
Total	4408	7438	59.3

Grössenklassen			
gkl	nsnetto	nsbrutto	Rate
1	285	497	57.3
2	1493	2733	54.6
3	1028	1721	59.7
4	962	1536	62.6
5	640	951	67.3
Total	4408	7438	59.3

5 Automatische Datenaufbereitung und Analyse

5.1 Datenübernahme

Die Daten wurden jeweils mit der SAS-Prozedur Import direkt aus den Tabellen der Access-Datenbank PRI_FINAL.mdb importiert. Kleinere Probleme ergaben sich, weil einige Kommentarfelder für SAS zu gross waren. Diese Kommentarfelder wurden nicht vollständig übernommen.

Die wichtigen Tabellen sind:

- T_Ent_Entreprises: Bruttostichprobe
- T_Données_Définitives: Nettostichprobe mit definitiven Daten

- T_INFO: Info über Verlauf Erhebung und Gruppen
- T_Groupe: Namen der Gruppen und Gruppen-Nummer

Die Tabelle T_INFO enthält alle Angaben zu den Gruppierungen und zum Verlauf der Erhebung (7505 Einträge). Die Variablen sind in Tabelle 16 aufgeführt.

Tabelle 16 Wichtige Tabellen der Datenbank

Variablenname	Bemerkung
T_INFO_ID_Q	ID Fragebogen
ID Entreprise	Unternehmensnummer
N° PDF"	Nummer des eingescannten Fragebogen (Seite 1. Für Seite 2 die Nummer um 1 erhöhen)
T_INFO_remplipar	
Rempli au tel	
Part.	
Part. PAS	
1er Rappel	
2ème Rappel	
Groupe	Name der Unternehmensgruppe
T_INFO_TêteGroupe	Unternehmen, dem die Ausgaben der Unternehmensgruppe zugeordnet wurden
Gd Compte	
Transmis à LINK	
Retourné par LINK	
Modifié par LINK	
Refusé poste	Durch Post zurückgeschickt (Adresse nicht auffindbar)
Retour sans Q	
Raison non part.	

Die Tabelle T_ENT_entreprises entspricht der Bruttostichprobe die von METH an UNL vor der Erhebung geliefert wurde. Allerdings sind die Einschlusswahrscheinlichkeiten und Gewichte nur als ganze Zahlen gespeichert. Darum wird später auf die Speicherung der Bruttostichprobe im SAS-Format zurückgegriffen (sample040330.sas7bdat).

Die Tabellen T_Données_xxx enthalten die Daten in verschiedenen Stadien. Die Variablen sind in Tabelle 17 aufgeführt. Im nachfolgenden Text sind die Namen zum Teil abgekürzt worden (z.B. anstatt T_QDON_C344 nur C344).

Die Tabelle vom 9. Dezember 2004 enthält 4434 Unternehmen bzw. Unternehmensgruppen.

Die für die Aufbereitung entwickelten Programme sind in Tabelle 28 im Anhang aufgeführt.

5.2 Strategie für die Aufbereitung

Die Strategie für die Aufbereitung der Daten ist, eine Kompatibilität von unten nach oben zu priorisieren (vgl. die Bilanzkontrollen im nächsten Abschnitt). Wir gehen davon aus, dass der

Tabelle 17 Variablennamen in der Access-Datenbank

T_INFO_ID_Q	ID Fragebogen
T_INFO_ID_Ent	Unternehmenskennzeichen
Unité fausse (qqf unes)	
Ttes unités fausses	Ufausses2 (Kilofehler)
T_INFO_AutresC	
T_INFO_Comcorrec	
T_INFO_Supprimer	
T_INFO_A105	Vollzeit
T_INFO_A106	Teilzeit 1
T_INFO_A107	Teilzeit 2
T_INFO_B201	Umweltschutzausgaben Ja
T_INFO_B202	Umweltschutzausgaben Nein
T_INFO_C301	Umweltschutzausgaben End-of-Pipe/Luftreinhaltung und Klimaschutz
....	Code siehe Abbildung 8
T_QDON_C346	Einnahmen aus Nebenprodukten Total 2003
T_QDON_C351	Personen mit Löhnen aus internen Ausgaben
T_QDON_C361	Umweltzertifiziert Ja
T_QDON_C362	Umweltzertifiziert Nein
T_QDON_D363	Umweltzertifiziert Teilweise
T_QDON_D402	Aufwand für Ausfüllen FB Stunden
T_QDON_D403	Aufwand für Ausfüllen FB Minuten

monetäre Teil des Fragebogens (Variablen C301 bis C346) zuerst bearbeitet werden soll. Dies macht auch für Unternehmensgruppen und ihre Teil-Unternehmen Sinn, da die monetären Variablen in sich konsistent sein sollten unabhängig von der Einheit. Dann kann ein erster Vergleich mit dem Personaleinsatz (C351) gemacht werden. Eventuell muss der Personaleinsatz zuerst mit den Fragen zur Beschäftigung (A105-A107) überprüft werden.

In diesem Zeitpunkt wird die Kontrolle der Beschäftigten laut Umfrage und laut BUR notwendig. Dies kann bei den Unternehmensgruppen nur geschehen, wenn sie konsolidiert sind. Nach der Kontrolle- und Einsetzung der monetären Fragen wurden also die Arbeiten für die Konsolidierung der Unternehmensgruppen durchgeführt (vgl. Abschnitt 5.7).

Ein weiteres Element der Strategie war, die Kontrollen (englisch edit checks) strikt von den Einsetzungen (englisch imputation) zu trennen. Bei den Kontrollen werden also ausschliesslich Regeln für die Überprüfung angewandt und Flaggen gesetzt. Diese Flaggen werden dann bei den Einsetzungen verwendet, um bei einzelnen Daten Werte einzusetzen.

Der Prozess der Datenaufbereitung ist natürlich iterativ. Zum Beispiel wurden die Beschäftigten (Abschnitt 5.6) mehrere Male untersucht: Vor und nach der Bildung von Unternehmensgruppen. Die Kontrolle der Beschäftigten legt nahe, zwei weitere Unternehmens-Gruppen zu bilden. Nach der ersten Kontrolle der Beschäftigten wurden z.T. die Angaben der Unternehmen durch die Angaben aus dem BUR ersetzt, weil offensichtliche Missverständnisse aufgetreten waren. Die Gruppenbildung und die Konsistenz-Tests mussten dann wiederholt werden.

Diverse individuelle Fehler mussten in einem Spezial-Programm individuell korrigiert werden. Es handelt sich um verifizierte Fehler (z.B. Ersetzung der weltweiten Ausgaben durch geschätzte Ausgaben in der Schweiz) oder Fehler die durch manuelle Kontrollen entdeckt wurden. Darunter fallen Kilofehler (Abschnitt 5.8), Zehner-Fehler (falsche zusätzliche 0), falsche vertikale Summierung der Einnahmen.

5.3 Bilanz-Kontrollen

Bilanz-Kontrollen werden am Beispiel der End-of-Pipe Ausgaben erklärt: Es werden die 4 Teilbeträge nach Anwendungsbereich (Luftreinhaltung und Klimaschutz = C301, Gewässerschutz = C311, Abfallbewirtschaftung = C321, Andere Umweltbereiche = C331) und das Total der End-of-Pipe Ausgaben 2003 = C341 abgefragt.

Der Code für fehlenden Wert, der von LINK eingeführt wurde (99999), und negative Werte werden für die Variablen C301, C311, C321, C331, und C341 auf fehlend (SAS Wert ".") gesetzt.

Für jede Variable wurde eine Indikator-Variable r (in SAS z.B. RC301 für C301) gesetzt. Wir verwenden den Index i für die Beobachtungen und den Index k für die Variablen. Die Variable y_{ik} bezeichnet also den Wert der k -ten Ausgabe bei der i -ten Unternehmung.

$$r_{ik} = \begin{cases} 0, & y_{ik} \text{ fehlt;} \\ \text{sig}(y_{ik}) + 1, & \text{sonst,} \end{cases} \quad (5)$$

wobei $\text{sig}()$ die Signum-Funktion den Wert -1 bei strikt negativen Werten, 0 bei Werten $= 0$ und 1 bei strikt positiven Werten annimmt. Da negative Werte vorgängig auf fehlend gesetzt wurden, ergibt $r_{ik} = \text{sig}(y_{ik}) + 1$ den Wert 1 , wenn $y_{ik} = 0$ ist, und sonst den Wert 2 .

Sei I_{EoP} die Menge der Indices für die Teilbeträge der End-of-Pipe Ausgaben. Die Summe der Teilbeträge über die nicht fehlenden Werte wird gebildet:

$$s_{i,EoP} = \sum_{k \in I_{EoP}, r_{ik} > 0} y_{ik}. \quad (6)$$

In SAS wird dies durch die SAS-Funktion `sum` ermöglicht, welche fehlende Werte ausschliesst. Anschliessend wird eine Indikator-Variable für fehlende Teilbetrags-Summe gebildet:

$$r_{i,EoP} = \begin{cases} 0, & \text{falls } \{k \in I_{EoP} : r_{ik} > 0\} = \emptyset. \\ \text{sig}(s_{i,EoP}) + 1, & \text{sonst;} \end{cases} \quad (7)$$

Der Indikator ist also 0 , falls die Subtotal-Summe fehlt, d.h. keinen Summanden enthält.

Erst jetzt wird die Teilbetrags-Summe mit dem erfragten Total y_{ik_0} verglichen, indem die Indikator-Variable $d_{i,EoP}$ berechnet wird:

$$d_{i,EoP} = \begin{cases} 0, & y_{ik_0} \text{ oder } s_{i,EoP} \text{ fehlt;} \\ \text{sig}(y_{ik_0} - s_{i,EoP}) + 2, & \text{sonst.} \end{cases} \quad (8)$$

Die Indikator-Variable $d_{i,EoP}$ nimmt also die Werte $0, 1, 2, 3$ an, je nachdem ob $y_{ik_0} - s_{i,EoP}$ fehlend, negativ, 0 , oder positiv ist.

Falls die Teilbetrags-Summe $s_{i,EoP}$ grösser ist als das erfragte Total y_{ik_0} , wird ein Flag f_{ik_0} gesetzt:

$$f_{ik_0} = \begin{cases} 1, & r_{i,EoP} > 0 \text{ und } r_{ik_0} = 0 \text{ (Total fehlt, aber Teilbetrags-Summe } \geq 0); \\ 2, & d_{i,EoP} = 1 \text{ und } r_{ik_0} > 0 \text{ (Total } < \text{ Teilbetrags-Summe);} \\ 0, & \text{sonst.} \end{cases} \quad (9)$$

5.4 Bilanz-Einsetzungen

Das Total y_{ik_0} wird wie folgt eingesetzt:

$$y_{ik_0} = \begin{cases} y_{ik_0}, & \text{falls } f_{ik_0} = 0; \\ s_{i,EoP}, & \text{falls } f_{ik_0} > 0. \end{cases} \quad (10)$$

Im Fall, dass die Teilbetrags-Summe kleiner ist als das Total, werden die angegebenen Teilbeträge unverändert belassen. Es wird aber zusätzlich eine Variable y_{ir} eingeführt, deren Wert wie folgt eingesetzt wird:

$$y_{ir} = \begin{cases} y_{ik_0} - s_{i,EoP}, & \text{falls } f_{ik_0} = 2; \\ 0, & \text{sonst.} \end{cases} \quad (11)$$

Es wird also eine neue Kategorie von Ausgaben künstlich erzeugt, die bei der Publikation mit "Uebrige Beiträge" ausgewiesen werden muss. Der Vorteil einer solchen Rest-Variable ist, dass kein Modell für die Verteilung des Überschusses auf die Teilbeträge gefunden und geschätzt werden muss.

Im Fall, dass einige Teilbeträge fehlen, andere nicht, können fehlende Werte bei den Totalen falsch ersetzt werden. De facto werden bei unserem Vorgehen in diesem Fall die fehlenden Teilbeträge mit 0 ersetzt und das Total entspricht dann der Summe der vorhandenen Teilbeträge, obwohl die tatsächlichen Teilbeträge grösser als 0 sein könnten. Dies kommt insbesondere dann vor, wenn ein Teilbetrag mit 0 angegeben worden ist, aber der Rest der Zeile (inklusive Total) nicht ausgefüllt worden ist. Eine Anzahl von Fragebogen hat in der ersten Kolonne eine 0 und lauter fehlende Werte in den restlichen Kolonnen. Hier handelt es sich vermutlich eher um eine Weigerung, in alle Felder eine 0 einzufüllen, als um echt fehlende Werte. Der für das Total eingesetzte Wert 0 wäre dann korrekt. Tatsächlich positive, aber fehlende Teilbeträge werden mit dem hier beschriebenen Vorgehen nicht korrigiert. Im Fragebogen wurde betont, dass "keine Ausgaben" mit 0 angegeben werden soll und nicht leer gelassen werden soll. Bei den Rückfragen durch das Marktforschungsinstitut LINK wurden die fehlenden Werte bei Teilbeträgen kontrolliert. Die Wahrscheinlichkeit für fälschlicherweise auf 0 gesetzte Teilbeträge wurde dadurch möglichst klein gehalten.

Eine Kontrolle der Einsetzungen von Totalen ergab, dass in den meisten Fällen die Befragten die Teilbeträge nicht korrekt summiert haben und damit die Einsetzung gerechtfertigt ist. Die Fälle, in denen das erfragte Total grösser ist als die Summe der Teilbeträge, stammen entweder von kleinen Additionsfehlern oder von Teilbetrags-Summen, die 0 sind. In beiden Fällen scheint es gerechtfertigt, das angegebene Total unverändert zu belassen.

5.5 Kontrolle und Einsetzung der monetären Variablen

Es wurde ein generelles Kontrollprogramm (`kontrollen.sas`) für die monetären Fragen und die Frage über den Personaleinsatz (C351) entwickelt, das nach jeder Einsetzung oder Korrektur Rückmeldungen über die Veränderungen in den Indikatoren r und d gibt. Flaggen f werden nur bei den Einsetzungen (`einsetzungen-geb-bes-tot.sas`) gesetzt, d.h. sie werden nicht mehr verändert. Hingegen werden die Antwort- und Kontroll-Indikatoren r und d aufgrund des gegenwärtigen Stands der Daten neu berechnet.

Zuerst werden die Wasser- und Abfallgebühren (C315 und C325) kontrolliert und in die entsprechenden Teilbeträge C314 und C325 integriert (analog zu einer Bilanz-Einsetzung). Dann werden die Bilanzgleichungen für die Ausgaben-Arten eingesetzt (also z.B. C304 bis C334 und C344).

Die Frage nach dem Personal-Einsatz (C351) ist nicht ganz klar formuliert. Die Regel $C351 * 100 > C343$ (Annahme, dass 1 Stelle ca. 100'000 Franken kostet) führt zu 683 Beobachtungen mit Fehlern (cf. Abbildung 1). Die Regel $C351 * 30 > C343$ führt immer noch zu 334 Fehlern. Es scheint nicht sinnvoll, eine Korrektur der totalen laufenden Ausgaben (C343) aufgrund des Personaleinsatzes vorzunehmen.

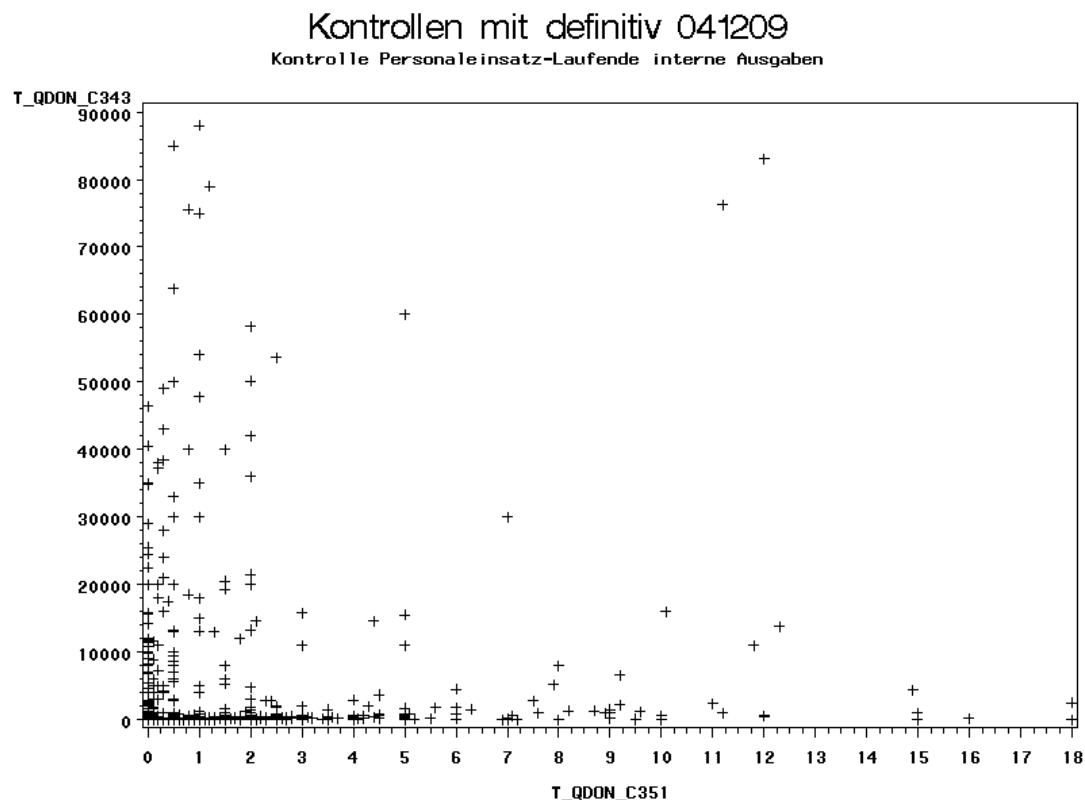


Abbildung 1 Total laufende Kosten gegen Personal-Einsatz (Nur Personal-Einsatz < 20)

Auch beim Vergleich der angegebenen Anzahl Beschäftigter mit dem Personaleinsatz für Umweltschutz zeigen sich bei einigen Unternehmen zu hohe Personaleinsätze. Tabelle 18 zeigt die Angaben von 15 Unternehmen deren Personaleinsatz für Umweltschutz 20% oder mehr

der Vollzeitäquivalente beträgt. Es gibt 35 Unternehmen, deren Anteil mehr als 10% und 79, deren Anteil mehr als 5% beträgt. Es wurde davon abgesehen, den Personaleinsatz zu bereinigen bzw. als Grundlage für eine Einsetzung heranzuziehen.

Tabelle 18 Beobachtungen mit Personaleinsatz $\geq 0.2 \cdot$ Vollz.-Äquivalente

VZ-Aeq.(FB)	Besch. (FB)	Besch. (BUR)	Personaleinsatz
aetot	atot	betot	C351
22.00	22	22	9.5
52.84	55	57	15.0
14.87	16	20	5.0
19.74	22	29	22.0
4.15	8	92	2.3
20.00	20	21	8.0
2.74	5	23	9.0
26.51	28	39	10.0
112.20	142	148	31.0
1.23	2	2	1.3
1.00	1	2	1.0
1.00	1	1	1.0
1.87	3	1	1.0
16.53	21	2	7.2
1.00	1	1	1.0

5.6 Anzahl Beschäftigte

In diesem Abschnitt soll die Konsistenz der Angaben über die Anzahl Beschäftigte untersucht werden. Die Kontrollen werden mehrere Male durchgeführt, z.B. vor und nach der Konsolidierung der Unternehmensgruppen.

Auf der ersten Seite des Fragebogens wird die Anzahl Beschäftigte nach Vollzeit, Teilzeit 1 und Teilzeit 2 abgefragt (Variablen $v = t_qdon_a105$, $t_1 = t_qdon_a106$, $t_2 = t_qdon_a107$). Die Summe ergibt die Anzahl Beschäftigte nach Fragebogen ($atot$: $a_i = v_i + t_{1i} + t_{2i}$), die mit der Anzahl Beschäftigten laut BUR zum Zeitpunkt der Ziehung ($betot=b_i$) verglichen werden kann. Ausserdem kann ein grobes Vollzeitäquivalent ($aetot$: e_i) nach folgender Formel berechnet werden:

$$e_i = v_i + 0.64t_{1i} + 0.23t_{2i}. \quad (12)$$

Die Variable a'_i entsteht aus $atot$, indem fehlende Werte durch $betot$ ersetzt werden.

Das Antwortverhalten wird in den Tabellen 19 und 20 beschrieben. Dabei ist die Variable $dabetot$ wie folgt definiert:

$$d_{abi} = \begin{cases} 1, & |a'_i - b_i| < \max(b_i/2, 100) ; \\ 0, & \text{sonst.} \end{cases} \quad (13)$$

Die Variable $dabetot$ ist also ein Indikator für die Erfüllung einer Kontroll-Regel, d.h. wenn $d_{ab} = 1$, dann passiert die Beobachtung diese Regel. Die Beschäftigten nach Fragebogen

sollten als nicht weniger als die Hälfte oder mehr als das Doppelte der Beschäftigten nach BUR ausmachen. Damit die Kontrolle aber bei kleinen Unternehmen nicht zu schnell anspricht, muss die Abweichung mindestens 100 betragen, dass die Kontrolle einen Fehler ergibt. Der Indikator dabetot wird auch 1 gesetzt, wenn atot fehlt. Die Anzahl Beschäftigte nach BUR fehlt nie und ist nie 0. Dank der zweiten Bedingung passieren grosse relative Veränderungen, die aber nur kleine absolute Veränderungen betreffen.

Der Indikator für fehlende Werte für die Anzahl Beschäftigte (r_{ai}) wird analog $r_{i,EoP}$ (7) berechnet.

Das Ergebnis der Kontrolle vor der Konsolidierung der Unternehmensgruppen ist in Tabelle 19 aufgeführt.

Tabelle 19 Vor Konsolidierung Unternehmensgruppen: Antworttypen für Anzahl Beschäftigte und Vergleich mit Angaben BUR

	d_{abi}		
r_{ai}	0	1	All
0	.	122	122
1	13	8	21
2	296	3995	4291
All	309	4125	4434

Wir müssen also in 296 Fällen den Abweichungen zwischen atot und betot nachgehen.

Folgende Probleme können sich hinter einer Abweichung verstecken.

- Es kann sich um ein anderes Unternehmen (andere Arbeitsstätten), eine Unternehmensgruppe, oder ein stark verändertes Unternehmen handeln.
- Es kann eine Unternehmensgruppe sein. Die Unternehmensgruppen haben auch einen Einfluss auf die Einschlusswahrscheinlichkeit.
- Die Anzahl Beschäftigte atot ist u.U. im Datensatz falsch. Z.B. wurden gewisse manuelle Einsetzungen von fehlenden A106 und A107 nicht korrekt ausgeführt oder Angaben auf dem Fragebogen, die sich auf eine Gruppe bezogen, wurden auf "fehlend" gesetzt.) Wir gehen davon aus, dass betot aus dem BUR korrekt ist.

Weitere Hinweise können sich eventuell durch den Vergleich mit dem Personaleinsatz (C351) und mit den Ausgaben ergeben.

Nach der Konsolidierung der Gruppen ergaben sich wesentlich weniger Abweichungen (vgl. Tabelle 20), die dann so belassen wurden.

5.7 Unternehmensgruppen

Die Zusammensetzung der Gruppen, wie sie während der Erhebung erfasst wurden, wurde anhand der Nettostichprobe untersucht. Genauer gesagt, wurden die Informationen über die

Tabelle 20 Nach Konsolidierung Unternehmensgruppen: Antworttypen für Anzahl Beschäftigte und Vergleich mit Angaben BUR

	d_{abi}		
r_{ai}	0	1	All
0	.	121	121
1	2	19	21
2	92	4174	4266
All	94	4314	4408

Gruppen aus der Datei T_INFO zu den Unternehmen in der Nettostichprobe hinzugefügt. Ausserdem wurde jeder Unternehmung der Indikator für die Schicht des Stichprobenplans und die Brutto-Auswahlwahrscheinlichkeit nach dem Stichprobenplan (Output von SAS PROC SURVEYSELECT) angefügt. Dabei wurde der Abgleich mit Hilfe der Variable t_info_id_ent, d.h. der Unternehmensnummer laut BUR, durchgeführt.

Bei drei Unternehmensgruppen waren die Angaben über den Umfang der Gruppe unklar. Sie wurde zum Teil telefonisch abgeklärt oder aus Geschäftsberichten eruiert und in separaten Listen erfasst (Excel-Datei gruppenezusatz.xls mit Variablen entid, gruppe, tgruppe (Indikator für Hauptunternehmen)). Andere Gruppen wurden mit Hilfe der Informationen des BUR (Feld "Gruppe 4") ergänzt.

Damit ergibt sich bei den Unternehmensgruppen folgende Situation: Von insgesamt 38 Unternehmensgruppen haben zwei Unternehmensgruppen keine Antwort geliefert. Der Grossteil der Unternehmensgruppen wurde bereits vollständig in der Bruttostichprobe erfasst. Drei Unternehmensgruppen wurden mit Hilfe von externer Information ergänzt. Wie im Abschnitt 5.6 erwähnt wurden zwei Unternehmensgruppen erst aufgrund der Abklärungen zu den Unterschieden zwischen der Anzahl Beschäftigte gemäss BUR und gemäss Fragebogen geschaffen.

Bei einigen Unternehmensgruppen haben sich auch Teil-Unternehmen an der Erhebung beteiligt. Es zeigt sich aber, dass die Qualität dieser Angaben nicht besser ist als die Gesamt-Angaben der Haupt-Unternehmung für die ganze Gruppe. Die Gesamt-Angaben wurden darum akzeptiert und alle Teil-Unternehmen von Unternehmensgruppen wurden aus der Netto-Stichprobe eliminiert.

An den Unternehmensgruppen sind insgesamt 230 Unternehmen beteiligt. Die Aufteilung auf Brutto- und Netto-Stichprobe ist in Tabelle 21 ersichtlich.

Tabelle 21 Unternehmen in Unternehmensgruppen nach Brutto- und Netto-Stichprobe

	innetto		
inbrutto	0	1	ALL
0	68	.	68
1	102	60	162
All	170	60	230

Die Analyse der Aufteilung der Beschäftigung der Unternehmensgruppen auf die Wirtschaftsgruppen zeigt, dass bei fast allen Unternehmensgruppen ein klarer wirtschaftlicher Schwer-

punkt erkennbar ist. Es wird darum von einer Aufteilung der Umweltschutzausgaben auf die Wirtschaftsgruppen der Unternehmensgruppe abgesehen. Diese Aufteilung hätte auf einem mehr oder weniger komplexen Modell beruht, das die Anzahl Beschäftigte und eventuell die Umweltschutzausgaben pro Kopf der Unternehmen, die nicht in Gruppen organisiert sind, berücksichtigt. Da das Modell für die Aufteilung nicht überprüfbar ist und ein klarer wirtschaftlicher Schwerpunkt für fast alle Unternehmensgruppen sichtbar ist, wurde keine Aufteilung durchgeführt. Es entsteht dadurch natürlich eine gewisse Unschärfe bei der Zuteilung der Umweltschutzausgaben zu den Wirtschaftsgruppen. Diese Art der Unschärfe ist aber bereits bei der Zuteilung der einzelnen Unternehmen zu einer Wirtschafts-Aktivität gegeben: Eine Unternehmung wird aufgrund der *Haupt*-Aktivität der beteiligten Arbeitsstätten eingeteilt.

Eine Unternehmensgruppe k mit Unternehmens - Einschlusswahrscheinlichkeiten $\pi_i, i \in G_k$, hat als Einschlusswahrscheinlichkeit die Gegenwahrscheinlichkeit zum Ereignis, dass kein einziges Unternehmen der Gruppe eingeschlossen wird.

$$\pi_k = P[k \in S] = 1 - \prod_{i \in G_k} (1 - \pi_i). \quad (14)$$

Falls also nur ein einziges Unternehmen der Gruppe mit Sicherheit in der Stichprobe ist, dann ist die Unternehmensgruppe mit Sicherheit in der Stichprobe, also $\pi_k = 1$. Es ergaben sich nur zwei Unternehmensgruppen, in denen alle Unternehmen Einschlusswahrscheinlichkeiten $\pi_i < 1$ hatten. Diesen wurde eine Einschlusswahrscheinlichkeit nach obiger Formel (14) zugeordnet.

Die Hilfsvariable x_i , "Anzahl Beschäftigte gemäss BUR zum Zeitpunkt der Ziehung", muss auch im Stichprobenrahmen entsprechend der Zusammensetzung und der Wirtschaftsaktivität der Unternehmensgruppen geändert werden. Anders gesagt müssen die Teilunternehmen der Unternehmensgruppen im Stichprobenrahmen entfernt werden und dem Hauptunternehmen die richtige Wirtschaftsgruppe (bzw. Wirtschaftsaktivität falls die Quotientenschätzer auf dieser Ebene angesetzt werden sollen) zugeordnet werden und die Hilfsvariable muss die Summe der Hilfsvariablen der Teil-Unternehmen enthalten: $x_k = \sum_{i \in G_k} x_i$. Unternehmen, die aufgrund externer Informationen den Unternehmensgruppen zugeteilt worden waren, aber nicht im Stichprobenrahmen vorhanden sind, wurden aus den Gruppen entfernt.

Die Zuordnung der Unternehmen zu ihren Gruppen, die Berechnung der Einschlusswahrscheinlichkeiten, die Bereinigung des Stichprobenrahmens und die Zuordnung der Informationen über Beschäftigung und Wirtschaftsaktivität zum Hauptunternehmen, sowie dessen entsprechende Einreihung in eine Wirtschaftsbranche, erfolgt in den Programmen *gruppen-brutto*, *rahmen2*, *gruppen-netto* (vgl. die Aufstellung der Programme in Tabelle 28).

5.8 Untersuchung Kilo-Fehler

Ein üblicher Fehler bei monetären Grössen ist, dass anstatt Angaben in 1'000 Franken Angaben in Franken gemacht werden. Wir nennen diesen Fehler den Kilo-Fehler. Alle Fragebogen wurden manuell darauf kontrolliert, ob ein Kilo-Fehler vorliegt. Normalerweise werden alle Angaben falsch aufgeführt. In Einzelfällen tritt der Kilo-Fehler aber nur bei einzelnen Angaben auf. Es gibt auch 10-er Fehler und 100-er Fehler.

Bei der manuellen Kontrolle wurden Fragebogen mit einem Kilo-Fehler-Flag markiert (T_Unités-fausses2). Zum Vergleich wurde die Summe der totalen Ausgaben (totaus = C341 + C342+

C343+ C344) gegen die Vollzeitäquivalente aufgetragen. Dabei wurden fehlende Vollzeitäquivalente durch die Anzahl Beschäftigte ersetzt. Abbildung 2 zeigt das Diagramm der Logarithmen der totalen Ausgaben gegen die Logarithmen der Vollzeitäquivalente für eine Wirtschaftsgruppe. Die eingezeichneten Geraden wurden wie folgt berechnet.

Der Achsenabschnitt der unteren Geraden ist $\text{med}(\log(y+1) - \log(x+1))$, wobei y die totalen Ausgaben und x die Vollzeitäquivalente sind. Abgesehen von der Verschiebung um $+1$ handelt es sich also um den Median der logarithmierten pro-Kopf Werte. Mit anderen Worten postulieren wir das Modell $y+1 = (x+1)^\beta$ und schätzen $\log(\beta)$ mit dem Median. Die Steigung beider Geraden ist 1. In der log-log-Skala wird ein Faktor $k = 1000$ eine vertikale Verschiebung auslösen: $\log(ky+1) \approx \log(k) + \log(y+1)$. Falls weniger als die Hälfte der Beobachtungen durch einen Kilo-Fehler betroffen sind, wird der Median durch den Faktor k nicht beeinflusst. Die obere Gerade hat die selbe Steigung wie die untere, ist aber um $2 * \text{med}(\log(y+1) - \log(x+1))$, die doppelte Median-Abweichung der logarithmierten Pro-Kopf-Werte, verschoben. Es zeigt sich, dass die manuell gefundenen Kilo-Fehler (in der Abbildung durch Punkte innerhalb der Kreise bezeichnet) sehr gut durch diese Abgrenzung identifiziert werden können. Allerdings musste in der Gruppe 21 die Grenze für den Kilo-Fehler von Hand festgelegt werden, da sie zu wenige Beobachtungen für eine effiziente Medianschätzung enthält. Einige manuell nicht entdeckte Kilo-Fehler wurden nachträglich als solche klassifiziert. Es bleiben aber auch einige durch dieses automatische Verfahren identifizierte Kilo-Fehler, welche nach manueller Abklärung nicht als Kilo-Fehler gelten konnten. Für die Imputation wurde auf die korrigierte manuelle Identifikation abgestützt. Kilo-Fehler, teilweise Kilo-Fehler sowie 10er und 100er Fehler wurden in einem Spezial-Programm eingesetzt (deterministische Korrekturen). Insgesamt hatten 279 oder 6.3% der Beobachtungen einen Kilo-Fehler.

5.9 Einsetzungen für Gebühren

Im Fragebogen wurde auf Seite 1 gefragt, ob die Unternehmen Ausgaben für Umweltschutz haben. Andererseits kann man davon ausgehen, dass ein Unternehmen bereits wenn nur eine Kehricht-Gebühr bezahlt werden muss, Umweltschutz-Ausgaben hat. Insofern erübrigt sich die Frage auf der ersten Seite oder sie muss dahingehend geändert werden, dass spezifisch nach Investitionen oder nach Ausgaben über die Abwasser- und Abfall-Gebühren hinaus gefragt wird. Bei Unternehmen, welche für die Gebühren für Abfall und Wasser nichts oder null angegeben haben, soll eine konservative Schätzung für die Gebühren eingesetzt werden. Als Hilfsvariable für diese Schätzung wurden Vollzeitäquivalente und Beschäftigte nach BUR geprüft. Da Vollzeitäquivalente keine klar besseren Resultate erbrachten, wurde mit der Anzahl Beschäftigte nach BUR gearbeitet. Für die Modellierung wurden nur die Unternehmen mit strikt positiven Ausgaben für Wasser bzw. Abfall herangezogen.

Die Modelle wurden getrennt nach Wirtschaftssektor 2 und 3 und ohne die grössten Unternehmen (ab 5000 Beschäftigte) angepasst. Ausserdem wurden die kleinen Unternehmen in den Wirtschaftsgruppen 20 und 21 separat behandelt. Es zeigte sich, dass diese zwei Wirtschaftsgruppen zu wenige Beobachtungen mit positiven Gebühren enthielten, um eine zuverlässige Modellierung zuzulassen.

Die Diagramme Gebühren gegen Beschäftigte zeigen, dass in den Daten immer noch viele Ausreisser sind. Um zu vermeiden, dass Ausreisser die Einsetzungen zu stark beeinflussen, wurde die Modellierung robust durchgeführt (Resultate siehe Tabelle 22).

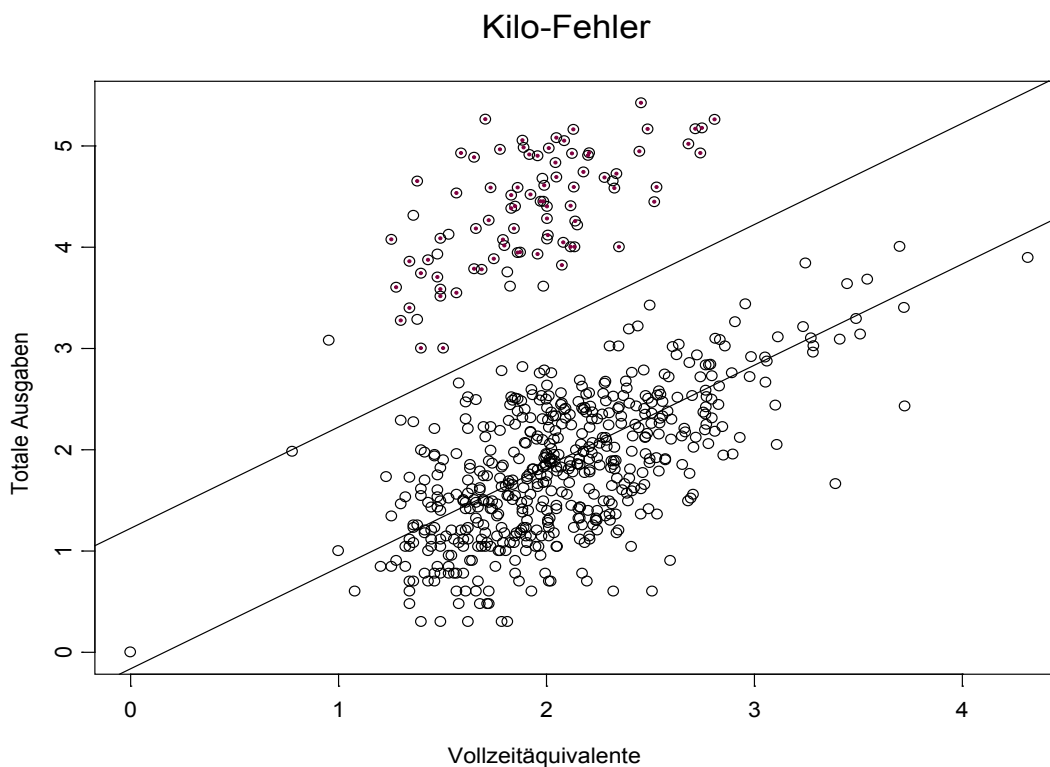


Abbildung 2 Totale Ausgaben vs Vollzeitäquivalente (je logarithmiert)

Die Gebühren wurden mit Hilfe des folgenden Modells eingesetzt:

$$\hat{y} = \max(0, \hat{\beta}x + \hat{\sigma}\sqrt{x}\epsilon). \quad (15)$$

Dabei stammen die Koeffizienten $\hat{\beta}$ und die Skalen $\hat{\sigma}$ aus der obigen Modellierung (siehe Tabelle 22). Der Fehlerterm mit $\epsilon \sim N(0, 1)$ wurde addiert, um die Variabilität der Einsetzungen realistischer zu machen.

Tabelle 22 Modellierung Gebühren mit Anzahl Beschäftigte

	Wasser = C315		Abfall = C325	
Sektor	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\beta}$	$\hat{\sigma}$
Sektor 2	0.144378	2.29043	0.111029	2.11912
Sektor 3	0.084789	1.88345	0.104306	1.91496

Nach der Einsetzung der Gebühren mussten die Bilanz-Kontrollen für die Gebühren noch einmal durchgeführt werden.

Der Einfluss der verschiedenen Kontrollen und Einsetzungen auf die Schätzungen ist schwierig zu verfolgen, weil die Entwicklung normalerweise iterativ erfolgt. Die eigentlichen Schätzverfahren werden auch erst im Nachhinein entwickelt. Die Tabellen 23 und 24 zeigen die Veränderungen einer provisorischen Schätzung, die allein auf Horvitz-Thompson Schätzern und auf einer einfachen Korrektur der Antwortausfälle basiert. In Tabelle 23 ist ersichtlich, dass der

Effekt der schlussendlich angewandten Kontrollen und Einsetzungen für die totalen Ausgaben beträchtlich ist: Bei den mittleren und grossen Unternehmen ($klein=0$) ergibt sich eine Erhöhung um knapp 19%. Die Erhöhungen bei den einzelnen Ausgabe-Arten ist aus Tabelle 24 ersichtlich. Sie sind moderat, ausser beim Einkauf von Dienstleistungen (C344), welcher die kommunalen Gebühren enthält: Diese Ausgaben-Art wird um 66% erhöht. Da die Erhöhungen alle in die gleiche Richtung gehen, ist der oben erwähnte Effekt für die totalen Ausgaben gut spürbar.

Tabelle 23 Effekt auf totaus

	betot		totaus vor Einsetzung		totaus nach Einsetzung	
klein	N	Sum	N	Sum	N	Sum
0	4123	1692980	2508	1840368	4123	2188832
1	285	952180	96	260044	99	262116
All	4408	2645160	2604	2100412	4222	2450948

Tabelle 24 Effekt auf Ausgaben-Arten und Einnahmen

Vor Einsetzung										
	C341		C342		C343		C344		C346	
klein	N	Sum	N	Sum	N	Sum	N	Sum	N	Sum
0	2472	304210	2474	391638	2488	652284	2480	492236	2432	61094
1	96	53358	95	43810	94	41376	94	121500	91	6194
All	2568	357568	2569	435448	2582	693660	2574	613736	2523	67287

Nach Einsetzung										
	C341		C342		C343		C344		C346	
klein	N	Sum	N	Sum	N	Sum	N	Sum	N	Sum
0	2492	305274	2491	402333	2517	661711	4123	819514	2448	63525
1	98	53358	98	44328	96	41376	97	123054	94	6712
All	2590	358631	2589	446661	2613	703087	4220	942568	2542	70236

6 Schätzverfahren

Als Grundbaustein für die Schätzverfahren dienen die üblichen Horvitz-Thompson Schätzer, welche erwartungstreu sind. Diese Schätzer gehen von der Einschlusswahrscheinlichkeit nach Stichprobenplan $\pi_i, i \in S$, aus. Der Horvitz-Thompson Schätzer für das Total einer Variable y ist dann:

$$T_y^{HT} = \sum_{i \in S} \frac{y_i}{\pi_i}. \quad (16)$$

Der Kehrwert der Einschlusswahrscheinlichkeit wird oft als Stichproben-Gewicht bezeichnet: $w_i = 1/\pi_i$. Damit lässt sich der Horvitz-Thompson Schätzer als gewichtete Summe schreiben:

$$T_y^{HT} = \sum_{i \in S} w_i y_i. \quad (17)$$

Für unseren Fall einer geschichteten Stichprobe ist $\pi_i = \tilde{n}_h / N_h$, wobei $\tilde{n}_h$ und N_h die Bruttostichprobengrösse und die Populationsgrösse der Schicht h , welche die Beobachtung i enthält, bezeichnen. Der Horvitz-Thompson Schätzer ist also das übliche geschichtete Mittel:

$$T_y^{HT} = \sum_{i \in S} w_i y_i = \sum_{h=1}^L \sum_{i \in S_h} \frac{N_h}{\tilde{n}_h} y_i. \quad (18)$$

Als erste Korrektur für die Antwort-Ausfälle wird die Bruttostichprobengrösse $\tilde{n}_h$ durch die Nettostichprobengrösse n_h ersetzt. Dieses Vorgehen liefert erwartungstreue Horvitz-Thompson Schätzer, falls die Antwort-Ausfälle innerhalb der Schichten einfache Zufalls-Stichprobe sind (entsprechend "Missing Completely at Random"). Wir bezeichnen diesen Schätzer im Folgenden immer noch als Horvitz-Thompson Schätzer oder geschichtetes Mittel.

6.1 Schichtung und selbst-repräsentierende Unternehmen

Wie üblich bei Unternehmenserhebungen bestehen enorme Grössenunterschiede unter den Unternehmen. Es stellt sich die Frage, ob die grössten Unternehmen mit den selben Modellen erfasst werden können, wie die mittleren und ob die Modelle für die mittleren Unternehmen auf die kleinsten Unternehmen mit weniger als 20 Beschäftigten übertragen werden können. Insbesondere um den Einfluss von extremen Werten bei den grössten Unternehmen einzuschränken, wurde für jede Wirtschaftsgruppe eine spezielle Schicht für die grössten Unternehmen geschaffen. Diese grössten Unternehmen werden als selbst-repräsentierend bezeichnet, weil ihre Ergebnisse nicht auf eine grössere Population hochgerechnet werden, d.h. jedes dieser selbst-repräsentierenden Unternehmen zählt im wesentlichen nur für sich selbst. Die Grenze für die selbst-repräsentierenden Unternehmen wurde wie folgt gewählt: Bei $\text{betot} > 5000$ sind 12 Unternehmen in dieser Klasse, bei $\text{betot} > 2000$ sind es 44. Bei einigen Unternehmen zwischen 2000 und 5000 kann man durchaus annehmen, dass ähnliche Unternehmen in der Population vorhanden sind. Die Unternehmen mit mehr als 5000 Beschäftigten sind hingegen vermutlich sehr speziell und können nicht hochgerechnet werden auf alle Unternehmen ab 250 Beschäftigte und können auch nicht ausgetauscht werden. Diese grössten Unternehmen hatten alle Einschlusswahrscheinlichkeit 1. Eine Gewichtung rein nach der Einschlusswahrscheinlichkeit würde also ein Gewicht von 1 ergeben. Nur in einer Wirtschaftsgruppe hat es bei diesen selbst-repräsentierenden Unternehmen tatsächlich Antwortausfälle gegeben. Da davon ausgegangen werden kann, dass die ausgefallenen Unternehmen tatsächlich Umweltschutzausgaben hatten, wurde in dieser Schicht vom Gewicht 1 abgewichen, d.h. die Gewichte nach Stichprobenplan wurden noch mit der Ausfallrate dieser selbst-repräsentierenden Schicht dividiert.

Da die Ausfallraten bei den kleineren und mittleren Unternehmen mit 20 bis 5000 Beschäftigten wesentlich grösser sind, hätte eine Korrektur der Antwortausfälle ohne die Separierung der selbst-repräsentierenden Unternehmen eine zu starke Korrektur zur Folge gehabt.

6.2 Wahl des Schätzmodells

Um die Schätzgenauigkeit zu erhöhen und um eine noch feinere Korrektur der Antwortausfälle zu erreichen, soll die Anzahl Beschäftigte nach BUR (betot) nicht nur bei der Bildung

der Schichten (vorgegeben nach Stichprobenplan plus die oben beschriebene Schicht der selbst-repräsentierenden Unternehmen) berücksichtigt werden, sondern eventuell auch bei einem Quotientenschätzer. Der Quotientenschätzer bezüglich einer Hilfsvariable x_i , die für jedes Unternehmen in der Stichprobe bekannt sein muss, und deren Populationstotale für alle Analysegruppen bekannt sein muss, sieht für eine Analysegruppe G wie folgt aus:

$$T_y^Q = x_{G+} \frac{T_{Gy}^{HT}}{T_{Gx}^{HT}} = \sum_{i \in G} x_i \frac{\sum_{i \in S \cap G} w_i y_i}{\sum_{i \in S \cap G} w_i x_i}. \quad (19)$$

Dabei bezeichnet G eine Analysegruppe in der Population, von der angenommen werden kann, dass die Korrelation zwischen y_i und x_i einigermassen konstant ist. Der Quotientenschätzer korrigiert also den Wert des Horvitz-Thompson Schätzers für y mit dem Verhältnis aus dem wahren und bekannten Total x_{G+} aus der Population für die Variable x und deren Horvitz-Thompson Schätzer T_{Gx}^{HT} . Der Quotient, der geschätzt wird, ist $\beta = y_{G+}/x_{G+}$ und der Schätzer ist $\hat{\beta} = T_{Gy}^{HT}/T_{Gx}^{HT}$.

Dank der Korrelation zwischen der interessierenden Variablen y_i , in unserem Fall zum Beispiel eine bestimmte Art von Umweltschutzausgaben, und der Hilfsvariable x_i , in unserem Fall die Anzahl Beschäftigten, ist die Varianz des Quotientenschätzers oft kleiner als der Horvitz-Thompson Schätzers.

Der Quotientenschätzer kann aber auch einen systematischen Fehler in die Schätzungen hineinbringen, nämlich dann, wenn die Analysegruppe zu klein oder zu inhomogen ist.

In unserem Fall bieten sich die Wirtschaftsgruppen als Analysegruppen an. Es kann aber nicht davon ausgegangen werden, dass die sich die selbst-repräsentierenden Unternehmen gleich verhalten wie die mittleren Unternehmen. Ausserdem hätten die kleinsten Unternehmen mit weniger als 20 Beschäftigten wegen ihrer kleinen Stichprobenrate ein so dominantes Gewicht, dass der Quotient allein von ihnen bestimmt würde. Die Analysegruppe wurde darum als Zusammenfassung der Unternehmen einer Wirtschaftsgruppe in den Schichten mit 20 bis 5000 Beschäftigten definiert.

Für die kleinsten Unternehmen unter 20 Beschäftigte und für die selbst-repräsentierenden Unternehmen wurden Horvitz-Thompson Schätzer verwendet. Das Gewicht der selbst-repräsentierenden Unternehmen ist dabei jeweils 1 ausser in einer Wirtschaftsklasse, wo es wegen Antwortausfällen etwas grösser ist.

6.3 Untersuchungsbereiche

Untersuchungsbereiche sind Teile der Population, für welche spezielle Auswertungen gemacht werden sollen. In unserem Fall stehen zwei Arten im Vordergrund: Grössenklassen und Grossregionen.

Untersuchungsbereiche sollten in den SAS-Programmen als Domains ausgewiesen werden, damit in den Varianzschätzern die Variabilität der Stichprobengrösse im Untersuchungsbereich berücksichtigt wird. Eine Ausnahme sind Untersuchungsbereiche, die sich als Aggregate von Schichten des Stichprobenplans ausdrücken lassen. Hier ist die Stichprobengrösse (brutto) festgelegt und in den SAS-Prozeduren kann mit dem BY-Statement gearbeitet werden. Damit eine einheitliche Prozedur angewendet werden kann, haben wir jeweils auch im zweiten Fall mit dem DOMAIN-statement gearbeitet.

Quotienten können aber mit der Version 9.1.3 von SAS noch nicht nach Untersuchungsbereichen berechnet werden. Das ist auch nicht unbedingt ratsam, weil die Aggregation von Resultaten nach Untersuchungsbereichen nur dann zum Resultat der aggregierten Untersuchungsbereiche wird, wenn derselbe Quotient verwendet wird.

Der Schätzer für einen Untersuchungsbereich wird also wie folgt berechnet:

$$T_{Dy}^Q = \frac{T_{Gy}^{HT}}{T_{Gx}^{HT}} x_{G \cap D+}. \quad (20)$$

Dies bedeutet, dass bei Quotientenschätzern der Quotient $y_{G \cap D+}/x_{G \cap D+}$ jeweils durch den Schätzer für die gesamte Analysegruppe G geschätzt wird (T_{Gy}^{HT}/T_{Gx}^{HT}) und nicht durch den naheliegenderen Schätzer $T_{GDy}^{HT}/T_{GDx}^{HT}$. Der Quotient über die ganze Analysegruppe kann stabiler geschätzt werden. Ein weiterer Vorteil dieses sogenannten synthetischen Schätzers ist die oben erwähnte Additivität der Resultate. Andererseits hat der synthetische Schätzer aber einen Bias zur Folge, wenn $y_{G \cap D+}/x_{G \cap D+} \neq y_{G+}/x_{G+}$.

Auswertungen nach Grössenklassen, die auf den ursprünglichen Schichten aufbauen, können relativ problemlos mit obigem Schätzer gemacht werden. Es entsteht jedoch ein Problem bei der Varianzschätzung, wenn verschiedene Grössenklassen für diese Auswertung Schichten mit dem selben kombinierten Quotientenschätzer umfassen: Der kombinierte Quotientenschätzer stützt sich auf Informationen aus allen Grössenklassen einer Analysegruppe und ist daher genauer, als ein separater Quotientenschätzer. Damit wird der Schätzer für eine Grössenklasse zu einer Art synthetischer Schätzer. Die Schätzer für die einzelnen Grössenklassen sind erwartungstreu unter der Voraussetzung das das gemeinsame Modell korrekt ist. Die Schätzer sind aber nicht unabhängig, da sie auf demselben Quotienten beruhen. Darum können zur Schätzung der Varianz von Summen über die Grössenklassen, insbesondere des Totals über eine Wirtschaftsklasse, die Varianzen der Totale der Grössenklassen nicht addiert werden, da der Kovarianzterm fehlt. Im Abschnitt 6.6 wird noch einmal anhand eines Beispiels auf dieses Problem eingegangen.

Auswertungen nach Grossregionen könnten im Prinzip auch so durchgeführt werden. Es wird aber aus zwei Gründen davon abgeraten. Erstens ist es schwierig, das Schätzmodell nach Grossregionen zu überprüfen, insbesondere, ob die Quotienten pro Analysegruppe konstant über die Grossregionen sind. Grossregionen haben u.U. schon sehr kleine Stichproben pro Analysegruppe. Der zweite Grund ist, dass es sich um eine Unternehmenserhebung handelt. Dabei wird bei Mehrbetriebsunternehmen als Standort des Unternehmens der Hauptsitz definiert. Die regionale Verteilung der Umweltschutz-Ausgaben wird damit verfälscht. Insbesondere gilt dies bei den Unternehmensgruppen, die wegen aggregierten Antworten gebildet werden mussten (siehe Abschnitt 5.7).

6.4 Ausreisser

Auch nach der Bereinigung der Kilo-Fehler bestehen noch Ausreisser. Die Abbildung 3 zeigt, dass die extrem grossen Beobachtungen in der Variablen C341 auch in der Variablen totaus gross sind. Ein Grund ist, dass totaus C341 als Summanden enthält. Allerdings könnten sich grosse C341 mit kleinen anderen Sub-Totalen kompensieren. In C341 oder in C301 wurden nämlich manchmal die gesamten Ausgaben eingetragen, ohne dass sie weiter unterteilt wurden. Die Einsetzungen übertragen C301 auf C341. Es ist daher nicht selbstverständlich, dass

keine Ausreisser im Diagramm C341 gegen totaus zu sehen sind. Zusätzlich sind Ausreisser im Diagramm C341 gegen betot auch Ausreisser im Diagramm totaus gegen betot. Damit könnte eine einfache Robustifizierung also von totaus bzw. der Beziehung zwischen totaus und betot ausgehen. Das Robustheitsgewicht $u_i(\text{totaus})$ könnte nachher auf alle Totale angewendet werden. Das Vorgehen ist nicht optimal und garantiert auch nicht Robustheit, ist aber wesentlich einfacher als ein univariates Robustheitsgewicht pro Variable zu berechnen und zu verwenden.

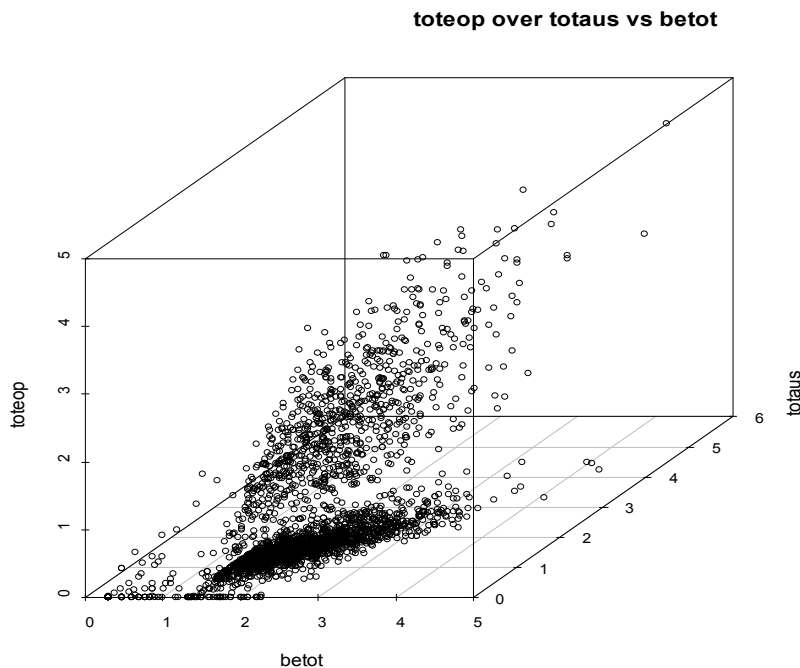


Abbildung 3 C341 (toteop) über totaus gegen betot

Mit Hilfe des Transformed-Rank-Correlation-Algorithmus ([Béguin and Hulliger 2004](#)) wurden auch multivariate Ausreisser gesucht. Der TRC-Algorithmus kann im Unterschied zu klassischen multivariaten Ausreisser-Entdeckungsmethoden auch Beobachtungen mit einzelnen fehlenden Werten (item-nonresponse) und Stichprobengewichte, wie in den vorliegenden Daten der UWSA-Erhebung, berücksichtigen. Dabei wurden die Totale C341, C342, C343 und C344 gemeinsam untersucht. Die Standardisierung auf pro-Kopf-Werte (d.h. Division durch die Anzahl Beschäftigte betot) ist dabei klar nicht genügend, um ein gutes Modell zu finden. Die Werte wurden logarithmiert. Variable C343 hatte so viele Beobachtungen gleich 0, dass nicht mit dem MAD standardisiert werden konnte. Der TRC-Algorithmus in der SAS-Implementation schlägt dann eine Quantils-Deviation für alle Variablen vor (90%-Quantil der absoluten Abweichungen vom Median).

Es wurde eine Variante mit der zusätzlichen Variable $l\text{betot} = \log_{10}(\text{betot} + 1)$ gerechnet, weil die Beschäftigten nie fehlen und damit im ad-hoc Imputationsverfahren des TRC-Algorithmus immer eine gute erklärende Variable zur Verfügung steht.

Die Imputations-Statistik zeigt denn auch, dass die logarithmierte Anzahl Beschäftigte lbetot

Tabelle 25 Imputation Statistics for the variables

Variable	N of obs imputed	using regressor	Correlation
lc341	186	lbetot	0.0998
lc341	1632	lc344	0.1902
lc342	186	lbetot	0.1014
lc342	1	lc341	0.1496
lc342	35	lc343	0.2565
lc342	1597	lc344	0.2247
lc343	187	lbetot	0.2255
lc343	1	lc342	0.2565
lc343	1607	lc344	0.3950
lc344	188	lbetot	0.5792

zwar nicht immer, aber doch bei jeder Variable, als Prädiktor benützt wurde (siehe Tabelle 25).

Tabelle 26 zeigt die erste Schätzung der Kovarianzmatrix des TRC-Algorithmus, welche nicht positiv-definit ist und noch keine ad-hoc Imputationen involviert. Die Korrelationen sind durchwegs positiv, wenn auch nicht sehr gross (Spannweite von 9.9% bis 57.9%). Die TRC- Kovarianzmatrix und der (robuste) Zentrumsvektor werden in Tabelle 27 aufgeführt. Es zeigt sich, dass sogar einige Kovarianzen das Vorzeichen gekehrt haben. Der Einfluss der ad-hoc Imputationen und der Transformation, um eine positiv-definite Matrix zu abzuleiten, ist also ziemlich gross.

Tabelle 26 Spearman-Kovarianzmatrix

lc341	lc342	lc343	lc344	lbetot
0.05517	0.01561	0.01736	0.04661	0.01041
0.01561	0.18022	0.04232	0.09943	0.01913
0.01736	0.04232	0.1386	0.15257	0.03724
0.04661	0.09943	0.15257	0.99556	0.25301
0.01041	0.01913	0.03724	0.25301	0.18022

Tabelle 27 TRC-Kovarianzmatrix und Zentrumsvektor

Kovarianzmatrix				
lc341	lc342	lc343	lc344	lbetot
0.00058	-0.00075	0.00341	0.00479	-0.00184
-0.00075	0.01152	-0.00949	0.00601	0.01977
0.00341	-0.00949	0.02472	0.0174	-0.01604
0.00479	0.00601	0.0174	0.08634	0.00876
-0.00184	0.01977	-0.01604	0.00876	0.06356
Zentrumsvektor				
lc341	lc342	lc343	lc344	lbetot
-0.00063	0.01226	0.00285	0.1143	0.55938

Die Abbildung 4 zeigt die Mahalanobis-Distanzen der Beobachtungen, wenn die TRC- Kova-

rianzmatrix und der TRC-Zentrumsvektor herangezogen werden. Die Punkte auf der waagerechten Komponenten der Kurve sind die Beobachtungen mit mehrheitlich Werten 0 in den Ausgaben. Der ungefähr linear ansteigende Teil zeigt, dass der Hauptteil der Beobachtungen (in logarithmischen Massstab) ungefähr multivariat normalverteilt ist. Die Abweichung von der Normalität kann entweder bei etwa 1'800 oder bei etwa 3'000 angenommen werden. Nimmt man als Grenze für die nicht-ausreissenden Beobachtungen das 99%-Quantil (ungewichtet) der Mahalanobis-Distanzen von 2'336.59, dann erhält man eine erste Liste von möglichen Ausreissern. Diese enthalten natürlich auch univariate Ausreisser (z.B. die grösste Unternehmung in der Erhebung), aber einige Beobachtungen stellen sich als echt multivariate Ausreisser heraus. Unter den Ausreissern hat es keine Beobachtungen mit fehlenden Werten.

Robust Multivariate Outlier Detection & Imputation with incomplete Survey Data: TRC

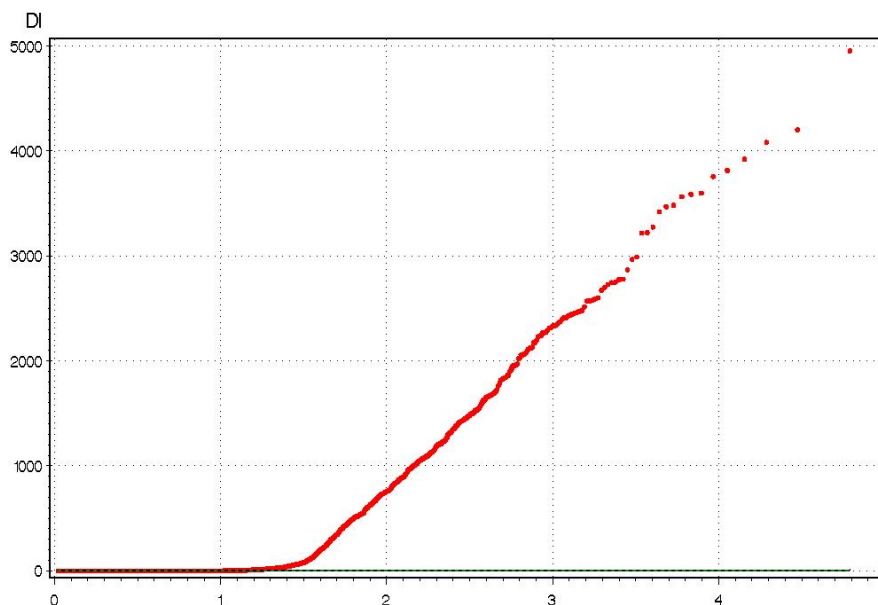


Abbildung 4 Standardisierte Mahalanobis-Distanzen gegen F-Quantile

Interessant ist auch der Vergleich mit den Ergebnissen des TRC-Algorithmus, wenn die Kovariable *lbetot* weggelassen wird. Es werden jetzt auch Beobachtungen mit fehlenden Werten als Ausreisser deklariert. Der TRC-Algorithmus eliminiert viele Beobachtungen von Anfang an, weil sie keine positive Ausgaben haben. Darum fällt beim Quantil-Plot der Mahalanobis-Distanzen auch der horizontale Teil der Kurve weg. Vergleicht man die Mahalanobis-Distanzen der Punkte individuell, dann zeigt sich klar, dass der Einschluss der Kovariablen in einigen Fällen bewirkt, dass fehlende Werte besser imputiert werden und dadurch Ausreisser verschwinden (siehe Punkte mit grossen *diexp* und kleinen *di* in der Grafik 5). Es scheint, dass der Einbezug der Kovariable hilft, gewisse extreme Werte in den Ausgaben zu erklären, wenn sie von grossen Unternehmen stammen. Im übrigen besteht eine gute Korrelation zwischen den beiden Mahalanobis-Distanzen.

Die Ausreisser, die mit TRC entdeckt wurden, könnten mit Hilfe von Nächsten-Nachbarn-Techniken oder linearen Modellen eingesetzt bzw. ersetzt werden. Die Entwicklung einer ent-

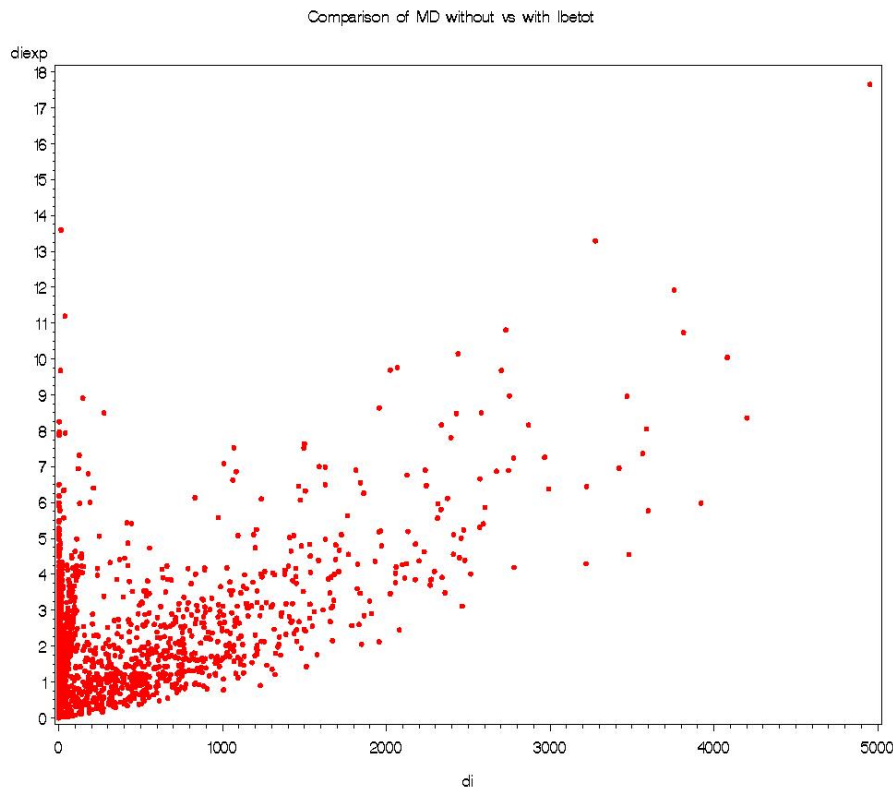


Abbildung 5 Mahalanobis-Distanzen ohne lbetot (diexp) gegen mit lbetot (di)

sprechenden Methode konnte leider aus Zeitgründen nicht weiter verfolgt werden.

6.5 Robustifizierung

Um zu testen, ob eine Schätzung mit Hilfe von robusten Quotientenschätzern wesentliche Abweichungen ergäbe, wurden Ein-Schritt-Quotientenschätzer angewendet (Hulliger 1999). Dabei wird von einer sehr robusten Schätzung der Steigung ein Robustheitsgewicht u_i abgeleitet, welches Beobachtungen mit grossen Residuen heruntergewichtet. Zuletzt wird ein Quotientenschätzer berechnet, der sowohl die Stichprobengewichte w_i als auch die Robustheitsgewichte u_i verwendet. Zusätzlich wurde im Programm `robrat` die Möglichkeit eingebaut, das totale Gewicht $w_i u_i$ nach unten zu beschränken. Die Varianz des robusten Schätzers wurde mit Hilfe der Residuen und der Prozedur `SURVEYMEANS` von SAS geschätzt.

Der Grad der Robustifizierung kann mit einer Abstimmungs-Konstanten festgelegt werden. Da bei asymmetrischen Verteilungen eine Robustifizierung üblicherweise zu einem Bias führt, ist die Wahl der Abstimmungs-Konstanten schwierig. Zwei Abstimmungskonstanten wurden getestet, nämlich $c = 10$ und $c = 20$. Dabei wurde jeweils das totale Gewicht nicht nach unten beschränkt ($tw=0$) bzw. nach unten auf 1 beschränkt ($tw=1$). Die Resultate in Tabelle 29 zeigen, dass der Einfluss der Robustifizierung drastisch ist. Bei einer Abstimmungs-Konstanten von $c = 10$ und ohne Beschränkung ergibt sich anstatt der Schätzung von 2'232'574 ohne Robustifizierung nur noch 1'375'391. Dabei verbesserte sich zwar die Varianz, aber der Varia-

Tabelle 28 Programme für Auswertungen

Programm	Input	Output	Bemerkung
ergebnisse	kontrolliert oder einge- setzt;nstrat	totals;extrap	Nachschichtungsgewichte, HT-Schätzer und Graphiken
rats3	extrap;totals		HT-Schätzer für kleinste und grösste Unternehmen, Quotienten-Schätzer pro Wirtschaftsgruppe für den Rest
robrat	extrap;totals		Robuster Einschnitt- Quotientenschätzer
rats4dom	extrap;totals		Auswertung nach Gröszen- klassen
trc050923.sas uwsatrc.sas	extrap		Macro TRC-Algorithmus Aufruf Macro TRC

tionskoeffizient blieb in der selben Grössenordnung. Die Beschränkung der totalen Gewichte nach unten dämpft die Robustifizierung und damit den Bias etwas. Aber selbst bei Beschränkung nach unten auf 1 und einer sehr grossen Abstimmungskonstanten von $c = 120$ ist die Differenz zwischen der robusten und nicht robusten Schätzung noch sehr gross (mehr als das Vierfache der Standardabweichung). Dies wurde als zu gross eingeschätzt.

Offensichtlich werden die grössten Beobachtungen (bei den mittleren Unternehmen) immer noch so stark heruntergewichtet, dass das Total zu stark beeinflusst wird. Dies kann so interpretiert werden, dass die Asymmetrie in den Daten durch den Quotientenschätzer bei den mittleren Unternehmen nicht genügend gedämpft wird. Mit anderen Worten müsste für eine gute Robustifizierung ein besseres Modell gefunden werden. Leider war dies aus Zeitmangel nicht möglich.

Tabelle 29 Ergebnisse mit robustem Quotientenschätzer

c	tw	totaus	sd	$\bar{u}$ in %	$\hat{u}$ in %
10	0	1375391	45450	97.6	98.71
10	1	1679666	55543	98.6	98.73
20	0	1555357	50441	98.8	99.75
20	1	1742157	57328	99.3	99.76
30	0	1658073	54133	99.2	99.97
30	1	1772187	58463	99.5	99.98
50	0	1793411	60920	99.6	99.99
120	1	1909017	68590	99.9	99.999

c: Abstimmungskonstante, tw: Beschränkung des totalen Gewichts nach unten, totaus: Schätzung der totalen Ausgaben, sd: Standardabweichung, $\bar{u}$: Mittelwert der Robustheitsgewichte, $\hat{u}$: Gewichteter Mittelwert der Robustheitsgewichte ($\sum_S u_i w_i / \sum_S w_i$).

6.6 Resultate und Genauigkeit

Bei der Entwicklung der Schätzverfahren zeigte sich, dass die Anwendung der in Abschnitt 5.9 entwickelten Einsetzung für die Gebühren bei den Kleinst-Unternehmen mit weniger als 20 Beschäftigten einen zu grossen Effekt hätte (Erhöhung des Gesamttotals *totaus* auf 1'016'737, bzw. um 280%). Darum wurde diese Einsetzung bei den Kleinst-Unternehmen *nicht* durchgeführt.

In Tabelle 30 sind die Resultate für die Variable *totaus*, d.h. die Summe der Ausgaben, aufgeführt. Wie in Abschnitt 6.2 ausgeführt, werden für die mittleren Unternehmen Quotientenschätzer bezüglich der Hilfsvariablen Anzahl Beschäftigte (*betot*) verwendet, während für die selbst-repräsentierenden Schichten und die Klein-Unternehmen Horvitz-Thompson Schätzer angewendet werden. Der Varianzgewinn dank dem Quotientenschätzer ist klein. Die Quotientenschätzer sollten natürlich trotzdem einen eventuellen Bias wegen Antwortausfällen verkleinern. Die Schätzungen mit Horvitz-Thompson (HT) und Quotientenschätzer (RE) weichen aber nicht stark voneinander ab. Die Variationskoeffizienten bei den mittleren Unternehmen sind zwischen 3% und 20%. Dies kann als befriedigend bezeichnet werden. Bei den selbst-repräsentierenden Schichten ist die Varianz jeweils 0 ausser in einer Wirtschaftsaktivität (Antwortausfälle). Die Variationskoeffizienten für die Kleinst-Unternehmen sind natürlich wesentlich grösser (32% und 39%). Hier müssen also die Ergebnisse mit grösster Vorsicht interpretiert werden. Der eigentlich Sinn der Erhebung bei diesen Kleinst-Unternehmen ist ja auch, die Grössenordnung des Beitrags der Kleinst-Unternehmen im Vergleich zu den Gesamtausgaben abzuschätzen.

Die Gesamtausgaben sind in Tabelle 31 aufgeführt. Die Kleinst-Unternehmen haben einen Anteil von 10.5% an den Gesamtausgaben. Die Standardabweichung der beiden Totale ist in der selben Grössenordnung, aber der Variationskoeffizient bei den mittleren und grossen Unternehmen (3%) ist natürlich wesentlich kleiner als bei den Kleinst-Unternehmen.

Die erzielten Variationskoeffizienten für die Investitionen (siehe Tabelle 32) können mit den in Tabelle 2 geplanten Variationskoeffizienten verglichen werden. Dabei ist zu beachten, dass bei der Planung weniger Aktivitäten unterschieden werden konnten. Die geplanten Genauigkeiten wurden recht gut eingehalten. Zum Beispiel wurde für den Dienstleistungsbereich (*ngrind* = 13, ..., 17) ein Variationskoeffizient von 0.13 geplant und erreicht wurde 0.125. Zu beachten sind allerdings auch die recht grossen Unterschiede in der Genauigkeit je nach Wirtschaftsaktivität und einzelne Wirtschaftsaktivitäten mit grossen Variationskoeffizienten. Z.B. hat die Holzindustrie (*ngrind*=4) einen Variationskoeffizienten von 25%. Ein grobes 95% Vertrauensintervall für die Investitionen der Holzindustrie ($\text{Total} \pm 2 \text{ mal Standardabweichung}$) würde also von 2622 bis 8146 reichen.

Für die Auswertungen nach Grössenklassen wurde eine neue Kategorisierung abgeleitet (Tabelle 33), welche die Vorgaben von Eurostat berücksichtigt. Die neue Grössenklasse *clt4*=3 ist also aus zwei Schichten aggregiert und benützt denselben Quotienten wie die Grössenklasse *clt4*=2 und, teilweise *clt4*=4. Die neue Grössenklasse *clt4*=4 aggregiert einen (synthetischen) Quotientenschätzer und einen Horvitz-Thompson Schätzer. Aggregate quer zu den Analysegruppen (Wirtschaftsaktivität) sind problemlos und die Variationskoeffizienten sind korrekt. Das Ergebnis ist in Tabelle 34. Die Summe der Schätzungen über die Grössenklassen 2, 3 und 4 entspricht dem vorherigen Total (ohne Grössenklassen). Die entsprechenden Varianzen dürfen aber jetzt nicht addiert werden. Der Variationskoeffizient, der sich bei Addition ergäbe

Tabelle 30 Totale Ausgaben nach Analysegruppen, selbst-repräsentierenden Schichten und Klein-Unternehmen

ngrind	self	klein	Ratio	HT	cv(HT)	RE	cv(RE)
1	0	0	3.4478	8524	0.1633	8171	0.1607
2	0	0	2.0610	107807	0.0696	114925	0.0722
3	0	0	1.2677	23852	0.1224	21786	0.1284
4	0	0	1.2589	16162	0.1483	15457	0.1463
5	0	0	2.1417	108812	0.1226	105885	0.1268
7	0	0	5.7412	403585	0.0930	400483	0.0744
7	1	0	3.3756	51850	0.0000	51850	0.0000
8	0	0	4.6288	62583	0.2099	63188	0.2099
9	0	0	3.1524	43485	0.1308	47620	0.1286
10	0	0	0.8106	244845	0.0401	238640	0.0349
10	1	0	0.9877	17648	0.0000	17648	0.0000
11	0	0	4.3013	93709	0.1604	96655	0.1633
12	0	0	1.0670	159690	0.0745	159895	0.0650
13	0	0	0.7399	205648	0.1196	203850	0.1095
13	1	0	0.5023	62650	0.5992	62650	0.5992
14	0	0	0.6125	59591	0.1244	60218	0.0784
15	0	0	1.4088	129755	0.2096	142551	0.2063
15	1	0	2.6772	309555	0.0000	309555	0.0000
16	0	0	0.2932	34085	0.0724	33305	0.0609
17	0	0	0.4132	79795	0.0609	77243	0.0602
17	1	0	0.1779	1000	0.0000	1000	0.0000
20	0	1	0.3455	99505	0.3177	99505	0.3177
21	0	1	0.2448	162611	0.3852	162611	0.3852

Variablen: ngrind: Wirtschaftsklasse, self: Indikator für Schicht der selbstrepräsentierenden Unternehmen, klein: Indikator für Schicht der Klein-Unternehmen, HT: Horvitz-Thompson Schätzer mit Variationskoeffizient, RE: Quotientenschätzer mit Variationskoeffizient.

ist zu klein. Der korrekte Variationskoeffizient ist in Tabelle 31 aufgeführt.

Tabelle 31 Gesamtausgaben Haupterhebung und Erhebung bei Kleinunternehmen

klein	total	sd_total	cv_total
0	2'232'574	67'916	0.03042
1	262'116	70'166	0.26769

Tabelle 32 Total der Investitionen (t_qdon_c341+t_qdon_c342)

ngrind	stotal	sd_total	cv_total
1	4414	848	0.1921
2	41732	4478	0.1073
3	6042	1267	0.2097
4	5384	1381	0.2565
5	33228	5173	0.1557
7	126891	11882	0.0936
8	42780	12425	0.2904
9	8404	1140	0.1357
10	90489	5593	0.0618
11	54671	7297	0.1335
12	56765	6587	0.1160
13	52956	13890	0.2623
14	9365	2330	0.2488
15	151730	25236	0.1663
16	4370	1241	0.2840
17	13970	2401	0.1718

Tabelle 33 Grössenklassen für Auswertung.

ngkl5	anzbe	clt4	Schätzmethode
1	0-19	1	HT
2	20-49	2	QS
3	50-99	3	QS
4	100-249	3	QS
5	250-5000	4	QS
6	5000+	4	HT

Tabelle 34 Auswertung der totalen Ausgaben nach Grössenklassen (clt4)

clt4	stotal	sd_total	cv_total
1	262116	70165.9	0.26769
2	352113	10926.37	0.03103
3	634575	19653.83	0.03097
4	1245886	46506.13	0.03733
Summe 2,3,4	2232574	—	—

Anhang

A.1 Construction of sampling weights for EPE93

The EUREDIT project used the data of the Swiss Environment Protection Expenditure Survey 1993 for development and testing of methods ([Charlton 2004](#)). The sampling weights were not registered on the original data set of EPE93. They were reconstructed in the following way: The sample design is two-way stratified. One stratification is according to economic activity (variable `actcl`) the other stratification is according to the number of employees of the enterprise in 5 classes (variable `sizecl`). The size classes are defined in table 35.

Tabelle 35 Size classes in EPE93

sizecl	number of employees
0	0-19
1	20-49
2	50-99
3	100-249
4	250+

The original sample design did not cover size class `sizecl=0` and therefore the population sizes for this size class is unknown. There are cases with unknown economic activity or size class. The population sizes have been taken from ([INFRAS 1996](#), page 190/191). However, industry 11 (ENERGIE) was not documented and the number of enterprises per size class was taken from the enterprise census of 1991. Where the stratification is known the weight is np/ns . Where only industry is known, i.e. for size class 0, we imputed an overall weight of the activity class. Thus the weight for size class 0 is the population size of the activity class divided by the total sample in the activity class. Where neither activity class nor size class is known (2 cases) the weight is set to 1. The sum of the weights is 19'968.53. We cannot calibrate this number to a population number since the population size is not documented in ([INFRAS 1996](#)). More precisely, the population in size class 0 is unknown. We could calibrate on the number of employees in the population 2'615'842.92 by calibrating the estimated number of employees on this figure. The estimated number of employees in the population is 2'592'536. Thus an overall correction factor of 1.00899 would result. Since it is close to 1 we did not apply this correction factor to the weights.

Since there are a lot of outliers in the data an ad hoc robustified ratio estimator has been used in ([INFRAS 1996](#)). Furthermore there are lots of zero values in the data. Therefore Hulliger and Kassab ([Hulliger and Kassab 1998](#)) developed robust estimators for the parallel survey on environment protection expenditures of municipalities. Other estimation procedures might be more appropriate.

A.2 Wirtschaftgruppen und Populationsgrößen

Die Umschlüsselung der Wirtschaftsaktivitäten der Erhebung über Umweltschutzausgaben 1993 auf die NOGA-Klassifikation ist in Tabelle A.2 aufgeführt. Diese Umschlüsselung ist ad-

hoc für die Erstellung des Stichprobenplans erstellt worden.

Tabelle 36 Eins-zu-eins Korrespondenz ASWZ85 zu NOGA95 auf Zweisteller

ASWZ2	NOGA2	ASWZ2	NOGA2
11	40	51	51
21	15	52	51
22	15	55	52
23	16	56	52
24	17	57	55
25	18	58	50
26	20	62	60
27	21	65	63
28	22	71	65
31	24	72	66
32	25	73	70
33	14	75	72
34	27	76	93
35	29	81	80
36	31	83	85
37	33	85	85
38	36	86	92
41	45	87	91
42	45	88	92

Die gewünschte Einteilung der Wirtschaftsaktivitäten in Gruppen für die Publikation ist in Tabelle [A.2](#) aufgeführt. Tabelle [38](#) zeigt die Vergrößerung, die für die Erstellung des Stichprobenplans verwendet wurde.

A.3 Programme

Die meisten Programme sind auf Anfrage erhältlich. Das Programm spezialkorr kann aus Datenschutzgründen nicht zur Verfügung gestellt werden.

5 propositions d'agrégation des branches NOGA en 7 à 17 groupements

Cinq ventilation NOGA envisageables - AC, AD, AE, BD, BE - en fonction du niveau de détail des données diffusées, à savoir par exemple : agrégats globaux, désagrégation par variable économique, par variable économique et domaine d'environnement, par domaine d'environnement et classe de taille, etc.

Note : Le groupement A du secteur II ainsi que les 3 classes de taille retenues correspondent aux exigences du questionnaire OCDE/Eurostat.

			Classes taille (nbre pers. occupées)			<- nbre d'entreprise en 2001	
Groupement			1 à 49	50 à 249	plus de 250		
Secteur II (10-45, sauf 37)	A	B	72'472	2'690	480	A priori au moins 70% des dépenses devraient se trouver ici.	
10 Extraction de houille, lignite, tourbe	1	1	1	-	-		
11 Extraction hydrocarbures, s. ann.			1	-	-		
14 Autres industries extractives			324	11	1		
15 Industrie alimentaires et boissons	2	2	2'420	154	39		
16 Industrie du tabac			6	3	3		
17 Industrie textile	635		69	4			
18 Industrie de l'habillement et fourrures	924		15	2			
19 Industrie du cuir, de la chaussure	280		7	-			
20 Travail du bois, fabrication d'articles en bois	6'446		54	5			
21 Industrie du papier et du carton	173		39	20			
22 Edition, impression, reproduction	5		4'380	144	22		
23 Cokéfaction	6		9	2	1		
24 Industrie chimique	7	3	791	134	46		
25 Fabrication d'articles caoutchouc, plastique			703	102	15		
26 Fabric. d'autres produits minéraux non métal.	8		1'326	62	11		
27 Métallurgie	9	209	45	18			
28 Travail des métaux	10	4	7'712	268	18		
29 Fabrication de machines et équipements			2'986	303	75		
30 Fabr. machines de bureau, équip. inform.			127	8	2		
31 Fabrication mach., appareils électriques nca			887	94	33		
32 Fabr. équipements radio-TV, communic.			585	53	18		
33 Fabrication instrum. de précision/horlogerie			3'095	228	50		
34 Industrie automobile			156	16	4		
35 Fabrication autres moyens de transport			369	21	8		
36 Fabrication de meubles, ind. diverses			3'751	83	9		
37 Récup., prép. au recyclage							
40 Production électricité, gaz et eau	11		5	278	61	17	
41 Captage, distribution d'eau		94		2	2		
45 Construction	12	6	33'804	712	57		

Secteur III (50-64)	C	D	E	183'174	2'193	444	<- nbre d'entreprise en 2001
50 Commerce d'auto, réparation, carburants	1	1	1	14'462	107	13	A priori 20% des dépenses environ
51 Commerce de gros et intermédiaires				17'911	488	61	
52 Commerce de détail				38'274	226	80	
55 Hôtellerie et restauration	2			24'719	317	30	
60 Transports terrestres et par conduites	3			7'186	158	19	-> se limiter aux dépenses des entreprises de + de 50 emplois ?
61 Transports par eau				84	9	-	
62 Transports aériens				122	11	3	
63 S. auxiliaires de transports				2'790	83	30	
64 Postes et télécommunications	4			2	561	29	15
65 Intermédiation financière		1'380			125	65	
66 Assurances		253			40	32	
67 Services auxiliaires financiers et assurances		3'504			30	5	
70 Activités immobilières		4'124			34	3	-> se limiter aux dépenses des entreprises de + de 50 voire de 250 emplois ?
71 Location de machines et d'équipements		873			10	1	
72 Activités informatiques		10'998			122	23	
73 Recherche et développement		447			31	11	
74 Autres services fournis aux entreprises				55'486	373	53	
75 Adm. publ., défense, social							
80 Enseignement							
85 Santé et activités sociales							
90 Assainissement, voirie							
91 Activités associatives							
92 Activités récréatives, culturelles et sportives							
93 Services personnels							

Abbildung 6 Vorschlag für Bildung von Wirtschaftsgruppen

Tabelle 37 Gruppierung der NOGA2 in die 17 von Eurostat vorgeschlagenen Analysegruppen.

NOGA2 Text	noga2	gruppe.a	USchA-Intensität
10 Extraction de houille, lignite, tourbe	10	1	3
11 Extraction hydrocarbures, s. ann.	11	1	3
14 Autres industries extractives	14	1	3
15 Industrie alimentaires et boissons	15	2	3
16 Industrie du tabac	16	2	3
17 Industrie textile	17	3	3
18 Industrie de l'habillement et fourrures	18	3	3
19 Industrie du cuir, de la chaussure	19	3	3
20 Travail du bois, fabrication d'articles en bois	20	4	3
21 Industrie du papier et du carton	21	5	3
22 Edition, impression, reproduction	22	5	3
23 Cokéfaction	23	6	3
24 Industrie chimique	24	7	3
25 Fabrication d'articles caoutchouc, plastique	25	7	3
26 Fabric. d'autres produits minéraux non métall.	26	8	3
27 Métallurgie	27	9	3
28 Travail des métaux	28	10	3
29 Fabrication de machines et équipements	29	10	3
30 Fabr. machines de bureau, équip. inform.	30	10	3
31 Fabrication mach., appareils électriques nca	31	10	3
32 Fabr. équipements radio-TV, communic.	32	10	3
33 Fabrication instrum. de précision/horlogerie	33	10	3
34 Industrie automobile	34	10	3
35 Fabrication autres moyens de transport	35	10	3
36 Fabrication de meubles, ind. diverses	36	10	3
40 Production électricité, gaz et eau	40	11	3
41 Captage, distribution d'eau	41	11	3
45 Construction	45	12	3
50 Commerce d'auto, réparation, carburants	50	13	2
51 Commerce de gros et intermédiaires	51	13	2
52 Commerce de détail	52	13	2
55 Hôtellerie et restauration	55	14	2
60 Transports terrestres et par conduites	60	15	2
61 Transports par eau	61	15	2
62 Transports aériens	62	15	2
63 S. auxiliaires de transports	63	15	2
64 Postes et télécommunications	64	15	2
65 Intermédiation financière	65	16	1
66 Assurances	66	16	1
67 Services auxiliaires financiers et assurances	67	16	1
70 Activités immobilières	70	17	1
71 Location de machines et d'équipements	71	17	1
72 Activités informatiques	72	17	1
73 Recherche et développement	73	17	1
74 Autres services fournis aux entreprises	74	17	1

Tabelle 38 Vergrößerung von Gruppierung ac nach bd.

gruppe . ac	gruppe . bd
1	1
2	2
3	2
4	2
5	2
6	3
7	3
8	4
9	4
10	4
11	5
12	6
13	7
14	7
15	7
16	8
17	8

Bei der Vergrößerung auf 7 Klassen (gruppe . bdp) wurde gruppe . bd=5 zu gruppe . bd=1 geschlagen.

Tabelle 39 Stichprobenrahmen: Populationsgrößen

noga2	gkl5					All
	1	2	3	4	5	
10	1	.	.	.	.	1
11	1	.	.	.	.	1
14	250	52	11	2	1	316
15	2172	205	86	71	41	2575
16	5	1	1	3	3	13
17	537	72	36	30	6	681
18	844	37	10	5	2	898
19	238	14	4	2	1	259
20	6271	243	47	9	3	6573
21	121	40	15	25	19	220
22	3917	280	98	52	24	4371
23	7	1	.	2	1	11
24	629	125	78	58	48	938
25	558	123	61	42	15	799
26	1205	106	41	26	9	1387
27	165	33	19	22	19	258
28	7153	637	183	97	19	8089
29	2626	394	170	140	77	3407
30	98	24	7	2	2	133
31	794	103	55	37	32	1021
32	509	65	25	29	16	644
33	2866	228	139	91	50	3374
34	132	26	9	7	3	177
35	338	19	9	8	7	381
36	3613	171	52	32	10	3878
40	188	85	39	26	18	356
41	38	5	2	.	1	46
45	33102	2004	484	214	53	35857
50	14454	364	79	35	13	14945
51	18021	1012	328	180	68	19609
52	38474	941	213	100	77	39805
55	24833	1251	267	113	40	26504
60	7192	381	116	58	22	7769
61	62	19	7	5	.	93
62	105	18	5	6	2	136
63	2687	164	60	34	33	2978
64	501	53	14	9	25	602
65	1225	165	70	50	73	1583
66	246	69	13	19	35	382
67	3383	149	24	8	5	3569
70	4295	110	34	13	6	4458
71	838	19	7	5	1	870
72	11729	311	81	35	21	12177
73	462	37	18	11	12	540
74	58648	1096	276	185	73	60278
All	255533	11252	3293	1898	986	272962

Tabelle 40 Programme und Daten für Rahmen und Ziehung (SAS)

Prog/Dat	Name	Beschreibung
p	Rahmen.sas	Erstellt den provisorischen Stichprobenrahmen
d	week_ent040222.sas7bdat	BUR-Auszug Unternehmen vom 22.2.04 für Stichprobenplanung
p	Ziehung.sas	Definitiver Stichprobenrahmen und Ziehung
d	week_ent040322.sas7bdat	BUR-Auszug Unternehmen vom 22.3.04 für Ziehung
p	plan.sas	Erstellt die Datei basestrat.xls mit den Angaben über die Population (Input für S-Plus)
d	basestrat040326.xls	Schichtindikatoren und Größen (Variablen: epeg, gkl5, nph=_FREQ_, sumbetot. Input für Stichprobanplan in S-Plus.)
d	gruppen.sas7bdat	Definition der 17 Analysegruppen mit NOGA2
p	listast.sas	Erstellt die Liste der Arbeitsstätten zu den gezogenen Unternehmen
d	week_loc040322.sas7bdat	BUR-Auszug Arbeitsstätten vom 22.3.04 für Ziehung
d	plan040329.sas7bdat	Stichprobenraten und Schicht (Input für SURVEYSELECT)
d	sample040330.sas7bdat	Stichprobe (Unternehmen)
d	listast.sas7bdat	Stichprobe (Arbeitsstätten)

Die Dateien sind in den Archiven progdat040330.zip und entloc040322.zip gespeichert.

Tabelle 41 Programme für die Aufarbeitung der Daten

Programm	Input	Output	Beschreibung
import	Access-Datenbank und Excel-Dateien (gruppenezusatz, atotbetot)	SAS-Dateien	Einlese-Programm.
gruppen-brutto	info; bur.alt; (bur.neu;) netto; sample040330	gruppenb; pistrat	Gruppeninformationen aus der Datenbank und Zusätze zusammenführen
rahmen2	rahmen2; gruppenb;	rahmen3; nstrat	Aufdatierung Rahmen bez. UG, Berechnung Einschlusswk. der Gruppen, Zwischendateien für die UG, Schichttotale herausschreiben
gruppen-netto	rahmen3; definitiv	netto	
spezialkorr	netto; atot2betot	netto2	Spezialkorrekturen aufgrund Abklärungen über Unternehmen mit grossen Unterschieden zwischen atot und betot und aufgrund Kontrolle Kilo-Fehler (Unitésfausses2)
kontrollen	netto2 oder eingesetzt	kontrolliert oder eingesetzt	Die Kontrollen können ab den Rohdaten oder nach den Einsetzungen durchgeführt werden
modell-geb	kontrolliert		Robuster Schätzung des Quotienten Gebühren zu Beschäftigte (-> einsetzungen-geb-tot)
einsetzungen-geb-bes-tot	kontrolliert	eingesetzt	Einsetzungen monetär

Die Programme sind in der Datei UWSA_Programme_061130.zip gespeichert

Tabelle 42 Programme und Daten für Stichprobenplan

Prog/Dat	Name	Beschreibung
f	weightsandstrata.ssc	Funktionen für geschichtete Stichproben mit fehlenden Werten
s	noga2undrecodes.ssc	Einlesen UschA93 Daten und Recodierungen
d	uscha93totale040303.xls	Auszug aus USchA93 Daten (Gewichtung aus Euredit)
s	splan_ac.SSC	Stichprobenplan auf 17 Gruppen basierend
f	power.ssc	Zuteilung auf Schichten proportional zu einer Variablen
f	meanwgalpha.ssc	Winsorisiertes Mittel
s	anova.SSC	Varianzanalyse der Variationskoeffizienten
d	plan040329.xls	Stichprobenraten und Schichten (Input für ziehung.sas)

f: S-Plus Funktion, s: S-Plus Skript, d: Daten



Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

OFS BFS UST

Sektion UWN / Espace de l'Europe 10 / 2010 Neuchâtel

FRAGEBOGEN

ERHEBUNG ÜBER DIE UMWELTSCHUTZAUSGABEN DER UNTERNEHMEN 2003

Für Auskünfte stehen Herr Jacques Roduit und seine Kollegen

Ihnen gerne zur Verfügung:

deutsch: 032 713 64 72

französisch: 032 713 66 12

italienisch: 032 713 67 80 oder unter umwelt@bfs.admin.ch.



Füllen Sie den Fragebogen von Hand aus, mit einem schwarzen oder dunkelblauen Kugelschreiber. Lassen Sie **keine Felder leer** und geben Sie das **Fehlen von Ausgaben mit einer «0»** (Null) an.

5 7 8 2



**Bitte lesen Sie
die Erläuterungen im Anhang**

Bitte bis zum **21. Mai 2004** im beigelegten Antwortcouvert oder an oben stehende Adresse zurücksenden.

A Allgemeine Informationen

Die vorliegende Testerhebung bezieht sich auf das Geschäftsjahr **2003** (Abschluss zwischen 01.05.2003 und 30.04.2004).

Die Erhebungseinheit ist das Unternehmen gemäss Betriebs- und Unternehmensregister (BUR) des BFS. Bitte kontrollieren und, sofern nötig, korrigieren Sie (Ergänzungen/Streichungen) den beigelegten Auszug aus dem BUR und senden Sie ihn mit dem Fragebogen zurück.

Bitte beantworten Sie den vorliegenden Fragebogen **für alle Betriebe Ihres Unternehmens**, die auf dem oben erwähnten BUR-Auszug aufgeführt sind.

Lassen Sie **keine Felder leer** und geben Sie das **Fehlen von Ausgaben mit einer «0»** (Null) an. Falls die Beantwortung Schwierigkeiten bereitet, setzen Sie bitte eine bestmögliche Schätzung ein.

Die Antworten von Unternehmen, die wenige oder keine Umweltschutzausgaben tätigen, sind für die Qualität der Ergebnisse dieser Erhebung ebenfalls sehr wichtig.

Auskunftsperson in Ihrem Unternehmen für zusätzliche Informationen

Name A101

Vorname A102

E-Mail A103

Telefon A104

Im Unternehmen beschäftigte Personen am 30. September 2003 (inkl. Inhaber/innen, Geschäftsführer/innen und Lehrlinge)

Mit 90% und mehr der betriebsüblichen Arbeitszeit

A105

Mit 50 bis 89% der betriebsüblichen Arbeitszeit

A106

Unter 50% der betriebsüblichen Arbeitszeit

A107

B Umweltschutzausgaben 2003

Ihre Ausgaben zum Schutz der Umwelt vor Auswirkungen Ihrer Betriebstätigkeit können 2003 aus Investitionen in Installationen oder Anlagen, aus betriebsinternen laufenden Kosten, aus Kosten für den Beizug externer Dienstleister oder aus der Bezahlung kommunaler Gebühren bestanden haben (⇒ nähere Einzelheiten in den Erläuterungen).

Haben Sie 2003 derartige Ausgaben zur Behandlung oder Verhinderung von Umweltverschmutzung durch Ihr Unternehmen getätigt?

B201 ☐ Wenn Ja, füllen Sie bitte **unter Beachtung der beiliegenden Erläuterungen** die Teile C und D des Fragebogens aus.

B202 ☐ Wenn Nein, füllen Sie bitte Teil D des Fragebogens aus und senden Sie diesen an uns zurück.

Abbildung 7 Fragebogen Seite 1

C Auflistung der Umweltschutzausgaben 2003

Für die Angabe Ihrer Ausgaben im Jahr 2003 in untenstehender Tabelle stützen Sie sich bitte auf die Definitionen, die Listen mit Beispielen und das Hilfsschema in den beigelegten **Erläuterungen**.

Bitte lassen Sie **keine Felder leer** und geben Sie das **Fehlen von Ausgaben mit einer «0»** (Null) an. Falls die Beantwortung Schwierigkeiten bereitet, setzen Sie bitte eine bestmögliche Schätzung ein.

Beträge in tausend Franken , ohne MWSt (keine Ausgabe = 0)					
	Luftreinhaltung und Klimaschutz	Gewässerschutz	Abfallbewirt- schaftung	Andere Umweltbereiche *	Total 2003
«End-of-Pipe»-Investitionen (Behandlung der Umwelt- verschmutzung)	C301	C311	C321	C331	C341
Integrierte Investitionen (Vermeidung der Umwelt- verschmutzung)	C302	C312	C322	C332	C342
Laufende interne Ausgaben	C303	C313	C323	C333	C343
Einkauf von Dienstleistungen bei Dritten (inkl. kommunale Gebühren)	C304	C314	C324	C334	C344
⇒ davon kommunale Gebühren		C315	C325		
Einnahmen aus Nebenprodukten (durch Ihre Umweltschutz- aktivitäten generiert)	C306	C316	C326	C336	C346

* Lärmschutz, Natur- und Landschaftsschutz, Bodenschutz, Grundwasserschutz, Forschung & Entwicklung, übrige Tätigkeiten.

Beträge, die mehr als einen Umweltbereich betreffen, sind - notfalls mittels einer Schätzung - auf die vier angegebenen Umweltbereiche aufzuteilen.

Anzahl Personen, deren Entlohnung teilweise oder vollständig in den oben erwähnten laufenden internen Ausgaben verrechnet wurde (z.B. 4,3 Personen)	C351
---	--

	Ja	Nein	Teilweise
Ist Ihr Unternehmen umweltzertifiziert (z.B. ISO 14000, EMAS usw.) ?	C361	C362	C363

D Bemerkungen zur Erhebung, zum Fragebogen und zu den Erläuterungen

D401

Zur Beantwortung des Fragebogens insgesamt aufgewendete Zeit	Stunden D402	Minuten D403
--	--	--

Wir danken Ihnen herzlich für Ihre wertvolle Mitarbeit!

Literatur

- Béguin, C. and B. Hulliger (2003). *Robust Multivariate Outlier Detection and Imputation with Incomplete Survey Data*. EUREDIT.
- Béguin, C. and B. Hulliger (2004). Multivariate outlier detection in incomplete survey data: The epidemic algorithm and transformed rank correlations. *J.R.Statist.Soc.A* 167(Part 2.), 275–294.
- BFS (1996). *Umweltausgaben und -investitionen in der Schweiz 1992/93, Ergebnisse einer Pilotstudie*. Bundesamt für Statistik.
- BFS (2005). *Umweltschutzausgaben der Unternehmen 2003, Erste Ergebnisse*. Bundesamt für Statistik, BFS Aktuell.
- BFS (2006). *Umweltschutzausgaben der Unternehmen, Ergebnisse der Erhebung 2003*. Bundesamt für Statistik.
- Charlton, J. (Ed.) (2004). *Methods and Experimental Results from the Euredit Project*. <http://www.cs.york.ac.uk/euredit/results/results.html>.
- Hulliger, B. (1999). Simple and robust estimators for sampling. In *Proceedings of the Section on Survey Research Methods*, pp. 54–63. American Statistical Association.
- Hulliger, B. and M. Kassab (1998). *Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes*. Swiss Federal Statistical Office.
- INFRAS (1996). *Umweltschutzausgaben in der Schweiz, Technischer Bericht zur Piloterhebung*. Bundesamt für Statistik.

Methodenberichte des Dienstes Statistische Methoden des BFS

Rapports de méthodes du Service de méthodes statistiques de l'OFS

Methodological reports of the Statistical Methods Unit of SFSO

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 Datenaufbereitung 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 Lohnindizes 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 élèves (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:

<i>Diffusionsmittel</i>	<i>Kontakt N° à composer</i>	<i>Moyen de diffusion</i>
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.

Der Stichprobenplan für die Erhebung über Umweltschutzausgaben der Unternehmen 2003 wurde mit Hilfe der Daten der entsprechenden Piloterhebung aus dem Jahr 1993 erstellt. Daten für Kleinst-Unternehmen wurden mit einer Kontrollstichprobe erhoben. Bei der Aufbereitung der Daten wurden die Daten von Unternehmensgruppen konsolidiert, Inkonsistenzen kontrolliert und, so weit möglich, bereinigt und fehlende Werte eingesetzt. Die Schätzverfahren benützen eine spezielle Schicht der grössten Unternehmen und berücksichtigen die Antwortausfälle.