Methodenbericht

Stichprobenplan für die Obstbaumzählung 2001

0

Statistische Grundlagen und Übersichten
Bases statistiques et vues d'ensemble
Basi statistiche e presentazioni generali

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

OFS BFS UST

Neuchâtel, 2002

Statistik der Schweiz

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

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geldpolitik, Finanzmärkte und -akteure
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Medien, Zeitverwendung
17 Politik
18 Öffentliche Finanzen
19 Rechtspflege
20 Gesellschaft in Bewegung (Querschnittsthemen)

Statistique de la Suisse

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

0 Bases statistiques et vues d'ensemble
1 Population
2 Espace et environnement
3 Travail et vie active
4 Economie nationale
5 Prix
6 Industrie et services
7 Agriculture et sylviculture
8 Energie
9 Construction et logement
10 Tourisme
11 Transports et communications
12 Politique monétaire, marchés et acteurs financiers
13 Protection sociale
14 Santé
15 Education et science
16 Culture, médias, emploi du temps
17 Politique
18 Finances publiques
19 Droit et justice
20 Société en mutation (thèmes interdisciplinaires)

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Stichprobenerhebung für die Obstbaumzählung 2001

Autoren **Daniel Kilchmann und Beat Hulliger**
Bundesamt für Statistik

Herausgeber **Bundesamt für Statistik**

OFS BFS UST

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

Neuchâtel, 2002

Vorwort

In den landwirtschaftlichen Betriebszählungen und in den Erhebungen über die Direktzahlungen an die Landwirtschaft werden viele Informationen über die Anzahl der Obstbäume in der Schweiz erhoben. Es stellt sich die Frage, ob diese Erhebungen anstatt einer speziellen Obstbaumzählung zur Schätzung der Obstbäume verwendet werden könnten. Dabei müsste die Anzahl der von diesen Erhebungen nicht erfassten Bäume mit einer Stichprobenerhebung bei den Gemeinden geschätzt werden. Die Sektion Land- und Forstwirtschaft des BFS hat den Dienst Statistische Methoden beauftragt, die Methoden und die Genauigkeit einer Stichprobenerhebung bei 200 Gemeinden abzuklären. Die Arbeiten wurden durch Daniel Kilchmann und Beat Hulliger vom Dienst Statistische Methoden ausgeführt und von Arthur Zesiger von der Sektion Land- und Forstwirtschaft begleitet. Der Auftrag musste leider frühzeitig zurückgezogen werden und demzufolge konnten nicht alle vorgesehenen Arbeiten abgeschlossen werden. Trotzdem gibt der vorliegende Bericht eine Grundlage, um eventuell in einem späteren Zeitpunkt die Arbeiten wieder aufzunehmen.

Zusammenfassung

Bis anhin wurden die Obstbäume anhand einer Vollerhebung erhoben. Die Resultate wurden jeweils nach Kantonen und Fruchtart aufgeteilt publiziert. Für die Obstbaumzählung 2001 wurde entschieden, nur noch in den landwirtschaftlichen Betrieben eine Vollerhebung durchzuführen. Für die nicht landwirtschaftlichen Betriebe war eine Stichprobenerhebung vorgesehen. Die Unterscheidung zwischen landwirtschaftlichen und nicht landwirtschaftlichen Betrieben wurde in den vorhergehenden Erhebungen nicht gemacht. Darum wurden zuerst verschiedene Modelle für die Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben entwickelt. Die Prognose wurde verwendet, um die Güte eines Quotientenschätzers der Anzahl Bäume zu beurteilen. Der Quotientenschätzer stützt sich auf die Korrelation der Anzahl Bäume auf nicht landwirtschaftlichen und landwirtschaftlichen Betrieben. Die Stichprobe der Gemeinden wurde nach Kantonen oder Gruppierungen von Kantonen (z.B. Grossregionen NUTS2) geschichtet. Das Ziel war, in sämtlichen Schichten einen Variationskoeffizient von weniger als 5% zu erreichen. Die Analyse der Schätzungen zeigt, dass eine Stichprobengrösse von 200 Gemeinden nicht ausreicht, um die Anzahl Bäume in nicht landwirtschaftlichen Betrieben nach Kantonen aufgeteilt mit genügender Genauigkeit zu schätzen. Sofern die Schichtung beibehalten wird, muss die Stichprobengrösse auf rund 750 Gemeinden erhöht werden. Ein gangbarer Mittelweg wäre vermutlich eine moderate Erhöhung der Stichprobe zusammen mit einer Reduktion der Ansprüche an die Genauigkeit in einigen Kantonen.

Schlüsselwörter

Methodenbericht; Obstbaumzählung; OZ2001; Stichprobe; Schichtung; Schätzung; Quotientenschätzer; Variationskoeffizient.

Auskunft: Daniel Kilchmann, Tel. 032 713 60 84
Daniel.Kilchmann@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: 338-0008

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Grafik/Layout: BFS

Copyright: BFS, Neuchâtel 2002
Abdruck – ausser für kommerzielle Nutzung –
unter Angabe der Quelle gestattet.

ISBN: 3-303-00241-X

Inhaltsverzeichnis

1 Einführung **5**

2 Grundgesamtheit und Datenquellen **6**
2.1 Datenqualität 7

3 Schichtung **7**

4 Schätzer **8**
4.1 Definitionen 8
4.2 Schätzung anhand einer Stichprobe 9
4.3 Quotientenschätzer 10
4.4 Variationskoeffizient 11
4.4.1 Quotientenschätzer vs. Schätzer mit geschichtetem Mittel 12

5 Stichprobenplan **12**

6 Zuordnung **12**
6.1 *Neyman*-Zuordnung 13
6.2 Proportionale Zuordnung 13
6.2.1 Proportionalität zur Anzahl Bäume 15
6.2.2 Proportionalität zur landwirtschaftlichen Nutzfläche 15

7 Schätzung mit geschichtetem Mittel **15**
7.1 Schichtung nach Kantonen 16
7.2 Schichtung nach Kantonsgruppen 16
7.2.1 Kantons Gruppierung 16
7.2.2 NUTS2 18

8 Schätzung anhand eines Quotientenschätzers **20**
8.1 Grundgesamtheit 22
8.2 Schätzer 23
8.2.1 Definitionen 23
8.2.2 Quotientenschätzer 24
8.2.3 Variationskoeffizient 25
8.3 Prognosen für die OZ2001 26
8.3.1 Prognose anhand der Daten der OZ81, der OZ91 und der DZ99 . . . 26
8.3.2 Prognose anhand der Daten der OZ51 bis OZ91 27
8.3.3 Begrenzte Prognose 29

9 Schlussfolgerung **33**

Anhang **37**

A Variablenliste **37**

B Graphiken **38**
B.1 Prognose anhand der OZ81, der OZ91 und der DZ99 38
B.2 Prognose anhand der OZ51 bis OZ91 40
B.3 Begrenzte Prognose 40

C S-PLUS Programme **42**
C.1 Stichprobenplan Funktionen 42
C.2 Erarbeitung des Stichprobenplans 44

Literatur **48**

1 Einführung

Seit 1951 führt das BFS in Abständen von 10 Jahren eine eidg. Obstbaumzählung durch. Die letzte Erhebung dieser Art erfolgte 1991. Bis anhin wurde eine Vollerhebung durchgeführt, wobei die Obstbäume in zwei Grössenklassen unterschieden wurden: die Niederstammbäume (Stammhöhe bis 1.0m) und die Halb- und Hochstammbäume (Halbstammbäume: beim Kernobst ist die Stammhöhe zwischen 1.0m und 1.6m und beim Steinobst 1.0m bis 1.2m). In der OZ2001 werden hingegen nur noch die Hochstammbäume verlangt, wobei aus Gründen der Vergleichbarkeit mit der OZ91 auch die Halbstammbäume erhoben werden müssen, da diese in der OZ91 gemeinsam mit den Hochstammbäumen erhoben wurden. Beim Steinobst wird in der OZ2001 der Messbarkeit wegen darauf verzichtet, einen Unterschied zwischen Halb- und Hochstammbäumen zu machen. Die erhobenen Bäume werden zusätzlich noch nach den Baumarten unterschieden (c.f. BFS, Eidg. Alkoholverwaltung [1]).
Um Kosten zu sparen, war für die OZ2001 vorgesehen, nur in landwirtschaftlichen Betrieben eine Vollerhebung durchzuführen und in den nicht landwirtschaftlichen Betrieben eine Stichprobe zu ziehen. Die Vollerhebung sollte gleichzeitig mit der Erhebung der Direktzahlungen 2001 (DZ01) erfolgen, welche alle zwei Jahre durchgeführt wird. In den früheren Erhebungen der Direktzahlungen des Bundes wurden nur die Anzahl Hochstammbäume erhoben, nicht aber Halbstammbäume und auch die verschiedenen Baumarten wurden nicht erhoben. Um Direktzahlungen zu erhalten, müssen Landwirtschaftsbetriebe (landwirtschaftliche Nutzfläche $>$ 1ha) mehrere Bedingungen erfüllen. Unter anderem müssen sie ein Teil ihrer Betriebsfläche (7% im Talgebiet, 4% im Berggebiet) als ökologische Ausgleichsflächen ausweisen. Pro Baum wird dem Landwirt eine Are ökologische Ausgleichsfläche angerechnet. Zudem erhält er vom Bund einen Beitrag zur Hochstammförderung. Dies ist einer der Gründe die dafür sprechen, die Anzahl Hochstammbäume zu erheben.
Ein weiterer Teil der Bäume, der bisher nicht erwähnt wurde, betrifft die Obstkulturen, welche jedoch vollumfänglich anhand der jährlichen Erhebung des BWL erfasst werden, und somit keinen Einfluss auf die Stichprobenerhebung haben.
Mit dem Ziel, das Total der Hochstammbäume zu schätzen, wurde von der Sektion AGR (J.-F. Fracheboud, A. Zesiger) der Auftrag an METH gegeben, einen Stichprobenplan für die OZ2001 auszuarbeiten. Dabei wurde die Vorgabe gegeben, dass die gesamte Stichprobengrösse 200 Gemeinden umfassen sollte, wobei Schätzwerte auf Kantonsebene gefragt wurden. METH versuchte einen Stichprobenplan zu erarbeiten, der für sämtliche Schätzer einen maximalen Variationskoeffizienten von 5% ergeben sollte.
Das Projekt Stichprobenerhebung der Obstbaumzählung 2001 wurde jedoch noch vor Beendigung des Stichprobenplans abgebrochen. Dies ist der Grund, weshalb wir nicht alle Einzelheiten vollumfänglich untersucht haben.

Nachfolgend werden die Grundgesamtheit und die Schätzer kurz beschrieben. Darauf folgen zwei Methoden zur Schätzung des Totals der Obstbäume, eine einfache Schätzung und eine Schätzung anhand eines Quotientenschätzers. Das Total wurde dabei auf Kantonsebene oder Gruppierungen von Kantonen (z.B. Grossregionen der Schweiz: NUTS2) geschätzt.

2 Grundgesamtheit und Datenquellen

In der Tabelle 1 präsentieren wir einen Überblick über die verschiedenen Erhebungen und ihren Eigenschaften. Die Analysedaten der OZ2001 setzen sich aus den Daten der Vollerhebung der OZ2001 und den hochgerechneten Daten der Stichprobe OZ2001 zusammen. In den Gemeinden, welche in der Stichprobe gezogen werden, wird eine Vollerhebung durchgeführt. Somit werden deren landwirtschaftliche Betriebe doppelt erhoben. Es muss bemerkt werden, dass in früheren Erhebungen die landwirtschaftlichen und die nicht landwirtschaftlichen Betriebe nicht unterschieden wurden. Deshalb existieren keinerlei Hilfsdaten zur Schätzung der Anzahl Bäume in nicht landwirtschaftlichen Betrieben.
Die Grundgesamtheit besteht aus den Gemeinden die in die OZ2001 miteinbezogen werden. Die Schätzwerte, die anhand der Stichprobe berechnet werden, beziehen sich immer auf die Grundgesamtheit oder Schichten der Grundgesamtheit (falls eine Schichtung vorgenommen wird). Wir haben versucht die Daten maximal auszuschöpfen, deshalb hängt die Grundgesamtheit von der Schätzmethode ab, die wir untersucht haben.
Die Stichprobengrösse wurde auf 200 Gemeinden begrenzt, in welchen jeweils eine Vollerhebung der Halb- und Hochstammbäume durchgeführt werden sollte. Im weiteren wurden aussagekräftige Resultate auf Kantonsebene angestrebt. Gemeinden wurden in die OZ2001 aufgenommen, wenn ein Teil ihrer Obstbäume in einer der folgenden Erhebungen bereits erhoben wurde :

1. Eidg. Landwirtschafts-Betriebszählung 2000 (LBZ2000)
2. Erhebung der landwirtschaftlchen Direktzahlungen 1999 (DZ99)
3. Obstbaumzählung 1991 (OZ91)
4. Bodennutzungsstatistik 1997 (Erhebung mit Luftbildern).

Bereits aus der OZ91 wurden Gemeinden mit dem gesamten Gemeindegebiet über 1000 M.ü.M. ausgeschlossen. Von diesen Gemeinden wurden jedoch einige in der OZ2001 wieder aufgenommen, da sie das Kriterium 1 oder 2 erfüllen. Die Sektion AGR lieferte uns somit die Daten von 2911 Gemeinden (resp. 2908 Gemeinden in einer zweiten Datenlieferung), die den obigen Kriterien entsprechen.
Als Identifikatoren der Gemeinden standen uns die Gemeindenummer sowie der Name der Gemeinde und der Gemeindetyp zur Verfügung. Ausserdem hatten wir Identifikatoren für den Bezirk und den Kanton. Weiter wurden in einer ersten Datenlieferung die Baumbestände der OZ91 und der DZ99 pro Gemeinde geliefert sowie die Anzahl landwirtschaftlicher Betriebe pro Gemeinde. In einer zweiten Lieferung wurden die Baumbestände der OZ81 sowie eine Prognose der OZ2001 aufgrund der OZ81 und der OZ91 geliefert. Weitere Variablen waren die landwirtschaftliche Nutzfläche sowie die Fläche von nicht landwirtschaftlichen Wiesen und Weiden und die Gemeindefläche. Die Daten der BZ2000 konnten nicht geliefert werden, da deren Erhebung bis im Oktober 2000 noch nicht abgeschlossen war. Die vollständige Variablenliste ist in Anhang A aufgeführt.

Tabelle 1: Darstellung der verschiedenen Erhebungsarten der Anzahl Bäume in Erhebungen, aufgeteilt in die Art der Betriebe. Ein "X"bedeutet, dass keine Daten in den entsprechenden Betrieben erhoben wurden/werden.

	Landwirtsch. Betriebe mit DZ	Landwirtsch. Betriebe ohne DZ	Nicht landw. Betriebe
OZ81	Vollerhebung	Vollerhebung	Vollerhebung
OZ91	Vollerhebung	Vollerhebung	Vollerhebung
DZ99	Vollerhebung	X	X
Vollerhebung OZ2001	Vollerhebung	Vollerhebung	X
Stichprobe OZ2001	Stichprobe	Stichprobe	Stichprobe

2.1 Datenqualität

Es hat sich herausgestellt, dass die Daten, welche geliefert wurden, nicht mit den Daten übereinstimmen, die im Konzept der Sektion AGR [4] erwähnt wurden. Nach Rückfragen stellte sich heraus, dass die Daten der DZ99 Daten beinhalten, die AGR als nicht vertrauenswürdig erschienen. Dies hat zur Folge, dass sämtliche Resultate mit Vorsicht genossen werden müssen, und dass bei einer allfälligen Erhebung mit Abweichungen zu rechnen ist. Die Zuordnung der Stichprobe konnte nur anhand dieser Daten berechnet werden, womit die optimale Zuordnung nicht berechnet werden kann (c.f. Abschnitt 6).
Es gilt noch zu bemerken, dass wir zwei Datenlieferungen erhalten haben, wobei die zweite Lieferung zusätzliche Daten aufwies (z.B. Daten der OZ81). In der zweiten Datenlieferung wurden 3 Gemeinden weniger geliefert als in der ersten. Wir haben darauf verzichtet, herauszufinden, welche Gemeinden in beiden Lieferungen übereinstimmten.

3 Schichtung

Die Resultate der OZ2001 sollten auf Ebene der Kantone veröffentlicht werden, womit bereits gesagt ist, dass ein Stichprobenplan mit einer Schichtung in Kantone angestrebt wird. Es werden ebenfalls Schichtungen in kleinere Einheiten, wie Gemeindeklassen bezüglich der Anzahl Bäume oder bezüglich der Grösse der landwirtschaftlichen Nutzfläche durchgeführt. Ebenso wird eine Schichtung in NUTS2 erstellt.
Für die Erstellung von Gemeindeklassen haben wir die Dalenius-Methode (*cum* $\sqrt{f}$ *rule*, c.f. [2]) benutzt.

Dalenius-Methode Die Methode ist in [2] anhand eines Beispiels genau erklärt. Es seien y_0 und y_L der kleinste und der grösste Wert der Variablen y. Wir gehen davon aus, dass die Anzahl Schichten, H, bekannt ist. Das Ziel ist es nun, die Schichtgrenzen $y_1, ..., y_{H-1}$ zu finden, so dass die Varianz des Totals über alle Schichten minimiert wird. Hierfür wird der folgende Algorithmus angewandt:

1. das Intervall $[y_0, y_L]$ in K Intervalle mit konstanter Länge unterteilen, wobei $K \gg H$;
2. für jedes Intervall, $k = 1, ...K$, die folgenden Werte berechnen: die Häufigkeit $f_k(y)$, die Wurzel der Häufigkeit $\sqrt{f_k(y)}$ und die kumulierte Wurzel der Häufigkeit $\phi_k = \sum_{j=1}^{k} \sqrt{f_j(y)}$;
3. das Intervall $[\phi_1, \phi_K]$ in H gleich grosse Intervalle unterteilen, was zu den Intervallsgrenzen $\phi_{[j]}$, $j = 1, ..., H-1$ führt;
4. die zu $\phi_{[j]}$, $j = 1, ..., H-1$ gehörenden Werte von y_j, $j = 1, ..., H-1$, als Schichtgrenzen festlegen.

Die "*Neyman*"-Zuordnung anhand von y ist die optimalste Zuordnung innerhalb der erhaltenen Schichten (c.f. [2]).

4 Schätzer

Das Ziel der OZ2001 ist es, den Halb- und Hochstammbaumbestand in der Schweiz und auf Kantonsebene zu schätzen. Der Baumbestand sollte in die verschiedenen Obstbaumarten unterteilt werden können. Wir haben zwei Schätzmethoden untersucht um das Total des Baumbestandes zu schätzen, einerseits einen einfachen Schätzer (gewichtetes Mittel), andererseits einen Quotientenschätzer. Dabei haben wir Modelle untersucht, welche die Grösse der Gemeinde oder die Anzahl Bäume in der Gemeinde bei der Schätzung berücksichtigt haben. Komplexere Stichprobenpläne mit ungleichen Einschlusswahrscheinlichkeiten der Gemeinden wurden aus Zeitgründen nicht untersucht.
Der Einfachheit halber haben wir die Halb- und Hochstamm Obstbäume manchmal nur kurz mit Bäumen bezeichnet, da keine Gefahr von Verwechslungen vorhanden ist.

4.1 Definitionen

Die Population $\Omega = \{1, ..., N\}$ besteht aus der Menge der N Gemeinden der Grundgesamtheit. Es sei:
y_i : die Anzahl Halb- und Hochstammbäume in der Gemeinde i $(i = 1, ..., N)$;
$t = \sum_{i=1}^{N} y_i$: das Total der Halb- und Hochstammbäume (Wert, den es zu schätzen gilt).

Bei der Schichtung in Kantone oder NUTS2 Regionen, kurz in H Schichten $\Omega_1, ..., \Omega_H$, so dass $\Omega = \bigcup_{h=1}^{H} \Omega_h$, definieren wir für $h = 1, ..., H$:
N_h : die Anzahl Gemeinden in der Schicht h;
y_{ih} : die Anzahl Bäume in der Gemeinde i der Schicht h;
$\bar{y}_h = \sum_{i=1}^{N_h} y_{ih}/N_h$: der Mittelwert der Anzahl Bäume in den Gemeinden der Schicht h;
$S^2_{y\Omega_h} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{ih} - \bar{y}_h)^2$: die empirische Varianz der Anzahl Bäume in den Gemeinden der Schicht h.

4.2 Schätzung anhand einer Stichprobe

In jeder Schicht $h = 1, ..., H$ wird eine Stichprobe s_h von Gemeinden gezogen. Wir definieren:
n_h : die Anzahl Gemeinden der Schicht h, welche in der Stichprobe gezogen werden;
$f_h = n_h/N_h$: der Stichprobenanteil in der Schicht h;
$\bar{y}_{s_h} = \sum_{i \in s_h} y_{ih}/n_h$: der Mittelwert der Anzahl Bäume in den Gemeinden der Stichprobe s_h.

Die gesamte Stichprobe s ist die Vereinigung der Stichproben in jeder Schicht, $s = \bigcup_{h=1}^{H} s_h$. Die Stichprobengrösse ist somit $n = \sum_{h=1}^{H} n_h$. Der einfachste erwartungstreue Schätzer für das Total des Baumbestandes der Grundgesamtheit ist das gewichtete Mittel:

$$\hat{t}_{SS} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i \in s_h} y_{ih} = \sum_{h=1}^{H} N_h \bar{y}_{s_h} \tag{1}$$

Die Varianz des Totals $\hat{t}_{SS}$ im Falle eines geschichteten Stichprobenplans mit einfacher Zufallsstichprobe innerhalb der Schichten ist durch folgende Formel gegeben:

$$V(\hat{t}_{SS}) = \sum_{h=1}^{H} N_h^2 (1 - f_h) \frac{S^2_{y\Omega_h}}{n_h} = \sum_{h=1}^{H} N_h (N_h - n_h) \frac{S^2_{y\Omega_h}}{n_h} \tag{2}$$

Unter den erwähnten Ziehungsbedingungen können wir $S^2_{y\Omega_h}$ anhand von $S^2_{ys_h}$ ohne Bias schätzen, wobei:

$$S^2_{ys_h} = \frac{1}{n_h - 1} \sum_{i \in s_h} (y_{ih} - \bar{y}_{s_h})^2 \tag{3}$$

Somit erhalten wir den Schätzer der Varianz:

$$\hat{V}(\hat{t}_{SS}) = \sum_{h=1}^{H} N_h (N_h - n_h) \frac{S^2_{ys_h}}{n_h} \tag{4}$$

4.3 Quotientenschätzer

Der Quotientenschätzer hat zum Ziel, die einfache Schätzung durch das gewichtete Mittel zu verbessern, d.h. die Varianz des Schätzers soll vermindert werden. Dies wird anhand einer Hilfsvariablen x bewerkstelligt, die mit der zu erhebenden Variablen y gut korreliert. Mit anderen Worten, es wird ein lineares Modell angepasst: $y = \beta x$. Wenn die Varianz von y_i bezüglich der Regressionsgeraden proportional zu x_i ist, ist der Quotientenschätzer nicht nur konsistent[1], sondern auch ein *best linear model-unbiased estimator*[2]. Hierbei muss bemerkt werden, dass der Quotientenschätzer einen Bias aufweist, der jedoch bei grossen Stichproben vernachlässigbar wird (c.f. p. 153 [2]), da er proportional zu $1/n_h$ ist. Bei kleinen Stichprobengrössen in den Schichten wird die Varianz innerhalb der Schicht unterschätzt und somit auch die gesamte Varianz, welche die Summe der Varianzen in den Schichten ist. Wir gehen davon aus, dass die Stichproben in den einzelnen Schichten unabhängig voneinander gezogen werden, womit die Schätzwerte innerhalb jeder Schicht unabhängig sind. Bei kleinen Stichproben und gleichzeitig grosser Varianz von x ist dieses Schätzverfahren nicht mehr verlässlich.
Bei geschichteten Stichproben können zwei Quotientenschätzer unterschieden werden, der separierte und der kombinierte. Wir beschränken uns hier auf die Formeln für das Total, somit ist der separierte Quotientenschätzer $\hat{t}_{Rs}$

$$\hat{t}_{Rs} = \sum_{h=1}^{H} \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}} N_h \bar{x}_h \tag{5}$$

Der Schätzer für β in der Schicht h ist somit $\hat{\beta}_{Rs,h} = \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}}$. Der kombinierte Quotientenschätzer, $\hat{t}_{Rc}$, von t ist:

$$\hat{t}_{Rc} = \frac{\sum_{h=1}^{H} N_h \bar{y}_{s_h}}{\sum_{h=1}^{H} N_h \bar{x}_{s_h}} \sum_{h=1}^{H} N_h \bar{x}_h = \frac{\hat{t}_{SSy}}{\hat{t}_{SSx}} N \bar{x} \tag{6}$$

Der Schätzer für β ist somit $\hat{\beta}_{Rc} = \frac{\sum_{h=1}^{H} N_h \bar{y}_{s_h}}{\sum_{h=1}^{H} N_h \bar{x}_{s_h}}$.
Folgend geben wir noch zwei gebräuchliche Approximationen für die entsprechenden Varianzen $V(\hat{t}_{Rs})$ und $V(\hat{t}_{Rc})$ wieder, die nur bei grossen Stichproben zuverlässig sind.

[1] $\lim_{n \longrightarrow \infty} P\left(\left|\hat{t} - t\right| < \varepsilon\right) = 0,\ \varepsilon > 0$

[2] Bei wiederholtem Ziehen in einer endlichen Population ist $E(\hat{y}) = E(y)$, wobei y als Zufallsvariable angesehen wird: $y_i = \beta x_i + \epsilon_i$.

$$V(\hat{t}_{Rs}) \approx \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h(n_h-1)} \sum_{i=1}^{n_h} \left(y_i - \hat{\beta}_{Rs,h} x_i\right)^2 = \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h} S^2_{r,Rs,s_h} \tag{7}$$

wobei S^2_{r,Rs,s_h} die Varianz der Residuen des separierten Quotientenschätzers in der Schicht h ist.

$$V(\hat{t}_{Rs}) \approx \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h} \left(S^2_{y\Omega_h} + R_h^2 S^2_{x\Omega_h} - 2R_h \rho_h S_{y\Omega_h} S_{x\Omega_h}\right) \tag{8}$$

$$V(\hat{t}_{Rc}) \approx \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h(n_h-1)} \sum_{i=1}^{n_h} \left(y_i - \hat{\beta}_{Rc} x_i\right)^2 = \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h} S^2_{r,Rc,s_h} \tag{9}$$

wobei S^2_{r,Rc,s_h} die Varianz der Residuen des kombinierten Quotientenschätzers ist.

$$V(\hat{t}_{Rc}) \approx \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h} \left(S^2_{y\Omega_h} + R^2 S^2_{x\Omega_h} - 2R \rho_h S_{y\Omega_h} S_{x\Omega_h}\right) \tag{10}$$

wobei $\bar{x}_h$ der Mittelwert der Hilfsvariablen x in der Schicht h ist, $R_h = \bar{y}_h/\bar{x}_h$, $R = \sum N_h \bar{y}_h / \sum N_h \bar{x}_h$ und ρ_h der Korrelationskoeffizient zwischen x und y in der Schicht h ist.

Aus den Formeln geht hervor, dass der Schätzer um so genauer ist, je besser x mit y korreliert, d.h. je grösser ρ_h ist.
Die Resultate des separierten Quotientenschätzers sind bei grossen Stichproben und wenn $N_h \bar{x}_h$ für sämtliche Schichten zur Verfügung steht, besser als jene des kombinierten Quotientenschätzers. Hingegen ist letzterer bei kleinen Stichproben in den einzelnen Schichten dem separierten Quotientenschätzer vorzuziehen. Sobald jedoch die Schichten grosse Unterschiede in der Steigung β_h aufweisen, wird die Varianz des kombinierten Quotientenschätzers grösser.

4.4 Variationskoeffizient

Der geschätzte Variationskoeffizient $cv(\hat{t})$ ist ein Mass für die relative Genauigkeit des Schätzers $\hat{t}$. Anhand der Stichprobe wird der Variationskoeffizient folgendermassen geschätzt:

$$cv(\hat{t}) = \frac{\sqrt{\hat{V}(\hat{t})}}{\hat{t}} \tag{11}$$

Unter der Normalitätshypothese für das Total ist somit das Vertrauensintervall des geschätzten Totals $[\hat{t} \pm z_{1-\alpha/2} cv(\hat{t})\hat{t}]$. Für $\alpha = 5\%$ haben wir somit das 95% Vertrauensintervall $[\hat{t} \pm 1.96\ cv(\hat{t})\hat{t}]$.

4.4.1 Quotientenschätzer vs. Schätzer mit geschichtetem Mittel

Eine Entscheidungshilfe, ob ein Quotientenschätzer effizienter ist als der einfache Schätzer, ist in [2] gegeben und lautet folgendermassen:
In grossen Stichproben mit einfacher Stichprobenziehung hat der Quotientenschätzer eine kleinere Varianz als der einfache Schätzer, wenn

$$\rho(x,y) > \frac{1}{2}\left(\frac{S_x}{\bar{x}}\right) \Big/ \left(\frac{S_y}{\bar{y}}\right) = \frac{cv_x\left(\hat{t}\right)}{2cv_y\left(\hat{t}\right)} \tag{12}$$

Mit anderen Worten, die Korrelation der Schätzvariable und der Hilfsvariablen müssen in jeder Schicht (12) genügen, damit die Verwendung eines Quotientenschätzers sinnvoll ist.
Wir haben eine Stichprobengrösse von 200 Gemeinden vorgegeben. Sofern wir zu viele Schichten kreieren, werden wir in einzelnen Schichten keine grossen Stichproben mehr ziehen können. In diesem Falle wäre die obige Entscheidungshilfe kein Garant mehr für einen effizienten Quotientenschätzer.

5 Stichprobenplan

Um einen Stichprobenplan zu erstellen, muss die Erhebungsvariable "Anzahl Bäume", y, durch eine ihr nahestehende Variable $\tilde{y}$ ersetzt werden, die einer Prognose von y nahe kommt (z.B. lineare Abhängigkeit: $\tilde{y} = \gamma + \delta y$). Hierzu wurden verschiedene Modelle untersucht, wobei jeweils die Daten der OZ91 und/oder der DZ99 als Hilfsvariablen verwendet wurden, um die Anzahl Bäume in landwirtschaftlichen und in nicht landwirtschaftlichen Betrieben für das Jahr 2001 zu prognostizieren. Diese werden in den entsprechenden Abschnitten weiter erläutert.

6 Zuordnung

Um die Stichprobengrösse zu berechnen, können verschiedene Methoden angewandt werden, deren Qualität von den Charakteristiken der Erhebungsmethode und der Qualität der zur Verfügung stehenden Daten abhängt. Von AGR wurde uns versichert, dass die non-response Rate bei 0% liegen würde, deshalb beziehen wir diese nicht in die Berechnung der Stichprobengrösse ein.
Die Erhebungsmethode innerhalb der Schichten wird eine einfache Zufallsstichprobe sein. Dabei wird innerhalb einer Schicht jede Gemeinde, welche ihr angehört, mit der gleichen Wahrscheinlichkeit in die Stichprobe aufgenommen.
Bezeichnen wir die Wahrscheinlichkeit, dass die Gemeinde G_i in der Stichprobe der Schicht h gezogen wird, mit $p_{i,h}$. Somit ist

$$\begin{aligned} p_{i,h} &= p_{j,h} = p_h = \frac{n_h}{N_h} && \text{wenn } G_i, G_j \in \Omega_h \\ p_{i,h} &= p_{j,h} = 0 && \text{wenn } G_i, G_j \notin \Omega_h. \end{aligned}$$

Wobei N_h die Anzahl Gemeinden in der Schicht h ist und n_h die Anzahl Gemeinden der Schicht h, welche in die Stichprobe aufgenommen werden.
Damit wir in jeder Schicht einen Varianzschätzer berechnen können, der auch eine gewisse Aussagekraft hat, haben wir in sämtlichen Schichten eine minimale Stichprobengrösse von 4 Gemeinden verlangt. Die Schichten, welche auf 4 Gemeinden angehoben werden mussten, wurden sodann aus der Berechnung der Zuordnung genommen, und die Zuordnung des Rests der Gemeinden der Stichprobe wurde wiederum anhand der gewählten Zuordnung zu berechnet. Dies hat zur Folge, dass den kleinen Kantonen proportional zu ihrer Grösse mehr Gemeinden für die Stichprobe zugesprochen werden als grösseren Kantonen.

6.1 *Neyman*-Zuordnung

Eine erste Methode zur Berechnung der Stichprobengrösse in den Schichten ist die "*Neyman*-allocation". Sie ist die optimale Zuordnungsmethode bei einer geschichteten Stichprobe, um die Varianz des geschätzten Mittels zu minimieren (siehe [3]).
Die Stichprobengrösse jeder Schicht berechnet sich folgendermassen:

$$n_h = n\frac{N_h S_{\tilde{y}\Omega_h}}{\sum_{h=1}^{H} N_h S_{\tilde{y}\Omega_h}} = n\frac{cv(\hat{t}_{SS\tilde{y}})\hat{t}_{SS\tilde{y}}}{\sum_{h=1}^{H} N_h S_{\tilde{y}\Omega_h}} \tag{13}$$

Somit ist das Produkt der Standardabweichung und der Schichtgrösse ausschlaggebend für die Stichprobengrösse in der Schicht. Dies ist in der Tabelle 2 illustriert, aus der ersichtlich ist, dass die grosse Streuung in der Schicht Zug einen verhältnismässig kleinen Einfluss auf die Stichprobengrösse hatte, da Zug nur 11 Gemeinden aufweist. Hingegen ist die Stichprobengrösse in der Schicht Basel (BS/BL) bei kleinerer Streuung grösser, da diese Schicht mehr Gemeinden aufweist.
Da wir mindestens 4 Gemeinden pro Schicht verlangt haben, ist diese Zuordnung nicht mehr über alle Schichten hinweg optimal. In der Tabelle 2 wurde die Zuordnung aufgrund der Daten der ersten Datenlieferung gemacht.

6.2 Proportionale Zuordnung

Bei dieser Zuordnungsmethode wird die Stichprobengrösse jeder Schicht proportional zur Populationsgrösse (siehe unten) berechnet. Wir haben zwei Methoden untersucht: Proportionalität zur Anzahl Bäume in der Schicht und Proportionalität zur landwirtschaftlichen Nutzfläche in der Schicht.
Wir haben nur nach Kantonen geschichtet, wobei Halbkantone zusammengelegt wurden (AR mit AI, BS mit BL, NW mit OW). Es muss bemerkt werden, dass nicht für alle Gemeinden die Daten der landwirtschaftlichen Nutzfläche vorhanden waren. Diese Gemeinden wurden somit aus der Grundgesamtheit ausgeschlossen.

Tabelle 2: *Neyman*-Zuordnung in Kantonsschichten. Die Anzahl Gemeinden in der Schicht stehen unter N_h, jene in der Stichprobe unter n_h. Die Streuung innerhalb der Schicht wurde mit $S_{\tilde{y}\Omega_h}$ bezeichnet.

Kanton	N_h	n_h	$S_{\tilde{y}\Omega_h}$
ZH	171	11	1032
BE	392	38	1489
LU	106	18	2561
UR	19	4	634
SZ	27	6	3438
OW/NW	16	4	2375
GL	27	4	309
ZG	11	4	3053
FR	243	7	459
SO	126	9	1158
BS/BL	88	7	1241
SH	34	4	486
AR/AI	26	4	1123
SG	87	18	3182
GR	114	4	353
AG	231	13	891
TG	80	16	3050
TI	158	4	96
VD	374	9	367
VS	120	4	190
NE	48	4	231
GE	41	4	135
JU	80	4	680
Total	2619	200	

6.2.1 Proportionalität zur Anzahl Bäume

Um die Stichprobengrösse für jeden Kanton zu bestimmen, wurde hier die gesamte Anzahl Bäume des Kantons verwendet, wobei die Daten aus der DZ99 stammen. Somit haben die einzelnen Gemeinden kein Gewicht mehr in dieser Berechnung, sondern nur noch das Total des Kantons.
Bezeichnen wir die Stichprobengrösse mit n (= 200), dann nimmt die Zuordnung folgende Form:

$$n_h = n\frac{\sum_{i=1}^{N_h} \tilde{y}_{ih}}{\sum_{h=1}^{H}\sum_{i=1}^{N_h} \tilde{y}_{ih}} = 200\frac{\sum_{i=1}^{N_h} \tilde{y}_{ih}}{\sum_{h=1}^{H}\sum_{i=1}^{N_h} \tilde{y}_{ih}} \tag{14}$$

Aus 13 geht hervor, dass diese Zuordnung optimal ist, wenn
$cv(\hat{t}_i) = cv(\hat{t}_j) \qquad \forall\ i, j \in \{1, ..., H\}$.
Wiederum haben wir ein Minimum von 4 Gemeinden pro Schicht verlangt und dann die übrig gebliebene Anzahl Gemeinden anhand der obigen Formel bestimmt.

6.2.2 Proportionalität zur landwirtschaftlichen Nutzfläche

Die Stichprobengrösse wurde hier ähnlich wie im vorhergehenden Abschnitt berechnet, nur dass anstatt der Anzahl Bäume der Wert der landwirtschaftlichen Nutzfläche aus der DZ99 verwendet wurde.
Es sei
Ln_{G_i} : die landwirtschaftliche Nutzfläche der Gemeinde i in Aren.

$$n_h = n\frac{\sum_{G_i \in \Omega_h} Ln_{G_i}}{\sum_{h=1}^{H}\sum_{G_i \in \Omega_h} Ln_{G_i}} \tag{15}$$

7 Schätzung mit geschichtetem Mittel

In diesem Abschnitt werden verschiedene Ansätze zur Erstellung eines Stichprobenplans untersucht. Dabei spielt die Schichtung die Hauptrolle; man kann unsere Schichtverfahren in zwei Ebenen aufteilen:

1. Schichtung nach Kantonen oder Kantonsgruppen (inkl. NUTS)

2. Schichtung nach der landwirtschaftlichen Nutzfläche der Gemeinden, oder nach der Anzahl Bäume in der Gemeinde.

Die zweite Schichtung erfolgte ausschliesslich innerhalb der ersten Schichtung. Je nach Schichtverfahren wurden verschiedene Zuordnungen berechnet, wie die 200 Gemeinden in der Stichprobe verteilt werden können.
Bei diesen Berechnungen haben wir uns auf die Daten der ersten Datenlieferung gestützt, somit haben wir 2911 Gemeinden als Grundgesamtheit.

7.1 Schichtung nach Kantonen

Wir werden eine*Neyman*-Zuordnung sowie die in 6.2 erwähnten Zuordnungen in Kantonsschichten untersuchen.
Anhand der gewählten Zuordnung wurden Teststichproben gezogen, um die Schätzgenauigkeit zu evaluieren.
Aus der Tabelle 3 ist ersichtlich, dass unabhängig von der Zuordnung, sämtliche Variationskoeffizienten entschieden zu gross sind. Es ist zudem offensichtlich, dass mit dem gewählten Ansatz selbst der Variationskoeffizient des Totals an Bäumen in der Schweiz zu gross ist, um verlässliche Resultate zu publizieren. Bei sämtlichen Ansätzen musste in 11 oder 12 Schichten, d.h. in rund der Hälfte der Schichten, die Stichprobengrössen bei der minimalen Anzahl 4 festgelegt werden.
In weiteren Ansätzen haben wir versucht, die*Neyman*- Zuordnung ausgeglichener zu gestalten, indem wir jede Schicht in zwei Gemeinde Grössenklassen (landwirtschaftliche Nutzfläche) aufgeteilt haben. Dies hatte zur Folge, dass Gemeinden, von denen wir die landwirtschaftliche Nutzfläche nicht kannten, ausgeschlossen wurden. Die Konstruktion dieser Grössenklassen wurde anhand der Dalenius-Methode mit Intervallen von 100 Aren berechnet. Diese Schichtung hatte zum Ziel, die Streuung innerhalb der Schichten merklich zu verringern, womit die Stichprobengrösse in kleineren Schichten und Schichten mit einer kleineren Streuung hätte ansteigen sollen. Wir verzichten auf die Wiedergabe der Resultate, da sie sich bezüglich der Variationskoeffizienten nicht vom obigen Ansatz unterscheiden. Ebenso hatten Ansätze mit unterschiedlicher Anzahl Grössenklassen innerhalb der Kantone keinen merklichen Einfluss auf die Variationskoeffizienten.

7.2 Schichtung nach Kantonsgruppen

Wir werden kurz eine Schichtungsmethode untersuchen, in welcher wir gewisse Kantone zusammen gefasst haben, um grössere Schichten zu erhalten. In einem weiteren Abschnitt werden wir Schichtungen in NUTS2 Regionen vornehmen.

7.2.1 Kantons Gruppierung

Wir haben die Kantone so gruppiert, wie das von AGR in [4] vorgeschlagen wurde (c.f. Tabelle 4) und haben eine *Neyman*-Zuordnung gewählt.

Tabelle 3: Zuordnung anhand der *Neyman*-Zuordnung, sowie proportional zur Anzahl Bäume (PropB) und proportional zur landwirtschaftlichen Nutzfläche (PropLN). N_h bezeichnet die Anzahl Gemeinden in der Schicht und n_h die Anzahl Gemeinden in der Stichprobe jeweils pro Schicht und Ansatz. Der Variationskoeffizient wurde mit cv_h abgekürzt und jeweils in Prozenten angegeben. In der untersten Zeile wurde das Total aufgeführt, sowie der jeweilige Variationskoeffizient des Totals.

		Neyman		PropB		PropLN	
Kanton	N_h	n_h	cv_h [%]	n_h	cv_h [%]	n_h	cv_h [%]
ZH	171	11	19.24	13	26.66	14	25.46
BE	392	38	18.28	29	19.30	35	20.72
LU	106	18	17.22	21	17.21	14	17.29
UR	19	4	78.31	4	44.18	4	54.88
SZ	27	6	48.92	5	32.20	4	32.99
NW	16	4	53.56	4	19.18	4	25.88
GL	27	4	53.41	4	50.75	4	30.00
ZG	11	4	18.90	3	26.32	4	19.49
FR	243	7	33.39	7	26.51	14	20.91
SO	126	9	30.92	8	22.49	6	40.02
BL	88	7	28.16	10	14.42	4	46.36
SH	34	4	15.79	4	18.71	4	31.69
AR	26	4	49.59	4	43.77	4	42.89
SG	87	18	23.33	18	20.07	14	16.28
GR	114	4	29.23	4	71.21	6	26.97
AG	231	13	28.41	15	22.80	12	29.52
TG	80	16	19.34	18	17.24	9	26.30
TI	158	4	55.22	4	41.63	4	76.51
VD	374	9	53.16	9	40.54	20	27.61
VS	120	4	54.07	4	90.99	6	44.75
NE	48	4	60.46	4	45.42	4	69.32
GE	41	4	16.57	4	33.17	4	23.53
JU	80	4	22.71	4	35.44	7	32.53
Total		200	8.00	200	7.10	201	7.75

Tabelle 4: Kantons Gruppierungen nach Vorschlag AGR (siehe [4]).

Gruppe	Kantone
1	BL, BS
2	LU, ZG
3	UR, OW, NW, SZ, GL, AI, AR
4	ZH
5	BE
6	FR
7	SO
8	SH
9	SG
10	GR
11	AG
12	TG
13	TI
14	VD
15	VS
16	NE
17	GE
18	JU

Aus der Tabelle 5 ist ersichtlich, dass die Variationskoeffizienten überaus gross sind, so dass diese Schichtungsmethode unter den gegebenen Voraussetzungen verworfen werden muss. Selbst der Schätzer für das Total der Bäume in der Schweiz hat mit einem Variationskoeffizienten von 7.39% eine nicht befriedigende Genauigkeit. Weiterhin musste in 7 Schichten die Stichprobengrösse bei den minimalen 4 Gemeinden festgelegt werden.

7.2.2 NUTS2

Eine weitere mögliche Gruppierung kann anhand der NUTS gemacht werden. Wir haben uns auf NUTS2 beschränkt. Die *Neyman*-Zuordnung, welche wir gewählt haben ist auch hier nicht optimal, da wir in der Region 17 die Stichprobengrösse bei 4 festlegen mussten, da sie anhand der ursprünglichen Zuordnung zu klein war (sieh Tabelle 6). Die Variationskoeffizienten sind bei dieser Schichtung nicht befriedigend.
Anhand einer zweiten Schichtungsmethode haben wir in jeder Schicht (NUTS2 Region) noch anhand der Gemeindegrösse (landwirtschaftlche Nutzfläche der Gemeinde) geschichtet. Dies hatte zur Folge, dass die Gemeinden ohne Angabe der landwirtschaftlichen Nutzfläche aus der Grundgesamtheit ausgeschlossen wurden. Die Grössenklassen wurden anhand der Dalenius-Methode (siehe Abschnitt 3) berechnet. In zwei NUTS2 Regionen (100014: Zürich, 100017: Tessin) wurde keine Grössenklassen Schichtung vorgenommen, da die resultierenden

Tabelle 5: Schätzungen innerhalb der Kantonsgruppen. N_h steht für die Schichtgrösse, n_h für die Stichprobengrösse und der Variationskoeffizient wurde mit cv_h bezeichnet. In der letzten Zeile wurde der cv des Gesamt Totals angegeben.

Kt. Gruppe	N_h	n_h	cv_h [%]
1	88	7	26.06
2	117	21	20.57
3	115	18	25.01
4	171	12	25.49
5	392	39	16.87
6	243	7	15.35
7	126	10	42.48
8	34	4	25.67
9	87	19	17.97
10	114	4	44.35
11	231	14	26.35
12	80	16	31.15
13	158	4	73.40
14	374	9	18.87
15	120	4	60.18
16	48	4	41.26
17	41	4	43.60
18	80	4	59.40
Total		200	7.39

Tabelle 6: Die Schichtgrösse wurde mit N_h, die Stichprobengrösse mit n_h und der Variationskoeffizient mit cv_h bezeichnet. In der letzten Zeile haben wir die Grösse der gesamten Stichprobe angegeben, sowie den Variationskoeffizienten vom Total Schätzer über alle Schichten hinweg.

NUTS2	N_h	n_h	cv_h [%]
11	554	11	28.35
12	894	64	13.10
13	320	21	18.96
14	171	11	32.84
15	375	58	23.43
16	183	31	16.21
17	207	4	65.00
Total	2704	200	8.76

Schichten immer noch eine grosse Streuung aufwiesen und da jeweils eine Schicht mit wenig Gemeinden entstand. Andererseits haben wir in anderen NUTS2 Regionen bis zu vier Grössenklassen gebildet (siehe Tabelle 7).

Obwohl sich der Variationskoeffizient des Totals gegenüber der einfachen Schichtung in NUTS2 um rund 2.5% verbessert hat, ist er für das Total und in sämtlichen Schichten unzureichend. Die Vergleichbarkeit der Tabellen 6 und 7 ist jedoch eingeschränkt, da sie sich nicht auf die gleiche Grundmenge beziehen. Es ist weiter zu bemerken, dass zwischen der zweitgrössten und der grössten Klasse die Schichtgrenze massiv ansteigt. Es wurde jedoch nicht untersucht, ob Ausreisser vorhanden sind, da jeweils eine beträchtliche Anzahl Gemeinden den grössten Klassen zugehören.

Fazit: Keiner der untersuchten Methoden für die Schätzung des Totals anhand eines geschichteten Mittels erbrachte befriedigende Resultate. Wir mussten auch feststellen, dass die Kantonsgruppierungen nur einen geringen Einfluss auf die Qualität der Schätzer innerhalb der Schichten hatte. Falls einer der Ansätze dieses Abschnitts beibehalten werden soll, muss die Stichprobengrösse merklich erhöht werden.

8 Schätzung anhand eines Quotientenschätzers

Diesem Ansatz liegt die Hypothese zu Grunde, dass das Total der Bäume in landwirtschaftlichen Betrieben durch die Daten der BZ2000 und der DZ2001 bekannt sind, das heisst, es muss nicht geschätzt werden. Einzig die Anzahl Bäume in nicht landwirtschaftlichen Betrieben muss somit geschätzt werden. Zur Erarbeitung des Stichprobenplans wurde die Erhebungsvariable durch eine Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben

Tabelle 7: Schichtung in NUTS2 und anhand von Grössenklassen der Gemeinden bezüglich landwirtschaftlicher Nutzfläche (Klassen). Diese Klassen bestehen aus kleinflächigen (kl), mittelgross flächigen (mittel), grossflächigen (gross) und sehr grossflächigen (sehr gross) Gemeinden. '-' bedeutet, dass keine Flächenschichtung vorgenommen wurde. Es wurde eine *Neyman*-Zuordnung angewandt. Die Schichtgrösse ist mit N_h bezeichnet, die Stichprobengrösse der Schichten mit n_h. Die Schichtgrenze der landwirtschaftlichen Nutzfläche wurde mit Ln_h abgekürzt und in Aren angegeben. Der Variationskoeffizient wurde mit cv_h bezeichnet und in Prozent angegeben.

Klasse	NUTS2	N_h	n_h	Ln_h [a]	cv_h [%]
kl	11	316	5	26276	33.94
gr	11	219	8	198721	30.10
kl	12	302	8	20081	40.92
mittel	12	292	17	38531	21.72
gross	12	177	14	71448	29.17
sehr gr	12	118	21	367121	15.85
kl	13	162	8	22913	41.44
gr	13	157	15	146433	17.08
-	14	171	15	168628	32.51
kl	15	139	6	28677	60.05
mittel	15	116	12	67141	28.93
gross	15	113	32	273356	14.44
kl	16	90	8	54088	18.22
gr	16	89	24	375869	12.35
-	17	158	4	52694	36.17
Total		2619	200		6.21

ersetzt. Ebenso wurde die Hilfsvariable (Anzahl Bäume in landwirtschaftlichen Betrieben) durch eine Prognose anhand der Daten vorangegangener Erhebungen ersetzt (siehe Abschnitt 8.3).
Wir haben die Gemeinden, deren Daten von 1981 bis 1999 vorhanden waren und die in der OZ2001 miteinbezogen werden, von den Daten der Gemeinden, von welchen diese Daten nicht vorhanden waren, getrennt. Dies mit der Absicht zwei Modelle aufzustellen. Für die Gemeinden von denen keine Daten aus früheren Erhebungen vorhanden sind, sollte eine einfache Stichprobe gezogen werden, ohne ihre Kantonszugehörigkeit oder andere Kriterien miteinzubeziehen. Für die Gemeinden, von denen Daten über deren Baumbestände bestehen, sahen wir die Möglichkeit anhand von Hilfsvariablen bessere Schätzer konstruieren zu können. Zudem konnten wir feststellen, dass die meisten Gemeinden, von welchen wenig bis gar keine Daten vorhanden waren, eher kleine Gemeinden sind. Damit sollte ein allfälliger Bias im Schätzverfahren einen kleinen Einfluss auf das Total der Kantone, respektive der Schweiz haben.

8.1 Grundgesamtheit

Bezeichnen wir: die Menge der Gemeinden, welche in der LBZ2000 erhoben wurden mit Ω_{BZ2000}; die Menge der Gemeinden, die in der DZ99 erhoben wurden mit Ω_{DZ99}; die Menge der Gemeinden, welche in der OZ91 (OZ81) erhoben wurden mit Ω_{OZ91} (Ω_{OZ81}). Von der Bodennutzungsstatistik 1997 wissen wir einzig welche Gemeinden erhoben wurden, nicht aber wie gross die Anzahl an Bäumen in den betreffenden Gemeinden ist, deshalb wurde diese Variable nicht berücksichtigt.
In diesem Abschnitt werden ausschliesslich die Daten der zweiten Datenlieferung verwendet. Es sei $\Omega = \bigcup_{i=1}^{2908} G_i$ die Menge der 2908 Gemeinden (G_i), deren Daten von AGR mit der zweiten Datenlieferung geliefert wurde und die somit in die OZ2001 miteinbezogen werden. Da die Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben anhand der Daten der OZ81, der OZ91 und der DZ99 berechnet wurde, sahen wir uns gezwungen die Grundgesamtheit nochmals neu zu definieren, da nicht alle Gemeinden in Ω in der OZ81, der OZ91 und in der DZ99 erhoben wurden. Dabei muss erwähnt werden, dass die Gemeindegrenzen zwischen 1981, 1991 und 1999 (resp. 2000/2001) Änderungen unterworfen waren, die nicht miteinbezogen wurden. Definieren wir:

$$\begin{aligned}\Omega_{D^*} &= \left(\bigcup_{G_j\in\Omega_{OZ81}} G_j\right) \cap \left(\bigcup_{G_j\in\Omega_{OZ91}} G_j\right) \cap \left(\bigcup_{G_j\in\Omega_{DZ99}} G_j\right) \\ &= \text{Gemeinden, welche in der OZ81, der OZ91 und in der DZ99 erhoben wurden.}\end{aligned}$$

$$\Omega_{\bar{D}^*} = \Omega \setminus \Omega_{D^*} = \text{Gemeinden,von denen keine Daten vorhanden sind}$$

Somit haben wir die Grundgesamtheit der Gemeinden, welche für den Stichprobenplan ausschlaggebend ist: $\Omega_* = \Omega_{D^*} \cup \Omega_{\bar{D}^*}$.
Wir sahen uns gezwungen die Schnittmenge der Gemeinden dieser drei Erhebungen als Basis zu nehmen, da sonst rund 700 fehlende Datenwerte entstanden wären. Wir haben darauf verzichtet zu untersuchen, ob diese oder wieviele dieser Gemeinden in die Kategorie der neuen oder zusammengelegten Gemeinden gehören. Es blieben uns somit 2202 Gemeinden in den 23 Kantonen (Halbkantone wurden zusammengelegt) in Ω_{D^*} übrig.
Wir werden uns in diesem Abschnitt ausschliesslich mit der Erarbeitung eines Stichprobenplans für die Gemeinden, von welchen bereits Daten über deren Baumbestände bestehen (Ω_{D^*}), befassen. Für die anderen Gemeinden ($\Omega_{\bar{D}^*}$) ist eine einfache Stichprobe vorgesehen, die jedoch nicht ausgearbeitet wurde, da das Projekt vorzeitig abgebrochen wurde.

8.2 Schätzer

Für die Errichtung des Stichprobenplans benutzen wir Daten aus früheren Erhebungen als Hilfsvariablen, um die Erhebungsvariable zu ersetzen. Da keine Daten der nicht landwirtschaftlichen Betriebe aus früheren Erhebungen vorhanden sind, haben wir mit Prognosen der Erhebungsvariablen sowie mit Prognosen des Totals der Bäume aller Betriebe gearbeitet.

8.2.1 Definitionen

Wir ergänzen die Definitionen in 4.1 mit Definitionen bezüglich der Prognosen:

- Prognose des Totals der OZ2001:

P_{T,G_i} : in der Gemeinde i
$P_{T,h}$: in der Schicht h
P_{T,s_h} : in der Stichprobe s_h der Schicht h

- Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben der OZ2001

P_{N,G_i} : in der Gemeinde i
$P_{N,h}$: in der Schicht h
P_{N,s_h} : in der Stichprobe s_h der Schicht h

- Prognose der Anzahl Bäume in landwirtschaftlichen Betrieben der OZ2001

P_{L,G_i} : in der Gemeinde i
$P_{L,h}$: in der Schicht h
P_{L,s_h} : in der Stichprobe s_h der Schicht h
Des weiteren wird der jeweilige Mittelwert mit $\bar{P}_{i,j}$ bezeichnet ($i \in \{T, N, L\}, j \in \{h, s_h\}$).
Die Prognose Methoden werden in den jeweiligen Abschnitten mathematisch genau mit den entsprechenden Formeln definiert.

Bezeichnen wir mit b_{E,G_i} die Anzahl Halb- und Hochstammbäume der Gemeinde i in der Erhebung E ($E \in \{OZ81, OZ91, DZ99\}$),
$\hat{t}$: Schätzer für das Total der Bäume in der OZ2001
$\hat{t}_N$: Schätzer für das Total der Bäume in nicht landwirtschaftlchen Betrieben
t_L : das Total der Bäume in landwirtschaftlchen Betrieben.
Somit haben wir $\hat{t} = t_L + \hat{t}_N$

8.2.2 Quotientenschätzer

Als Hilfsvariable für den Quotientenschätzer fungiert die Prognose der Anzahl Bäume in landwirtschaftlichen Betrieben, P_L. Um das Total der Bäume in nicht landwirtschaftlichen Betrieben zu schätzen wird aus (6):

$$\hat{t}_{N,Rc} = \frac{\sum_{h=1}^{H} N_h \bar{P}_{N,s_h}}{\sum_{h=1}^{H} N_h \bar{P}_{L,s_h}} \sum_{h=1}^{H} N_h \bar{P}_{L,h} \tag{16}$$

Für jede Schicht $h = 1, ..., H$ haben wir somit:

$$\hat{t}_{N,Rc,h} = \frac{\sum_{h=1}^{H} N_h \bar{P}_{N,s_h}}{\sum_{h=1}^{H} N_h \bar{P}_{L,s_h}} N_h \bar{P}_{L,h}. \tag{17}$$

Der kombinierte Quotientenschätzer für das Total aller Bäume, über alle Schichten hinweg, ist somit:

$$\hat{t}_{Rc} = t_L + \hat{t}_{N,Rc} \tag{18}$$

Wollen wir das Total auf Schichtebene schätzen, so müssen wir den separierten Quotientenschätzer (5) benutzen:

$$\hat{t}_{N,Rs} = \sum_{h=1}^{H} \frac{N_h \bar{P}_{N,s_h}}{N_h \bar{P}_{L,s_h}} N_h \bar{P}_{L,h} \tag{19}$$

Für die einzelnen Schichten $h = 1, ..., H$ haben wir somit:

$$\hat{t}_{N,Rs,h} = \frac{\bar{P}_{N,s_h}}{\bar{P}_{L,s_h}} N \bar{P}_{L,h} \ . \tag{20}$$

über alle Schichten hinweg, haben wir somit folgenden separierten Quotientenschätzer für das Total der gesamten Anzahl Bäume in der OZ2001:

$$\hat{t}_{Rs} = t_L + \hat{t}_{N,Rs} \tag{21}$$

Und für jede Schicht:

$$\hat{t}_{Rs,h} = t_{L,h} + \hat{t}_{N,Rs,h} \tag{22}$$

Da in landwirtschaftlichen Betrieben eine Vollerhebung gemacht wird, können wir laut AGR davon ausgehen, dass $Var(t_L) = 0$. Mit anderen Worten, es wird davon ausgegangen, dass die non-response Rate 0% beträgt. Dies hat zur Folge, dass

$$Var(\hat{t}) = Var(t_L + \hat{t}_N) = Var(\hat{t}_N) \tag{23}$$

Somit ist die Approximation der Varianz für den kombinierten Quotientenschätzer (siehe (7)):

$$\hat{V}\left(\hat{t}_{Rc}\right) = \sum_{h=1}^{H} \frac{N_h^2\left(1 - f_h\right)}{n_h} S^2_{r,Rc,s_h} \tag{24}$$

und die Approximation der Varianz für den separierten Quotientenschätzer (siehe (7)):

$$\hat{V}\left(\hat{t}_{Rs}\right) = \sum_{h=1}^{H} \frac{N_h^2\left(1 - f_h\right)}{n_h} S^2_{r,Rs,s_h} \tag{25}$$

8.2.3 Variationskoeffizient

Aufgrund von (23) ist der Variationskoeffizient des Totals somit

$$cv(\hat{t}) = \frac{\sqrt{var(y)}}{\hat{t}}$$

Anhand des kombinierten Quotientenschätzers ergibt dies für das Total

$$cv\left(\hat{t}_{Rc}\right) = \frac{\sqrt{var(\hat{t}_{N,Rc})}}{\hat{t}_{Rc}} \tag{26}$$

und für den separierten Quotientenschätzer

$$cv\left(\hat{t}_{Rs}\right) = \frac{\sqrt{var(\hat{t}_{N,Rs})}}{\hat{t}_{Rs}} \tag{27}$$

Anhand des separierten Quotientenschätzers haben wir somit:

$$cv\left(\hat{t}_{Rs,h}\right) = \frac{\sqrt{var(\hat{t}_{N,Rs,h})}}{\hat{t}_{Rs,h}} \tag{28}$$

Somit können wir mit einem kleinen Variationskoeffizienten rechnen, wenn $t_L \gg \hat{t}_{N,Rc}$, respektive $t_L \gg \hat{t}_{N,Rs}$.

8.3 Prognosen für die OZ2001

Wie bereits erwähnt, mussten wir die Daten der OZ2001 prognostizieren, da keine anderen Daten zur Verfügung standen. Wir zeigen hier drei Prognose Methoden auf, ohne auszuschliessen, dass andere Methoden existieren. Für die Berechnungen werden die Formeln aus 8.2.2 und 8.2.3 verwendet.

8.3.1 Prognose anhand der Daten der OZ81, der OZ91 und der DZ99

Dieses Prognosemodell wurde uns von AGR geliefert. Der Trend wurde dabei als prozentuale Veränderung auf Gemeindeebene zwischen den Daten der OZ81 und der OZ91 berechnet und auf die Daten der OZ91 angewandt. Wie mit neuen Gemeinden oder Gemeinden, welche zusammengelegt wurden, verfahren wurde, ist uns nicht bekannt. Die Anzahl Bäume in landwirtschaftlichen Betrieben der OZ2001 wurde in diesem Modell der Anzahl Bäume in der DZ99 gleichgesetzt.

$$P_{T,G_i} = b_{OZ91,G_i} + \frac{100 \cdot (b_{OZ91,G_i} - b_{OZ81,G_i})}{b_{OZ81,G_i}} \cdot b_{OZ91,G_i} \qquad \forall G_i \in \Omega_{D^*} \qquad (29)$$

$$P_{L,G_i} = b_{DZ99,G_i} \qquad \forall G_i \in \Omega_{D^*}$$

Aus dieser Prognose haben wir eine Prognose für die Anzahl Bäume in nicht landwirtschaftlichen Betrieben erstellt:

$$P_{N,G_i} = b_{OZ91,G_i} + \frac{100 \cdot (b_{OZ91,G_i} - b_{OZ81,G_i})}{b_{OZ81,G_i}} \cdot b_{OZ91,G_i} - b_{DZ99,G_i} \qquad \forall G_i \in \Omega_{D^*} \qquad (30)$$

Die Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben kann somit folgendermassen beschrieben werden:

$$P_N = P_T - b_{DZ99}$$

Diese Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben ergab in mehr als einem drittel der Daten einen negativen Wert, da in der DZ99 mehr Bäume vorhanden waren als für die OZ2001 prognostiziert war!
Damit ein Quotientenschätzer gute Eigenschaften hat, sollten P_N und P_L miteinander korrelieren.Aufgrund der negativen Werte kann nicht von einer guten Korrelation zwischen P_N und P_L gesprochen werden (siehe Abbildung 1). Somit kann der kombinierte Quotientenschätzer bei diesem Ansatz nicht angewandt werden.
In den Abbildungen im Anhang B.1 wurde die prognostizierte Anzahl Bäume in nicht landwirtschaftlichen Betrieben gegen die DZ99 pro Kanton abgebildet. Halbkantone wurden zusammengelegt und jeweils nur mit einem der beiden AKZ (Autokennzeichen) beschriftet.

Kantone mit wenigen Gemeinden in der OZ2001 müssten zusammengelegt werden (UR, SZ, NW/OW, GL, ZG, AI/AR, GR). Eine lineare Abhängigkeit kann in wenigen Kantonen nur schwerlich festgestellt werden. Damit ein separierter Quotientenschätzer angewandt werden könnte, müsste in jeder Schicht eine lineare Abhängigkeit ersichtlich sein, was nicht der Fall ist. Wir haben anhand von (12) überprüft, ob der Quotientenschätzer effizienter ist als der einfache Schätzer (siehe Tabelle 8). Nur gerade in den Schichten SO, SH, GR und AG ist unter diesem Modell der Quotientenschätzer effizienter. Somit ist dieses Modell in der vorliegenden Problemstellung nicht von Nutzen.
In diesem Modell wurde das Problem mit den negativen Werten von P_N nicht angegangen, wobei wir uns bewusst sind, dass dies an und für sich eine Quotientenschätzung anhand dieses Modells bereits in Frage stellt, da ein Anteil nicht negativ sein kann. Eine andere Methode zur Prognose von P_N, die sich der negativen Werte annimmt, ist in 8.3.3 wiedergegeben.

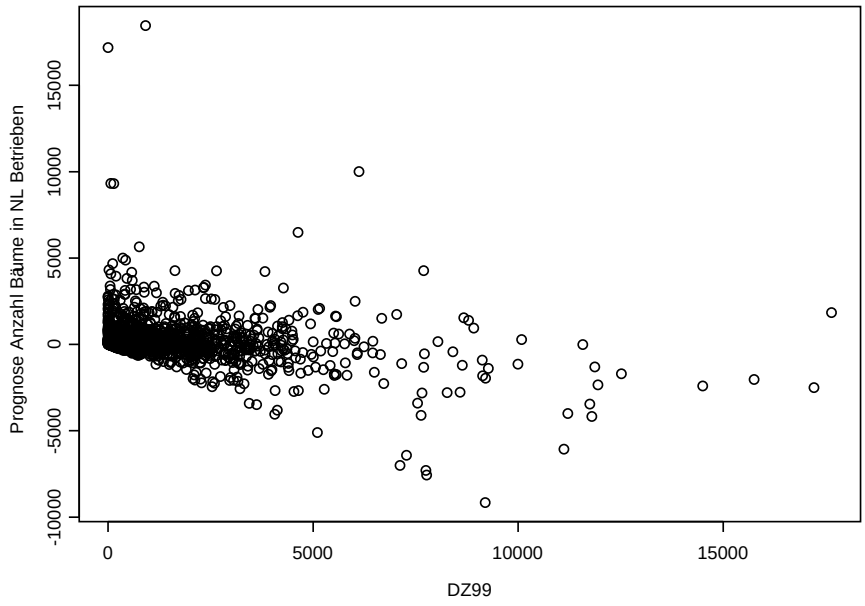


Abbildung 1: Prognose der Anzahl Bäume nicht landwirtschaftlicher Betriebe gegen die DZ99. Die Prognose basiert auf den Daten der OZ81, OZ91 und der DZ99.

8.3.2 Prognose anhand der Daten der OZ51 bis OZ91

Für diese Prognose hatten wir nur die Daten auf Kantonsebene aus der Publikation [1] zur Verfügung. Die Veränderungswerte für den Kanton Jura vor 1981 wurden von Bern

Tabelle 8: Für jede Kantonsschicht, die Korrelation zwischen der Prognose der Anzahl Bäume in nicht landw. Betrieben, ρ_h, das Verhältnis der Variationskoeffizienten, $cv_{b_{DZ91}}/2cv_{P_N}$ (siehe Formel (12)), sowie die Evaluation zwischen der Korrelation und dem Verhältnis der Variationskoeffizienten. In der Evaluationskolonne steht 'ok.', wenn der Quotientenschätzer effizienter ist als der Schätzer anhand des Mittels.

Kanton	$\rho_h(P_N, b_{DZ99})$	$cv_{b_{DZ99}}(\hat{t})/2cv_{P_N}(\hat{t})$	Evaluation
ZH	-0.15	0.06	
BE	-0.19	0.15	
LU	-0.30	-0.10	
UR	-0.60	-0.09	
SZ	-0.71	-0.17	
OW/NW	-0.59	-0.08	
GL	-0.05	0.31	
ZG	-0.87	-0.25	
FR	-0.02	0.14	
SO	0.51	0.20	ok.
BS/BL	0.15	0.24	
SH	0.46	0.21	ok.
AR	-0.63	-0.18	
SG	-0.12	-0.02	
GR	0.21	0.18	ok.
AG	0.29	0.23	ok.
TG	-0.52	-0.27	
TI	0.10	0.36	
VD	-0.18	0.18	
VS	0.23	0.32	
NE	-0.06	0.24	
GE	-0.48	0.10	
JU	0.06	0.14	
Total	-0.26	0.14	

übernommen. Für den Kanton Wallis wurde der fehlende Wert aus dem Jahr 1951 nicht ersetzt.
Aus der Abbildung im Anhang B.2 ist ersichtlich, dass kein gesamtschweizerischer Trend bestimmt werden kann. Um einen Trend auf Kantonsebene zu berechnen, sind einerseits zu wenig Daten vorhanden, andererseits sind die Schwankungen zu gross. Deshalb wurde dieser Ansatz nicht weiter verfolgt.

8.3.3 Begrenzte Prognose

Um das Problem mit den negativen Werten von P_N zu lösen, haben wir folgende Prognose der Anzahl Bäume der nicht landwirtschaftlichen Betriebe gemacht:

$$\begin{aligned} P'_{N,G_i} &= \max(b_{OZ91,G_i}, b_{DZ99,G_i}) - b_{DZ99,G_i} \qquad \forall G_i \in \Omega_{D^*} \\ P'_{L,G_i} &= b_{OZ91,G_i} \qquad \forall G_i \in \Omega_{D^*} \end{aligned}$$

Diesem Ansatz liegt die Hypothese zu Grunde, dass die Baumzahl zwischen 1991 und 2001 abgenommen hat. Da diese Hypothese nicht allgemein gültig ist, sind die Werte von P'_{N,G_i} 0, sobald in der $OZ91$ weniger Bäume angegeben sind als in der $DZ99$ (c.f. Abbildung 2). Aus der Abbildung 2 geht ebenfalls hervor, dass das gewählte Modell bedingt, dass die Anzahl Bäume in nicht landwirtschaftlichen Betrieben durch die Anzahl Bäume der OZ91 beschränkt ist, was sich als eine "diagonale" obere Grenze bemerkbar macht. Somit haben wir ein Modell, das nach oben und unten begrenzt ist, womit bereits gesagt ist, dass die Residuen eines linearen Modells, nicht normal verteilt sein können.

Aus den Abbildungen im Anhang B.3 ist ersichtlich, dass in den meisten Fällen ein lineares Modell durchaus realistisch ist. Dies scheint jedoch fraglich, wenn wir die Korrelationskoeffizienten in der Tabelle 9 betrachten. In unserem Modell bedeutet, dass $\rho = 1$, wenn z.B. $b_{DZ99,G_i} = 0 \quad \forall G_i \in \Omega_{D^*}$. In mehreren Kantonsschichten ist der Korrelationskoeffizient erfreulich gross, wir haben aber darauf verzichtet, dies weiter zu untersuchen. Hingegen ist die Korrelation in GL derart klein, dass es erstaunlich wäre, wenn wir anhand unseres Modells einen effizienten Schätzer bilden könnten. Es ist jedoch klar ersichtlich, dass in den meisten Schichten, ausser SZ, GL und ZG, der Quotientenschätzer dem einfachen Schätzer vorzuziehen ist. Aus diesen Gründen haben wir anhand der Formel (19) einen separierten Quotientenschätzer konstruiert:

$$\hat{t}_{N,Rs} = \sum_{h=1}^{H} \frac{N_h \bar{P}'_{N,s_h}}{N_h \bar{P}'_{L,s_h}} N_h \bar{P}'_{L,h} \tag{31}$$

wobei $\bar{P}'_{N,s_h} = \frac{1}{n_h} \sum_{G_i \in s_h} P'_{N,G_i}$, $\bar{P}'_{L,s_h} = \frac{1}{n_h} \sum_{G_i \in s_h} P'_{L,G_i}$ und $\bar{P}'_{L,h} = \frac{1}{N_h} \sum_{G_i \in h} P'_{L,G_i}$.

Tabelle 9: Für jede Kantonsschicht, die Korrelation zwischen der Prognose der Anzahl Bäume in nicht landw. Betrieben, ρ, das Verhältnis der Variationskoeffizienten, $cv_{b_{OZ91}}/2cv_{P'_N}$ (siehe Formel (12)), sowie die Evaluation zwischen der Korrelation und dem Verhältnis der Variationskoeffizienten. In der Evaluationskolonne steht 'ok.', wenn der Quotientenschätzer effizienter ist als der Schätzer anhand des Mittels; 'ρ_h klein' bedeutet, dass die Korrelation fraglich ist.

Kanton	$\rho_h(P'_N, b_{OZ91})$	$cv_{b_{OZ91}}(\hat{t})/2cv_{P'_N}(\hat{t})$	Evaluation
ZH	0.79	0.50	ok.
BE	0.71	0.50	ok.
LU	0.60	0.50	ok. ρ klein
UR	0.60	0.21	ok. ρ klein
SZ	0.37	0.41	
OW/NW	0.38	0.26	ok. ρ klein
GL	0.09	0.44	
ZG	-0.40	0.20	
FR	0.62	0.41	ok. ρ klein
SO	0.91	0.44	ok.
BS/BL	0.84	0.42	ok.
SH	0.87	0.41	ok.
AR/AI	0.32	0.20	ok. ρ klein
SG	0.55	0.41	ok. ρ klein
GR	0.92	0.29	ok.
AG	0.87	0.41	ok.
TG	0.70	0.40	ok.
TI	0.98	0.47	ok.
VD	0.84	0.43	ok.
VS	1.00	0.47	ok.
NE	0.92	0.38	ok.
GE	0.80	0.34	ok.
JU	0.76	0.28	ok.
Total	0.63	0.47	ok.

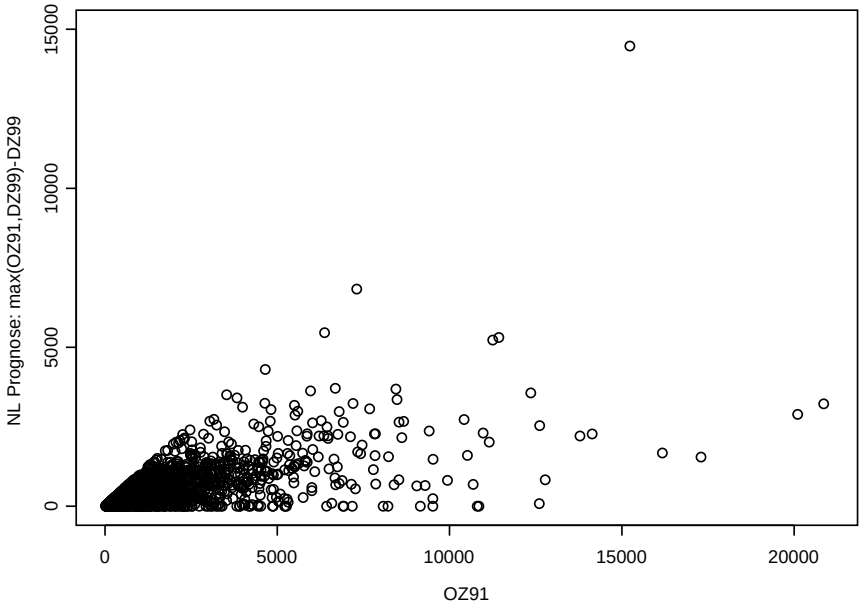


Abbildung 2: Begrenzte Prognose der Anzahl Halb- und Hochstammbäume in nicht landwirtschaftlichen Betrieben im Jahr 2001 gegen die Daten der OZ91 aufgezeichnet.

Aus (7) und (25) haben wir somit den Varianzschätzer des separierten Quotientenschätzers des Totals:

$$V(\hat{t}_{Rs}) \approx \sum_{h=1}^{H} \frac{N_h^2(1-f_h)}{n_h(n_h-1)} \sum_{i=1}^{n_h} \left(P'_{N,G_i} - \hat{\beta}_{Rs,h} b_{OZ91,G_i} \right)^2 \tag{32}$$

wobei $\hat{\beta}_{Rs,h} = \frac{\bar{P}'_{N,s_h}}{\bar{P}_{L,s_h}}$.
Die Stichprobengrössen in den Schichten haben wir anhand einer *Neyman*-Zuordnung bestimmt und den Variationskoeffizienten des separierten Quotientenschätzers in jeder Schicht analog zu (28) geschätzt (siehe Tabelle 10).
Aus dieser Tabelle ist ersichtlich, dass der Variationskoeffizient in rund der Hälfte der Schichten, sowie für den Totalschätzer, kleiner als 5% ist. Es fällt jedoch auch auf, dass Zürich, Luzern und Freiburg weiterhin einen sehr grossen Variationskoeffizienten aufweisen. Dies erstaunt nicht, wenn man die Abbildungen im Anhang B.3 genauer betrachtet. Denn in diesen Kantonen ist die Streuung recht gross. Hingegen ist der Variationskoeffizient in einigen Kantonen mit einer grossen Streuung kleiner als 5% (z.B. ZG, GL, UR, etc.). Dabei handelt es sich aber vor allem um Kantone, welche zu wenige Daten aufweisen, um ein gutes lineares Regressionsmodell anwenden zu können. Man beachte auch die grossen Unterschiede in den Skalen der Graphiken, das heisst, bei einer grösseren Skala kann die Streuung einen grösseren Einfluss auf die Varianz der Residuen haben.

Im weiteren muss darauf hingewiesen werden, dass die Residuen der linearen Modelle nicht normal verteilt sein können, da sie nach oben und unten begrenzt sind.
Es scheint fraglich, ob die Stichprobengrösse von 200 Gemeinden ausreichend ist, da, wie aus der Tabelle 10 hervorgeht, in 10 Schichten die Stichprobengrösse auf 4 angehoben werden musste. Folglich ist denn auch keiner der Variationskoeffizienten in Schichten mit mehr als 100 Gemeinden kleiner als 5%, mit Ausnahme des TI.

Wir haben untersucht, wie gross die Stichprobengrösse sein muss, dass in sämtlichen Schichten ein Variationskoeffizient von weniger als 5% vorhanden ist. Wie aus der Tabelle 11 hervorgeht, konnte mit einer Stichprobengrösse von 751 Gemeinden dieses Ziel erreicht werden, ausser in FR. In einer allfälligen Erhebung könnten die Variationskoeffizienten in den Kantonen Bern und Luzern dennoch die 5% Grenze überschreiten, da die Variationskoeffizienten in diesen Schichten nahe bei 5% sind.
Wenn wir das Total auf gesamtschweizerischer Ebene ohne Schichtung Schätzen wollen, so sind der separierte und der kombinierte Quotientenschätzer gleich. Anhand einer Stichprobe mit $n_1 = 200$ Gemeinden erhalten wir einen Variationskoeffizienten $cv_1(\hat{t}_{N,Rs}) = cv_2(\hat{t}_{N,Rc}) = 16.18\%$. Der Qualitätsverlust beruht auf dem Verzicht auf die Schichtung. Wir wenden die folgende Formel an, um die Stichprobengrösse n_2 zu berechnen, damit ein Variationskoeffizient $cv_2(\hat{t}_{N,Rs}) = cv_2(\hat{t}_{N,Rc}) = 5\%$ erreicht werden kann:

$$\begin{aligned} \frac{cv_1}{cv_2} &= \frac{\sqrt{n_2}}{\sqrt{n_1}}\sqrt{\frac{N-n_1}{N-n_2}} \\ n_2 &= \left(\frac{cv_1}{cv_2}\right)^2 \frac{n_1}{N-n_1}(N-n_2) \end{aligned} \tag{33}$$

wobei $N = 2202$ Gemeinden. Somit muss mit $n_2 = 1130$ Gemeinden gerechnet werden. Verfolgen wir jedoch den obigen Ansatz des separierten Quotientenschätzers mit Schichtung, so können wir mit einer Stichprobengrösse von $n_2 = 30$ Gemeinden das Total der Bäume in nicht landwirtschaftlichen Betrieben mit einem Variationskoeffizienten unter 5% schätzen. Die Zuordnung der 30 Gemeinden in die Schichten ist jedoch nicht mehr möglich. Denn wir wären dazu gezwungen, weniger als 4 Gemeinden pro Schicht zu erheben. Es ist realistischer den obigen Ansatz beizubehalten und die Stichprobengrösse zu erhöhen, wobei die Ansprüche an die Schätzung in gewissen Schichten reduziert werden müssten.

Fazit: Der separierte Quotientenschätzer mit dem Modell der begrenzten Prognose ist der bisher vielversprechenste, auch wenn er uns bei einer Stichprobengrösse von 200 unseren Ansprüchen nicht genügt. Eine Zusammenlegung der Kantone scheint unausweichlich zu sein. Aber wie aus dem Abschnitt 7 ersichtlich ist, ist dies noch kein Garant für markante Verbesserungen für den Schätzer des Totals innerhalb der Schichten. Da das Projekt vorzeitig abgebrochen wurde, haben wir diesen Ansatz nicht weiter verfolgt. Wenn nur eine Schätzung des Totals für die ganze Schweiz verlangt wir, so sollte trotzdem die Schichtung miteinbezogen werden. Denn anhand der Schichtung kann die Varianz verkleinert werden. Wir haben keine weiteren Untersuchungen angestellt, wie die Zuordnung bei einer kleinen

Stichprobengrösse gemacht werden könnte. Ebenso wurde die Schätzmethode der Varianz nicht der Stichprobengrösse angepasst.
In einem weiteren Modell, könnte man zum Beispiel auch die Differenz der Gemeindefläche und der Landwirtschaftlichen Nutzfläche der Gemeinde berechnen. Anhand der Anzahl Gemeindeeinwohner könnte ein Dichtemass für die Gemeinden berechnet werden. Falls die Anzahl Bäume der OZ91 und dieses Dichtemass korrelieren, könnte es als Hilfsvariable oder Schichtvariable verwendet werden. Dieser Ansatz wurde einerseits wegen fehlender Daten und andererseits wegen des Abbruchs der OZ2001 nicht genauer studiert.

9 Schlussfolgerung

Wir haben zwei Schätzmethoden unter verschiedenen Bedingungen untersucht. Keines der Resultate war vollständig befriedigend. Die Anwendung eines separierten Quotientenschätzers war am vielversprechendsten. Der separierte Quotientenschätzer wurde anhand einer begrenzten Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben und anhand der Daten der OZ91 als Hilfsvariable konstruiert. Für dessen Verwendung müsste jedoch eine Kantonsgruppierung vorgenommen werden. Im Abschnitt 7 (Schätzung mit gewichtetem Mittel) hatten wir verschiedene Kantonsgruppierungen vorgenommen. Hieraus lässt sich schliessen, dass eine Schichtung in NUTS2, verbunden mit einer Schichtung anhand der landwirtschaftlichen Nutzfläche, den anderen untersuchten Schichtungsmethoden vorzuziehen ist. Es ist jedoch fraglich, ob eine solche Schichtung mit der vorgegebenen Stichprobengrösse von 200 Gemeinden in sämtlichen Schichten zu einer qualitativ befriedigenden Schätzung ($cv < 5\%$) führt, denn wir mussten auch bei dieser Schichtungsmethode die Stichprobengrösse in zwei Schichten auf 4 Gemeinden anheben.
Diese Schätzmethoden wurden aufgrund des Abbruchs des Projekts nicht zu Ende geführt. Inwiefern die Datenqualität einen Einfluss auf die Schätzer hat, konnte nicht untersucht werden. Es scheint jedoch nicht unrealistisch zu sein, dass anhand einer Stichprobe von Gemeinden, interessante Schätzungen der Anzahl Obstbäume in nicht landwirtschaftlichen Betrieben erhalten werden können. Dabei müsste jedoch die Stichprobengrösse grösser als 200 sein, um in sämtlichen Schichten einen Variationskoeffizienten von weniger als 5% zu erhalten. Bei einer Stichprobengrösse von rund 750 Gemeinden kann mit einer recht guten Qualität der Schätzung gerechnet werden. Dies entspricht einem Drittel der Gemeinden, in welchen eine Vollerhebung gemacht werden müsste. Ein gangbarer Mittelweg wäre vermutlich eine moderate Erhöhung der Stichprobe zusammen mit einer Reduktion der Ansprüche an die Genauigkeit in einigen Kantonen.
Wenn auf Resultate pro Kanton verzichtet wird, dann wird eine kleinere Stichprobe benötigt. Es ist dennoch angebracht eine Schichtung vorzunehmen; denn bei einer guten Schichtung verringert sich die Varianz und somit die nötige Stichprobengrösse, um eine vorgegebene Genauigkeit zu erhalten.

Tabelle 10: Für die Schätzung des Totals wurde ein separierter Quotientenschätzer verwendet, der anhand einer begrenzte Prognose der Baumanzahl in nicht landwirtschaftlichen Betrieben und der Hilfsvariablen 'Anzahl Bäume in der OZ91' gebildet wurde. Die Schichtgrösse wurde mit N_h, die Stichprobengrösse pro Schicht mit n_h und der Variationskoeffizient des Schätzers des Schichttotals mit cv_h bezeichnet. In der letzten Zeile wurde das Total der Grundgesamtheit, sowie der gesamten Stichprobengrösse und der Variationskoeffizient des Totalschätzers über alle Schichten hinweg angegeben.

Kanton	N_h	n_h	cv_h [%]
ZH	171	11	11.29
BE	358	30	7.63
LU	104	8	12.19
UR	11	4	0.75
SZ	20	4	8.77
OW/NW	17	4	9.91
GL	15	4	1.64
ZG	11	4	0.99
FR	209	8	20.29
SO	126	14	5.78
BS/BL	60	8	4.56
SH	34	4	2.58
AI/AR	12	4	1.21
SG	79	9	4.92
GR	9	4	0.37
AG	216	23	7.83
TG	63	7	6.01
TI	166	10	3.48
VD	356	15	8.15
VS	31	13	0.94
NE	37	4	1.99
GE	38	4	4.03
JU	59	5	8.37
Total	2202	201	1.76

Tabelle 11: Die Stichprobengrösse wurde auf 751 Gemeinden erweitert. Für die Schätzung des Totals wurde ein separierter Quotientenschätzer verwendet, der anhand einer begrenzte Prognose der Baumanzahl in nicht landwirtschaftlichen Betrieben und der Hilfsvariablen 'Anzahl Bäume in der OZ91' gebildet wurde. Die Schichtgrösse wurde mit N_h, die Stichprobengrösse pro Schicht mit n_h und der Variationskoeffizient des Schätzers des Schichttotals mit cv_h bezeichnet. In der letzten Zeile wurde das Total der Grundgesamtheit, sowie der gesamten Stichprobengrösse und der Variationskoeffizient des Totalschätzers über alle Schichten hinweg angegeben.

Kanton	N_h	n_h	cv_h [%]
ZH	171	49	3.91
BE	358	132	4.77
LU	104	35	4.97
UR	11	4	0.75
SZ	20	12	4.47
OW/NW	17	10	4.04
GL	15	4	1.64
ZG	11	4	0.99
FR	209	35	5.95
SO	126	63	1.98
BS/BL	60	34	2.10
SH	34	9	1.83
AI/AR	12	4	1.21
SG	79	39	3.30
GR	9	6	0.70
AG	216	101	2.44
TG	63	31	3.80
TI	166	45	1.52
VD	356	67	4.05
VS	31	31	0.00
NE	37	11	1.69
GE	38	4	4.03
JU	59	21	3.94
Total	2202	751	0.76

Anhang

A Variablenliste

Tabelle 12: Variablenliste des Datensatzes OZ2001, der unter Access mit "Gemeinde-Info-OZ2001"geliefert wurde.

	Variable	Beschreibung
1	Kanton	Nummerierung 1 bis 26
2	AKZ.Kanton	Autokennzeichen des Kantons
3	Bezirk	
4	Gmde.Nr	Gemeinde Nummer
5	Name	Gemeindename
6	inOZ2001	In der OZ2001 vorgesehen 1=ja
7	OZ91.Befreit	Gemeinde in der OZ91 befreit, 1=ja
8	OZ81.B.ume	Bäume in der OZ1981
9	OZ91.B.ume	Bäume in der OZ91
10	DZ99.B.ume	Bäume von Betrieben mit Direkt-zahlungsbeiträgen 1999
11	Betriebe.DZ99	Anzahl Betriebe mit Direkt-zahlungsbeiträgen 1999
12	Diff.81.91	Differenz in Prozent OZ81-OZ91
13	Sch.tzung.2001	Schätzung für 2001 aus Trend 81-91
14	DZ99AnteilOZ01	Anteil-DZ99-Bäume in Prozent von OZ2001 Schätzung
15	ObstinASCH97	Obstbäume in der Arealstatistik, 1=ja
16	BZ2000.Betriebezusatz	Betriebe mit Bäumen in der BZ2000-Zusatz
17	BZ2000.ZusatzB.ume	Bäume aus der BZ2000-Zusatz
18	.uber1000m.M	ganzes Gemeindegebiet über 1000mü.M., 1=ja
19	DZ99oderOZ91	Bäume in OZ91 oder in DZ99, > 0 entspricht ja
20	A.DZ99oderOZ91oderASCH	Bäume in OZ91 oder in DZ99 oder in ASCH, > 0 entspricht ja
21	Typ.Gemeinde	Gemeinde-Typologie: A-C in OZ2001 Grundge samtheit, D-E nicht in OZ2001 Grundgesamtheit
22	Total.Fl.che.in.Aren	Total Gemeindefläche nach Arealstatistik
23	LN.in.Aren	Landw. Nutzfläche nach Eidg. Betriebszählung
24	NWiesenWeidenTotal.in.Aren	Total Naturwiesen und Weiden nach Eidg. Betriebszählung
25	DZ.nach.Gmde	Kantone welche in den DZ die Bäume in der Standortgemeinde erfassen, 1=ja

B Graphiken

B.1 Prognose anhand der OZ81, der OZ91 und der DZ99

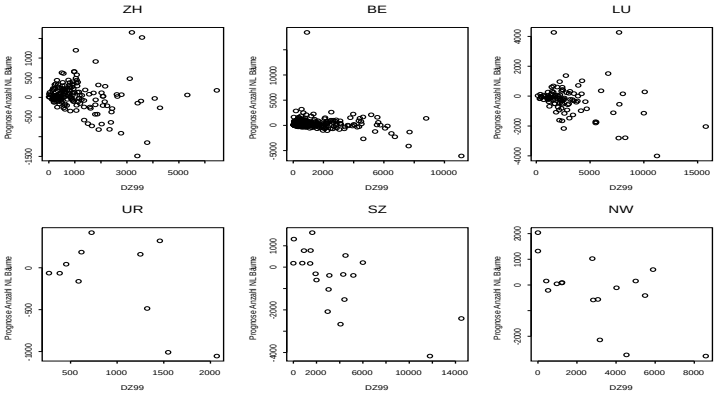


Abbildung 3: Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen die Prognose der gesamten Baumzahl der DZ99 gezeichnet. NW steht für NW und OW zusammen.Die Prognosen basieren auf den Daten der OZ81, OZ91 und der DZ99.

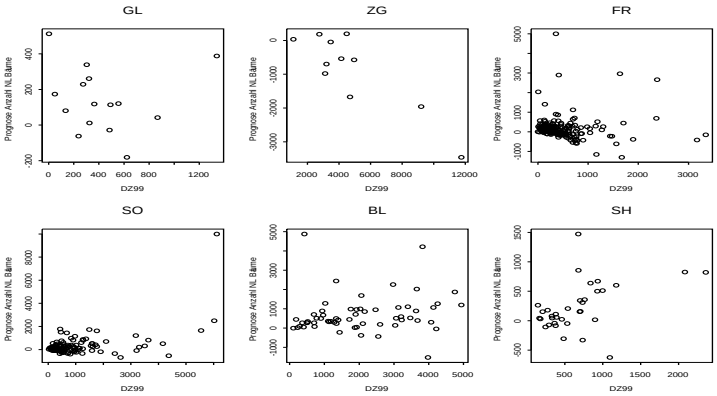


Abbildung 4: Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen die Prognose der gesamten Baumzahl der DZ99 gezeichnet. BL steht für BL und BS.Die Prognosen basieren auf den Daten der OZ81, OZ91 und der DZ99.

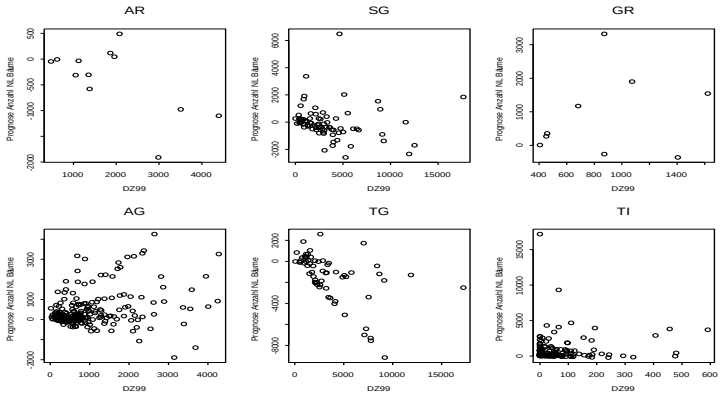


Abbildung 5: Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen die Prognose der gesamten Baumzahl der DZ99 gezeichnet. AR steht für AR und AI zusammen.Die Prognosen basieren auf den Daten der OZ81, OZ91 und der DZ99.

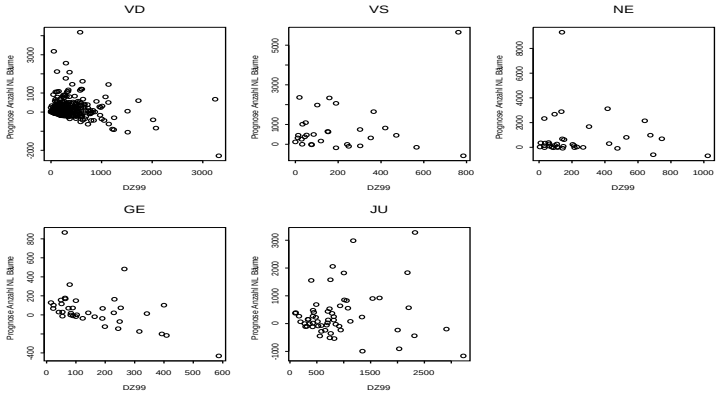


Abbildung 6: Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen die Prognose der gesamten Baumzahl der DZ99 gezeichnet.Die Prognosen basieren auf den Daten der OZ81, OZ91 und der DZ99.

B.2 Prognose anhand der OZ51 bis OZ91

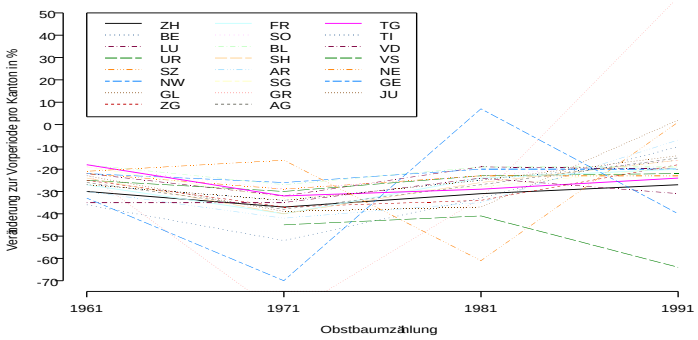


Abbildung 7: Trend der Veränderung der Anzahl Bäume in der Obstbaumzählung zwischen 1961 und 1991 in den Kantonen, gegenüber der Anzahl der vorhergehenden Obstbaumzählung.

B.3 Begrenzte Prognose

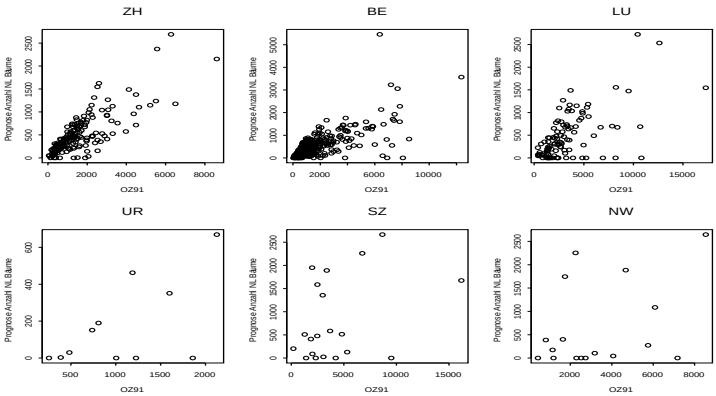


Abbildung 8: Begrenzte Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen Anzahl Bäume der OZ91 gezeichnet. Die Kantone Ob- und Nidwalden wurden zusammen gelegt (NW).

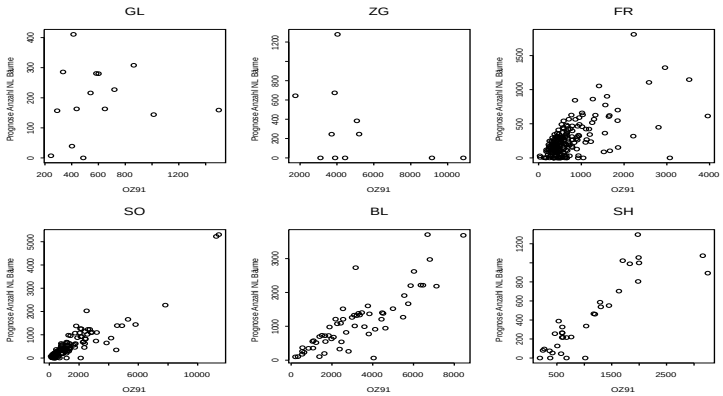


Abbildung 9: Begrenzte Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen Anzahl Bäume der OZ91 gezeichnet. Die Kantone Basel Land und Basel Stadt wurden zusammen gelegt (BL).

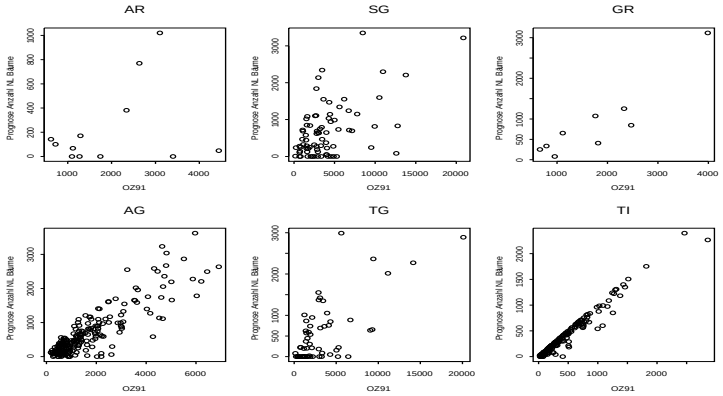


Abbildung 10: Begrenzte Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen Anzahl Bäume der OZ91 gezeichnet. Die Kantone AI und AR wurden zu AR zusammen gefasst.

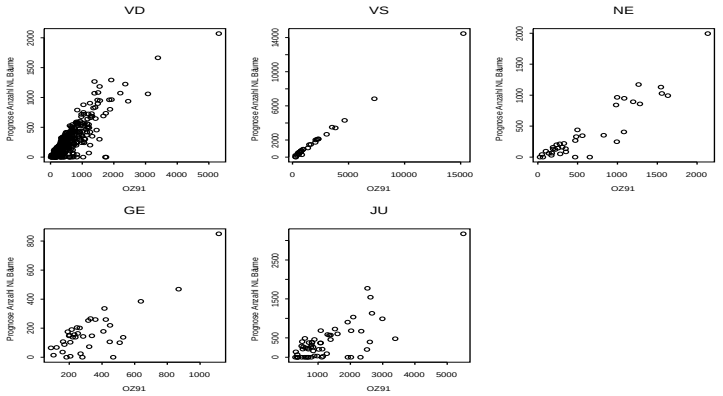


Abbildung 11: Begrenzte Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben (NL) gegen Anzahl Bäume der OZ91 gezeichnet.

C S-PLUS Programme

Wir haben nur die für diesen Bericht notwendigen S-PLUS Programme eingefügt. Diese wurden unter Windows auf S-PLUS 2000 entwickelt. Zuerst geben wir das Script von allgemeinen Stichprobenplanfunktionen wieder. Diese wurden im Script des zweiten Abschnitts mehrmals verwendet. Im zweiten Abschnitt sind die Programme für die eigentlichen Berechnungen des vorliegenden Berichts aufgeführt.

C.1 Stichprobenplan Funktionen

```
### Allgemeine Funtionen zu Stichprobenplänen:
### 1) 'neyman' bildet eine Neyman Zuordnung
### 2) 'propAlloc' bildet eine proportionale Zuordnung
### 3) 'stichprobe' zieht eine Zufallsstichprobe
### 4) 'estim.ht' berechnet die Horvitz-Thompson Schätzer
###    für den Erwartungswert und die Varianz für
###    geschichtete Stichproben mit einfacher Ziehung
###    innerhalb der Schichten
### 5) 'cumWurzFreq' berechnet die Schichtgrössen anhand
###    der Dalenius Methode (oder kummulierte Wurzel-
###    Frequenz Methode)

#neyman berechnet eine "Neyman-allocation", bei der die
#Stichprobengrösse 'n' gegeben ist. 'variable' ist die
#Erhebungsvariable, 'schichtvar' ist die Schichtungs-
#variable (optional), 'schichtN' ist die Anzahl Elemente
#jeder Schicht (optional, wenn angegeben, dann >1!)
#und 'minH' gibt eine minimale Schichtgrösse an(optional)
#es muss entweder 'schichtvar' oder 'schichtN' gegeben
#werden 'maxH' gibt eine maximale Schichtgrösse pro
#Schicht an (optional), wenn maxH=c(0), dann wird die
#maximale Schichtgrösse automatisch durch die Schicht-
#grösse begrenzt.
neyman_function(n,variable,schichtvar=c(-1),
       schichtN=c(1),minH=0,maxH=c(-1))
  {rowDim_length(variable)
    if (((schichtN==c(1))[1])&&((schichtvar!=c(-1))[1]))
    {
        N.h_rep(1,rowDim)
        n.num_0
        j_1
        for (i in 2:rowDim){
            if (schichtvar[i] == schichtvar[i-1])
                N.h[j]_N.h[j]+1
            else
                j_j+1
        }
        n.num_j
    }
    else if ((schichtN==c(1))[1])
    {
        return(NA)
    }
    else
    {
        N.h_schichtN
        n.num_length(schichtN)
    }
    if (maxH==c(0))
        maxH_N.h[1:n.num]
    s.h_rep(0,n.num)
    j_1
    for (i in 1:n.num)
    {
        s.h[i]_sqrt(var(variable[j:(j+N.h[i]-1)],
                                          unbiased=T))
        j_j+N.h[i]
    }
    n.h_rep(0,n.num)
    totNeym_as.numeric(N.h[1:n.num]%*%s.h)
    n.h_n*N.h[1:n.num]*s.h/totNeym
    n2_n
    if (min(maxH)>0)
    {
        index2_rep(1,n.num)
        for (i in 1:n.num)
            if (n.h[i]>maxH[i])
            {
                n2_n2-maxH[i]
                index2[i]_0
            }
        totNeym_as.numeric((index2*N.h[1:n.num])%*%s.h)
        n.h_n2*(index2*N.h[1:n.num])*s.h/totNeym
        n.h_n.h+((1-index2)*maxH)
    }
    if (minH>0)
    {
        index1_rep(1,n.num)
        for (i in 1:n.num)
            if (n.h[i]<minH)
            {
                if (N.h[i]>=minH)
                    n2_n2-minH
                else
                    n2_n2-N.h[i]
                index1[i]_0
            }
        if (min(maxH)>0)
        {
            totNeym_as.numeric((index2*(index1*
                                N.h[1:n.num]))%*%s.h)
            n.h_n2*(index2*(index1*N.h[1:n.num]))*s.h/
                                              totNeym
            n.h_n.h+((1-index1)*minH)+((1-index2)*maxH)
        }
        else
        {
            totNeym_as.numeric((index1*N.h[1:n.num])%*%s.h)
            n.h_n2*(index1*N.h[1:n.num])*s.h/totNeym
            n.h_n.h+((1-index1)*minH)
        }
        for (i in 1:n.num)
            if (N.h[i]<minH)
                n.h[i]_N.h[i]
    }
    x_matrix(ncol=3,nrow=n.num)
    dimnames(x)_list(NULL,c("N.h","n.h","S.h"))
    x[,1]_N.h[1:n.num]
    x[,2]_n.h
    x[,3]_s.h
    return(x)
}

#propAlloc berechnet eine proportionelle Zuordnung
propAlloc_function(n,variable,schichtvar,minH=0){
    s_sum(variable)
    rowDim_length(variable)
    N.h_rep(1,rowDim)
    nDim_1
    for (i in 2:rowDim)
        if (schichtvar[i]==schichtvar[i-1])
            N.h[nDim]_N.h[nDim]+1
        else
            nDim_nDim+1
    n.h_rep(0,nDim)
    j_1
    n2_n
    index1_rep(1,nDim)
    for (i in 1:nDim)
    {
        n.h[i]_n*sum(variable[j:(j+N.h[i]-1)])/s
        j_j+N.h[i]
```

```
        if (n.h[i]<minH)
        {
            n2_n2-minH
            index1[i]_0
        }
    }
    if (minH>0)
    {
        s1_0
        s2_rep(0,nDim)
        j_1
        lim_N.h[j]
        for (i in 1:rowDim)
            if (i<=lim)
            {
                if (index1[j]>0)
                {
                    s2[j]_s2[j]+variable[i]
                    s1_s1+variable[i]
                }
            }
            else if (i<rowDim)
            {
                j_j+1
                lim_lim+N.h[j]
                if (index1[j]>0)
                {
                    s2[j]_s2[j]+variable[i]
                    s1_s1+variable[i]
                }
            }
        n.h_n2*s2/s1+minH*(1-index1)
    }
    x_matrix(ncol=2,nrow=nDim)
    x[,1]_N.h[1:nDim]
    x[,2]_n.h
    return(x)

}

#Stichprobe kreiert eine Stichprobe anhand der Schicht-
#grössen und der Anzahl Elemente pro Schicht in der
#Stichprobe.
#'data' bezeichnet den Datensatz, 'N2.h' die Schicht-
#grösse und 'nNeymanKt2.h' bezeichnet die Stichproben-
#grösse der Schicht h 'stichVect' ist der Vektor mit der
#Schichtvariablen
stichprobe_function(data,N2.h,nNeymanKt2.h,stichVect=0,
            seed=205){
    set.seed(seed)
    print(seed)
    nKt_length(nNeymanKt2.h)
    if (stichVect != 0)
    {
        stichpNey_matrix(ncol=(dim(data)[2]+1),
                                nrow=sum(nNeymanKt2.h))
        dimnames(stichpNey)_list(NULL,
                    c(dimnames(data)[[2]],"itemNo"))
    }
    else
    {
        stichpNey_matrix(ncol=dim(data)[2],
                                nrow=sum(nNeymanKt2.h))
        dimnames(stichpNey)_list(NULL,dimnames(data)[[2]])
    }
    stichpNey_as.data.sheet(stichpNey)
    index1_0
    index2_0
    for (i in 1:nKt)
    {
        x_rep(0,nNeymanKt2.h[i])
        y_rep(0,N2.h[i])
        y_runif(N2.h[i],0,1)
        x_order(y)[1:nNeymanKt2.h[i]]
        x_x[order(x)]
        stichpNey[(index1+1):(index1+nNeymanKt2.h[i]),]_
                                        data[(index2+x),]
        if (stichVect != 0)
            stichpNey[(index1+1):(index1+nNeymanKt2.h[i]),
                (dim(data)[2]+1)]_index2+x
        index1_index1+nNeymanKt2.h[i]
        index2_index2+N2.h[i]
    }
    return(stichpNey)
}

# Horvitz-Thompson Schätzer

#'varible' bezeichnet die Erhebungsvariable, 'N2.h' die
#Schichtgrösse und 'nNeymanKt2.h' bezeichnet die Stich-
#probengrösse der Schicht h 'alpha' ist der Fehler
#erster Ordnung für das Konfidenzintervall
estim.ht_function(variable,N2.h,n.h,alpha=0.05){
    nKt_length(n.h)
    qalpha_qnorm(1-alpha/2)
    # Horvitz-Thompson Schätzer für das Total
    mittelH_rep(0,nKt)
    index1_1
    for (i in 1:nKt)
    {
        mittelH[i]_mean(variable[index1:(index1+n.h[i]-1)])
        index1_index1+n.h[i]
    }
    totBNeyman_as.integer(round(mittelH%*%N2.h,0))
    totBNeyman.h_mittelH*N2.h

    ##### Horvitz-Thompson Schätzer für die Varianz des
    ##### geschätzten Totals
    empVar_rep(0,nKt)
    index1_1
    for (i in 1:nKt)
    {
        x_sum((as.numeric(variable[index1:(index1+n.h[i]-
        1)])-mittelH[i])^2)
        x_x/(n.h[i]-1)
        empVar[i]_x
        index1_index1+n.h[i]
    }
    esBKt2.h_sqrt(empVar*N2.h*(N2.h-n.h)/n.h)
    varTotKt2_sum(esBKt2.h^2)

    cvKt2_sqrt(varTotKt2)/totBNeyman
    cvKt2.h_esBKt2.h/totBNeyman.h

    x_matrix(data=0,ncol=10,nrow=nKt)
    dimnames(x)_list(NULL,c("N.h","n.h","t.h","cv.h [%]",
                        "l.h","u.h","t","cv [%]","l","u"))
    x[,1]_N2.h
    x[,2]_n.h
    x[,3]_totBNeyman.h
    x[,4]_round(100*cvKt2.h,2)
    x[,5]_totBNeyman.h-qalpha*esBKt2.h
    x[,6]_totBNeyman.h+qalpha*esBKt2.h
    x[1,7]_totBNeyman
    x[1,8]_round(100*cvKt2,2)
    x[1,9]_totBNeyman-qalpha*sqrt(varTotKt2)
    x[1,10]_totBNeyman+qalpha*sqrt(varTotKt2)
    x_as.data.sheet(x)
    return(x)

}



### CumWurzelFrequenz Funktion: "variable" ist die
### Schichtungsvariable (nach den bestehenden Schichten
### geordnet!!), "nStrata" die Anzahl bestehender
### Schichten, "elPerStrata" der Vektor mit der Schicht-
### grösse der bestehenden Schichten, "nInStrata" der
### Vektor mit der Anzahl zu bildender Schichten inner-
### halb der bestehenden Schichten (Anzahl Schichten in
### der ersten Schicht = erstes Element von nInStrata,
### etc.)
### "nF" Anzahl initialer Schichten
### ("nF" >> "nInStrata[i]" für jedes i !!!)
cumWurzFreq_function(variable,nStrata=1,elPerStrata=c(1),
                                nInStrata=c(1),nF=1){
    if (length(nInStrata)!=nStrata)
        return(NA)
    strataLimite_matrix(nrow=length(variable),ncol=3)
    cnames_c("L","N","ordre")
    r2names_c(1:sum(nInStrata))
    for (i in 1:nStrata)
    {
        r2names[((i>1)*sum(nInStrata[1:i-1])+1):
        sum(nInStrata[1:i])]_paste("L,N",1:nInStrata[i])
    }
    rnames_c(paste(rep(paste("strate",1:nStrata),
                times=nInStrata),r2names),rep("",
                times=(length(variable)-sum(nInStrata))))
    dimnames(strataLimite)_list(rnames,cnames)
    strataLimite_as.data.sheet(strataLimite)
    q_1
    for (p in 1:nStrata)
    {
```

```
    y_variable[q:(q+elPerStrata[p]-1)]
    x_order(y)
    y_y[x]
    yL_max(y)
    y0_y[1]
    long_(yL-y0)/nF
    freq_rep(0,nF)
    i_1
    k_1
    while (i <= length(y))
    {
        if (y[i] <= (y0+k*long))
            freq[k]_freq[k]+1
        else
        {
            k_k+1
            freq[k]_freq[k]+1
        }
        i_i+1
    }
    cumSqrtFreq_rep(0,nF)
    cumSqrtFreq[1]_sqrt(freq[1])
    for (i in 2:nF)
        cumSqrtFreq[i]_sum(sqrt(freq[1:i]))
    lCumSqrtFreq_(cumSqrtFreq[nF]-
                          cumSqrtFreq[1])/nInStrata[p]
    strataNumber_rep(0,nInStrata[p])
    i_1
    j_1
    while ((i <= nF)&&(j<=nInStrata[p]))
    {
        while (cumSqrtFreq[i]<
                        (cumSqrtFreq[1]+j*lCumSqrtFreq))
            i_i+1
        strataNumber[j]_sum(freq[1:i])
        j_j+1
    }
    p2_(max(nInStrata)+nInStrata[p])
    if (p>1)
        p3_sum(nInStrata[1:(p-1)])+1
    else p3_1
    p4_sum(nInStrata[1:p])
    p5_c(0,strataNumber[1:nInStrata[p]][-nInStrata[p]])
    strataLimite[p3:p4,1]_y[strataNumber]
    strataLimite[p3:p4,2]_strataNumber[1:nInStrata[p]]
                                                    -p5
    strataLimite[q:(q+elPerStrata[p]-1),3]_q+x-1
    q_q+elPerStrata[p]
}
return(strataLimite)
}
```

C.2 Erarbeitung des Stichprobenplans

```
############################################################
###                                                      ###
### OZ2001: Berechnungen zum Stichprobenplan Schätzung###
### anhand des gewichteten Mittels und anhand eines   ###
### und anhand eines Quotientenschätzers (mit         ###
### vorhergesagten Daten)                             ###
### Autor: Daniel Kilchmann, METH                     ###
### Letzte Aenderung: 23.05.2001                      ###
###                                                      ###
############################################################
#  Es werden nur die Programme aufgeführt, welche für   #
#  den Methodenbericht relevant sind.                   #
#  Die Daten wurden im Access Format geliefert, wir     #
#  haben sie in ein Excel exportiert, um sie dann       #
#  Menugesteuert in S-PLUS (Windows) zu importieren.    #
#  Aus technischen Gründen mussten wir einige           #
#  Variablennamen ändern, die daraus resultierenden     #
#  Daten wurden wiederum unter Excel gespeichert:       #
#  1. Datenlieferung: "Gemeinde99.xls"                  #
#  2. Datenlieferung: "OZ2001.xls"                      #
#  Im weiteren haben wir unter S-PLUS eine Trendmatrize #
#  kreiert, die als "trend.sdd" abgespeichert wurde.    #
#  Es sollte darauf geachtet werden, dass zuerst säm-   #
#  tliche Funktionen, die in 'functions.ssc' aufgeführt #
#  sind durchlaufen zu lassen, erst dann kann           #
#  'final.ssc' durchgeführt werden.                     #
############################################################

# Stichprobengrösse
n_200
# Anzahl Kantone
nKt_26
# Kantone nummeriert
ktNr_seq(from=1, to=nKt, by=1)
# Abkürzungen der Kantone
ktNamen_vector(mode="character",length=nKt)
# Anzahl Gemeinden in OZ01
M_dim(Gemeinde99)[1]
ktNamen[1]_as.character(Gemeinde99$AKZ.Kanton[1])
j_2
for (i in 2:M)
{
    if (Gemeinde99$AKZ.Kanton[i] != ktNamen[j-1])
    {
        ktNamen[j]_as.character(Gemeinde99$AKZ.Kanton[i])
        j_j+1
    }
}

N_1
interm1_rep(0,times=M)
for (i in 1:M)
    if (Gemeinde99$inOZ2001[i] == 1)
    {
        interm1[N]_i
        N_N+1
    }
N_N-1
interm2_rep(0,times=N)
interm2_interm1[1:N]

### Bereinigung der Daten, indem die Gemeinden, welche
### nicht in die OZ01 einbezogen werden sollen,
### ausgeschlossen werden
Gmde99_matrix(ncol=dim(Gemeinde99)[2],nrow=N)
dimnames(Gmde99)_list(NULL,dimnames(Gemeinde99)[[2]])
Gmde99_as.data.sheet(Gmde99)
Gmde99_Gemeinde99[interm2,]

# Total der Landw. Nutzzone
TotLN_sum(as.numeric(Gmde99$LN.in.Aren))

# Anzahl Gemeinden pro Kt
N.h_rep(1,nKt)
j_1
for (i in 2:N)
{
    if (Gmde99$Kanton[i] == Gmde99$Kanton[i-1])
        N.h[j]_N.h[j]+1
    else
        j_j+1
}

ktNom_vector("character",23)
ktNom_ktNamen[-c(6,12,16)]

################
### Kreierung der Gemeinden, die Daten in der DZ99 hatten
### und deren landw. Nutzfläche bekannt ist, Zusammen-
### legung der Halbkantone!!
################
N2.h_rep(0,nKt-3)
N2.h[c(1:5,7:10,12,14:23)]_N.h[c(1:5,8:11,14,17:26)]
N2.h[c(6,11,13)]_N.h[c(6,12,15)]+N.h[c(7,13,16)]
indexOhneLN2_rep(0,times=dim(Gmde99)[[1]])
j_1
for (i in 1:dim(Gmde99)[[1]])
{
    if ((Gmde99$LN.in.Aren[i]!=0)&&
                (Gmde99$DZ99.B.ume[i]!=0))
    {
        indexOhneLN2[j]_i
        j_j+1
    }
}
```

```
N2ohneLN0.h_N2.h
k_0
for (i in 1:23)
{
    for (j in 1:N2.h[i])
    {
        if ((Gmde99$LN.in.Aren[k+j]==0)||
                    (Gmde99$DZ99.B.ume[k+j]==0))
            N2ohneLN0.h[i]_N2ohneLN0.h[i]-1
    }
    k_k+N2.h[i]
}
indexOhneLN_rep(0,times=sum(N2ohneLN0.h))
indexOhneLN_indexOhneLN2[1:sum(N2ohneLN0.h)]

Gmde99oLN0_matrix(ncol=dim(Gemeinde99)[2],
                          nrow=sum(N2ohneLN0.h))
dimnames(Gmde99oLN0)_list(NULL,
                       dimnames(Gemeinde99)[[2]])
Gmde99oLN0_as.data.sheet(Gmde99oLN0)
Gmde99oLN0_Gmde99[indexOhneLN,]


### Einfache Stichprobe (geschichtetes Mittel),
### Kantonsschichten

# Neyman Zuordnung
nNeyKt.h_neyman(200,Gmde99oLN0$DZ99.B.ume,
            schichtvar=rep(c(1:23),times=N2ohneLN0.h),
                                          minH=4)
sNeyKt.h_stichprobe(Gmde99oLN0,nNeyKt.h[,1],
                                round(nNeyKt.h[,2],0))
estNeyKt.h_estim.ht(sNeyKt.h$DZ99.B.ume,nNeyKt.h[,1],
                                round(nNeyKt.h[,2],0))
dimnames(estNeyKt.h)_list(ktNom, dimnames(estNeyKt.h)[[2]])

# Zuordnung proportional zur landw. Nutzfläche
nPropLnKt23.h_propAlloc(201,Gmde99oLN0$LN.in.Aren,
                           rep(c(1:23),times=N2ohneLN0.h),4)
sLNprop.h_stichprobe(Gmde99oLN0,N2ohneLN0.h,
                                round(nPropLnKt23.h[,2],0))
estLNprop.h_estim.ht(sLNprop.h$DZ99.B.ume,N2ohneLN0.h,
                                round(nPropLnKt23.h[,2],0))
dimnames(estLNprop.h)_list(ktNom, dimnames(estLNprop.h)[[2]])

# Zuordnung proportional zur Anzahl Bäume in der Gemeinde
nPropBKt23.h_propAlloc(200,Gmde99oLN0$DZ99.B.ume,
                           rep(c(1:23),times=N2ohneLN0.h),4)
sBprop.h_stichprobe(Gmde99oLN0,N2ohneLN0.h,
                                round(nPropBKt23.h[,2],0))
estBprop.h_estim.ht(sBprop.h$DZ99.B.ume,N2ohneLN0.h,
                                round(nPropBKt23.h[,2],0))
dimnames(estBprop.h)_list(ktNom, dimnames(estBprop.h)[[2]])


###############
### Kantonsgruppierungen: 1) nach Vorschlag AGR, 2) NUTS2
###############

###Kantonsgruppen: 1. BL&BS, 2. LU&ZG,
###3. UR&OW&NW&SZ&GL&AI&AR
Gmde99oLn02_matrix(ncol=(1+dim(Gmde99oLN0)[[2]]),
                                nrow=dim(Gmde99oLN0)[[1]])
dimnames(Gmde99oLn02)_list(NULL,
                       c(dimnames(Gmde99oLN0)[[2]],"strat"))
Gmde99oLn02_as.data.sheet(Gmde99oLn02)
Gmde99oLn02[,1:20]_Gmde99oLN0[,1:20]
Ngr.h_rep(0,18)
for (i in 1:dim(Gmde99oLn02)[[1]])
{
    if ((Gmde99oLn02$AKZ.Kanton[i]=="BS")||
        (Gmde99oLn02$AKZ.Kanton[i]=="BL"))
    {
        Gmde99oLn02$strat[i]_1
        Ngr.h[1]_Ngr.h[1]+1
    }
    else if ((Gmde99oLn02$AKZ.Kanton[i]=="LU")||
        (Gmde99oLn02$AKZ.Kanton[i]=="ZG"))
    {
        Gmde99oLn02$strat[i]_2
        Ngr.h[2]_Ngr.h[2]+1
    }
    else if ((Gmde99oLn02$AKZ.Kanton[i]=="UR")||
        (Gmde99oLn02$AKZ.Kanton[i]=="OW")
        ||(Gmde99oLn02$AKZ.Kanton[i]=="NW")||
        (Gmde99oLn02$AKZ.Kanton[i]=="SZ")
        ||(Gmde99oLn02$AKZ.Kanton[i]=="GL")||
        (Gmde99oLn02$AKZ.Kanton[i]=="AI")||
        (Gmde99oLn02$AKZ.Kanton[i]=="AR"))
    {
        Gmde99oLn02$strat[i]_3
        Ngr.h[3]_Ngr.h[3]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="ZH")
    {
        Gmde99oLn02$strat[i]_4
        Ngr.h[4]_Ngr.h[4]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="BE")
    {
        Gmde99oLn02$strat[i]_5
        Ngr.h[5]_Ngr.h[5]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="FR")
    {
        Gmde99oLn02$strat[i]_6
        Ngr.h[6]_Ngr.h[6]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="SO")
    {
        Gmde99oLn02$strat[i]_7
        Ngr.h[7]_Ngr.h[7]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="SH")
    {
        Gmde99oLn02$strat[i]_8
        Ngr.h[8]_Ngr.h[8]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="SG")
    {
        Gmde99oLn02$strat[i]_9
        Ngr.h[9]_Ngr.h[9]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="GR")
    {
        Gmde99oLn02$strat[i]_10
        Ngr.h[10]_Ngr.h[10]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="AG")
    {
        Gmde99oLn02$strat[i]_11
        Ngr.h[11]_Ngr.h[11]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="TG")
    {
        Gmde99oLn02$strat[i]_12
        Ngr.h[12]_Ngr.h[12]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="TI")
    {
        Gmde99oLn02$strat[i]_13
        Ngr.h[13]_Ngr.h[13]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="VD")
    {
        Gmde99oLn02$strat[i]_14
        Ngr.h[14]_Ngr.h[14]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="VS")
    {
        Gmde99oLn02$strat[i]_15
        Ngr.h[15]_Ngr.h[15]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="NE")
    {
        Gmde99oLn02$strat[i]_16
        Ngr.h[16]_Ngr.h[16]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="GE")
    {
        Gmde99oLn02$strat[i]_17
        Ngr.h[17]_Ngr.h[17]+1
    }
    else if (Gmde99oLn02$AKZ.Kanton[i]=="JU")
    {
        Gmde99oLn02$strat[i]_18
        Ngr.h[18]_Ngr.h[18]+1
    }
}
Gmde99oLn02_Gmde99oLn02[order(Gmde99oLn02$strat),]
nNeyKtGr.h_neyman(200,Gmde99oLn02$DZ99.B.ume,
                    schichtvar=Gmde99oLn02$strat,minH=4)
sKtGrney.h_stichprobe(Gmde99oLn02,nNeyKtGr.h[,1],
                                    round(nNeyKtGr.h[,2],0))
```

```
estKtGrney.h_estim.ht(sKtGrney.h$DZ99.B.ume,
                      nNeyKtGr.h[,1],round(nNeyKtGr.h[,2],0))
dimnames(estKtGrney.h)_list(as.character(c(1:18)),
                                dimnames(estKtGrney.h)[[2]])

#########################################################
## NUTS2 als Gebietsbegrenzungen: Région lémanique (VD,##
##     VS, GE), Espace Mittelland (BE, FR, SO, NE, JU) ##
##     Nordwestschweiz (BS, BL, AG), Zürich (ZH),      ##
##     Ostschweiz (GL, SH, AR, AI, SG, GR, TG),        ##
##     Zentralschweiz (LU, UR, SZ, OW, NW, ZG),        ##
##     Ticino (TI)                                     ##
#########################################################
ozNUTS2_matrix(ncol=(1+dim(Gemeinde99)[2]),nrow=N)
dimnames(ozNUTS2)_list(NULL,c("NUTS2",
                                dimnames(Gemeinde99)[[2]]))
ozNUTS2_as.data.sheet(ozNUTS2)
ozNUTS2[,2:21]_Gmde99
# Anzahl NUTS2 Regionen
nNuts2_7
for (i in 1:N)
{
    if ((ozNUTS2$AKZ.Kanton[i] == "VD") ||
        (ozNUTS2$AKZ.Kanton[i] == "VS")
        || (ozNUTS2$AKZ.Kanton[i] == "GE"))
        ozNUTS2$NUTS2[i]_11
    else if ((ozNUTS2$AKZ.Kanton[i] == "BE") ||
        (ozNUTS2$AKZ.Kanton[i] == "FR")
        || (ozNUTS2$AKZ.Kanton[i] == "SO") ||
        (ozNUTS2$AKZ.Kanton[i] == "NE")
        || (ozNUTS2$AKZ.Kanton[i] == "JU"))
        ozNUTS2$NUTS2[i]_12
    else if ((ozNUTS2$AKZ.Kanton[i] == "BS") ||
        (ozNUTS2$AKZ.Kanton[i] == "BL")
        || (ozNUTS2$AKZ.Kanton[i] == "AG"))
        ozNUTS2$NUTS2[i]_13
    else if (ozNUTS2$AKZ.Kanton[i] == "ZH")
        ozNUTS2$NUTS2[i]_14
    else if ((ozNUTS2$AKZ.Kanton[i] == "GL") ||
        (ozNUTS2$AKZ.Kanton[i] == "SH")
        || (ozNUTS2$AKZ.Kanton[i] == "AR") ||
        (ozNUTS2$AKZ.Kanton[i] == "AI")
        || (ozNUTS2$AKZ.Kanton[i] == "SG") ||
        (ozNUTS2$AKZ.Kanton[i] == "GR")
        || (ozNUTS2$AKZ.Kanton[i] == "TG"))
        ozNUTS2$NUTS2[i]_15
    else if ((ozNUTS2$AKZ.Kanton[i] == "LU") ||
        (ozNUTS2$AKZ.Kanton[i] == "UR")
        || (ozNUTS2$AKZ.Kanton[i] == "SZ") ||
        (ozNUTS2$AKZ.Kanton[i] == "OW")
        || (ozNUTS2$AKZ.Kanton[i] == "NW")||
        (ozNUTS2$AKZ.Kanton[i] == "ZG"))
        ozNUTS2$NUTS2[i]_16
    else if (ozNUTS2$AKZ.Kanton[i] == "TI")
        ozNUTS2$NUTS2[i]_17
}
ozNUTS2_ozNUTS2[order(ozNUTS2$NUTS2),]
ozNUTS2$NUTS2_as.numeric(ozNUTS2$NUTS2)

# Neyman Zuordnung
nNeyNu2.h_neyman(200,ozNUTS2$DZ99.B.ume,
                     schichtvar=ozNUTS2$NUTS2,minH=4)
sNUTSney.h_stichprobe(ozNUTS2,nNeyNu2.h[,1],
                                round(nNeyNu2.h[,2],0))
estNUTSney.h_estim.ht(sNUTSney.h$DZ99.B.ume,
                      nNeyNu2.h[,1],round(nNeyNu2.h[,2],0))

## Schichtung nach der Grösse der landw. Nutzfläche,
## Gemeinden ohne Angabe der landw. Nutzfläche werden
## aus der Grundgesamtheit ausgeschlossen
ozNU2oLN0_matrix(ncol=(1+dim(Gmde99oLN0)[2]),
                                nrow=dim(Gmde99oLN0)[1])
dimnames(ozNU2oLN0)_list(NULL,c("NUTS2",
                                dimnames(Gmde99oLN0)[[2]]))
ozNU2oLN0_as.data.sheet(ozNU2oLN0)
ozNU2oLN0[,2:21]_as.data.sheet(Gmde99oLN0)
# Anzahl NUTS2 Regionen
nNuts2_7
for (i in 1:dim(ozNU2oLN0)[1])
{
    if ((ozNU2oLN0$AKZ.Kanton[i] == "VD") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "VS") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "GE"))
        ozNU2oLN0$NUTS2[i]_11
    else if ((ozNU2oLN0$AKZ.Kanton[i] == "BE") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "FR")
        || (ozNU2oLN0$AKZ.Kanton[i] == "SO") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "NE")
        || (ozNU2oLN0$AKZ.Kanton[i] == "JU"))
        ozNU2oLN0$NUTS2[i]_12
    else if ((ozNU2oLN0$AKZ.Kanton[i] == "BS") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "BL")
        || (ozNU2oLN0$AKZ.Kanton[i] == "AG"))
        ozNU2oLN0$NUTS2[i]_13
    else if (ozNU2oLN0$AKZ.Kanton[i] == "ZH")
        ozNU2oLN0$NUTS2[i]_14
    else if ((ozNU2oLN0$AKZ.Kanton[i] == "GL") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "SH")
        || (ozNU2oLN0$AKZ.Kanton[i] == "AR") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "AI")
        || (ozNU2oLN0$AKZ.Kanton[i] == "SG") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "GR")
        || (ozNU2oLN0$AKZ.Kanton[i] == "TG"))
        ozNU2oLN0$NUTS2[i]_15
    else if ((ozNU2oLN0$AKZ.Kanton[i] == "LU") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "UR")
        || (ozNU2oLN0$AKZ.Kanton[i] == "SZ") ||
        (ozNU2oLN0$AKZ.Kanton[i] == "OW")
        || (ozNU2oLN0$AKZ.Kanton[i] == "NW")||
        (ozNU2oLN0$AKZ.Kanton[i] == "ZG"))
        ozNU2oLN0$NUTS2[i]_16
    else if (ozNU2oLN0$AKZ.Kanton[i] == "TI")
        ozNU2oLN0$NUTS2[i]_17
}
ozNU2oLN0_ozNU2oLN0[order(ozNU2oLN0$NUTS2),]

# Anzahl Gemeinden pro NUTS2 Region
Nn2oLn0.h_rep(0,nNuts2)
for (i in 1:dim(ozNU2oLN0)[1])
{
    if (ozNU2oLN0$NUTS2[i] == 11)
        Nn2oLn0.h[1]_Nn2oLn0.h[1]+1
    else if (ozNU2oLN0$NUTS2[i] == 12)
        Nn2oLn0.h[2]_Nn2oLn0.h[2]+1
    else if (ozNU2oLN0$NUTS2[i] == 13)
        Nn2oLn0.h[3]_Nn2oLn0.h[3]+1
    else if (ozNU2oLN0$NUTS2[i] == 14)
        Nn2oLn0.h[4]_Nn2oLn0.h[4]+1
    else if (ozNU2oLN0$NUTS2[i] == 15)
        Nn2oLn0.h[5]_Nn2oLn0.h[5]+1
    else if (ozNU2oLN0$NUTS2[i] == 16)
        Nn2oLn0.h[6]_Nn2oLn0.h[6]+1
    else if (ozNU2oLN0$NUTS2[i] == 17)
        Nn2oLn0.h[7]_Nn2oLn0.h[7]+1
}
Nn2oLn0.h_as.integer(Nn2oLn0.h)
ozNU2oLN0_ozNU2oLN0[order(ozNU2oLN0$NUTS2),]

# 4 Schichten (anhand der landw. Nutzfläche) innerhalb
# jeder NUTS2 Schicht
nNuts2.4h_cumWurzFreq(as.numeric(ozNU2oLN0$LN.in.Aren),7,
                          Nn2oLn0.h,rep(4,times=7),1000)
nNeyNu2.4h_as.data.sheet(neyman(200,
                        ozNU2oLN0$DZ99.B.ume[nNuts2.4h$ordre],
    schichtvar=rep(1:28,times=nNuts2.4h$N[1:28]),minH=4))
sNUTSney2.4h_stichprobe(ozNU2oLN0[nNuts2.4h$ordre,],
                        nNeyNu2.4h$N.h,round(nNeyNu2.4h$n.h,0))
estNUTSney2.4h_estim.ht(sNUTSney2.4h$DZ99.B.ume,
    nNeyNu2.4h$N.h,round(nNeyNu2.4h$n.h,0))

# Unterschiedliche Anzahl Schichten anhand der landw.
# Nutzfläche innerhalb der NUTS2 Schichten
# Anzahl Schichten innerhalb der NUTS2 Schichten
nGmdeK1Nu2_c(2,4,2,1,3,2,1)

nNuts2.mh_cumWurzFreq(ozNU2oLN0$LN.in.Aren,7,Nn2oLn0.h,
                                        nGmdeK1Nu2,1000)
nNeyNu2.mh_neyman(200,
                      ozNU2oLN0$DZ99.B.ume[nNuts2.mh$ordre],
                          schichtN=nNuts2.mh$N[1:15],minH=4)
nNeyNu2.mh[,1:3]_as.integer(round(nNeyNu2.mh[,1:3],0))
nNeyNu2.mh_as.data.sheet(nNeyNu2.mh)
sNoLNney.mh_stichprobe(ozNU2oLN0[nNuts2.mh$ordre,],
                               nNeyNu2.mh$N.h,nNeyNu2.mh$n.h)
estNoLNney.mh_estim.ht(sNoLNney.mh$DZ99.B.ume,
                               nNeyNu2.mh$N.h,nNeyNu2.mh$n.h)

#########################################################
#########################################################
##                                                     ##
## Schätzungsdatensatz basierend auf 2. Datenlieferung! ##
##                                                     ##
#########################################################
#########################################################
```

```
## Bereinigung der Daten (nur Daten die in OZ01 vorkommen
## beibehalten)
M_dim(OZ2001)[1]
N_0
index2_rep(0,dim(OZ2001)[2])
j_1
for (i in 1:M)
    if ((OZ2001$inOZ2001[i] == 1)&&
        (OZ2001$OZ91.Befreit[i] == 0)&&
        (OZ2001$Typ.Gemeinde[i] !="D")&&
        (OZ2001$Typ.Gemeinde[i] !="E")&&
        (!is.na(OZ2001$Diff.81.91[i])))
    {
        N_N+1
        index2[j]_i
        j_j+1
    }
index_rep(0,N)
index_index2[1:N]

OZ01_matrix(ncol=dim(OZ2001)[2],nrow=N)
dimnames(OZ01)_list(NULL,dimnames(OZ2001)[[2]])
OZ01_as.data.sheet(OZ01)
OZ01_OZ2001[index,]

for (i in 1:N)
{
    if ((OZ01$DZ99.B.ume[i]==0)||
        (is.na(OZ01$DZ99.B.ume[i])))
        OZ01$DZ99AnteilOZ01[i]_0
}
for (i in 1:N)
{
    if (OZ01$Sch.tzung.2001[i]==0)
        OZ01$DZ99AnteilOZ01[i]_100
}

Nkt_rep(1,26)
ktName_vector("character",26)
ktName[1]_as.character(OZ01$AKZ.Kanton[1])
j_1
for (i in 2:N)
{
    if (OZ01$Kanton[i]==OZ01$Kanton[i-1])
        Nkt[j]_Nkt[j]+1
    else {
        j_j+1
        ktName[j]_as.character(OZ01$AKZ.Kanton[i])
    }
}
ktNom_vector("character",23)
ktNom_ktName[-c(6,12,16)]
Kt.h_rep(0,23)
Kt.h_Nkt[-c(6,12,16)]
Kt.h[6]_Kt.h[6]+Nkt[6]
Kt.h[11]_Kt.h[11]+Nkt[12]
Kt.h[13]_Kt.h[13]+Nkt[16]

#Begrenzte Prognose
ymax_pmax(OZ01$OZ91.B.ume,OZ01$DZ99.B.ume)-
                                    OZ01$DZ99.B.ume
Nbegr.h_rep(1,26)
j_1
for (i in 2:dim(OZ01)[[1]])
    if (OZ01$AKZ.Kanton[i]==OZ01$AKZ.Kanton[i-1])
        Nbegr.h[j]_Nbegr.h[j]+1
    else
        j_j+1
Nbegr.h[6]_Nbegr.h[6]+Nbegr.h[7]
Nbegr.h[12]_Nbegr.h[12]+Nbegr.h[13]
Nbegr.h[15]_Nbegr.h[15]+Nbegr.h[16]
Nbegr.h[7:11]_Nbegr.h[8:12]
Nbegr.h[12:13]_Nbegr.h[14:15]
Nbegr.h[14:23]_Nbegr.h[17:26]
interm_rep(0,23)
interm_Nbegr.h[1:23]
Nbegr.h_rep(0,23)
Nbegr.h_interm
rm(interm)

n_200
neyBegrProg.h_neyman(n,ymax,schichtvar=rep(c(1:23),
                    times=Nbegr.h),minH=4,maxH=c(0))
sNeyBegrProg_stichprobe(OZ01,neyBegrProg.h[,1],
                    round(neyBegrProg.h[,2],0),seed=205)

# Separierter Quotientenschätzer für das Total Bäume in
# n. landw. Betr.
ymaxs_pmax(sNeyBegrProg$OZ91.B.ume,
        sNeyBegrProg$DZ99.B.ume)-sNeyBegrProg$DZ99.B.ume

tN.h_rep(0,23)
tL.h_rep(0,23)
tNrs.h_rep(0,23)
j_1
for (i in 1:23)
{
    tN.h[i]_sum(ymaxs[j:(j-1+
                        round(neyBegrProg.h[i,2],0))])
    tN.h[i]_neyBegrProg.h[i,1]*tN.h[i]/
                        round(neyBegrProg.h[i,2],0)
    tL.h[i]_sum(sNeyBegrProg$OZ91.B.ume[j:(j-1+
                        round(neyBegrProg.h[i,2],0))])
    tL.h[i]_neyBegrProg.h[i,1]*tL.h[i]/
                        round(neyBegrProg.h[i,2],0)
    j_j+round(neyBegrProg.h[i,2],0)
}

j_1
for (i in 1:23)
{
    tNrs.h[i]_sum(OZ01$OZ91.B.ume[j:(j+
            neyBegrProg.h[i,1]-1)])*tN.h[i]/tL.h[i]
    j_j+neyBegrProg.h[i,1]
}

# Residuen Berechnung
tt.h_rep(0,23)
j_1
for (i in 1:23)
{
    k_j+round(neyBegrProg.h[i,2],0)-1
    tt.h[i]_sum((ymaxs[j:k]-
            sNeyBegrProg$OZ91.B.ume[j:k]*tN.h[i]/tL.h[i])^2)
    j_k+1
}

# Varianzberechnung des Totals in n. landwirtschaftlichen
# Betr. anhand der Residuen, pro Schicht
f.h_rep(0,23)
SnRs.h_rep(0,23)
f.h_round(neyBegrProg.h[,2])/neyBegrProg.h[,1]
SnRs.h_sqrt(tt.h*(neyBegrProg.h[,1]^2)*(1-f.h)/
        (round(neyBegrProg.h[,2])*(neyBegrProg.h[,2]-1)))
j_1
for (i in 1:23)
{
    k_j+neyBegrProg.h[i,1]-1
    Trs.h_tNrs.h+sum(OZ01$OZ91.B.ume[j:k])
    j_k+1
}
CVtRs.h_matrix(ncol=1,nrow=23)
CVtRs.h_as.data.sheet(CVtRs.h)
dimnames(CVtRs.h)_list(ktNom,"cv.h")
CVtRs.h[,1]_100*SnRs.h/Trs.h
CVtRs_100*sqrt(sum(SnRs.h^2))/sum(Trs.h)

# Ohne Schichtung (nur Total für die Schweiz)
# kombinierter Quotientenschätzer
tNrc_sum(tN.h)*sum(OZ01$OZ91.B.ume) /sum(tL.h)
VtNrc_2202^2*(1-201/2202)/(201*200)*sum((ymaxs
                -sNeyBegrProg$OZ91.B.ume*sum(ymaxs)/
                sum(sNeyBegrProg$OZ91.B.ume) )^2)
CVtNrc_sqrt(VtNrc)/tNrc
#tttrc_(CVtNrc/0.05)^2*201/2001
tttrc_(CVtRs/5)^2*201/2001
nMin_2202*tttrc/(1+tttrc)
```

Literatur

[1] BFS, Eidg. Alkoholverwaltung (1993) Der Obstbau in der Schweiz: Feldobstbau/Obstkulturen: Schweizerische Obstbaumzählung 1991.

[2] Cochran, W. G. (1977) Sampling Techniques, 3rd ed. Wiley & Sons.

[3] Särndal, .C.-E., Swensson, B. and Wretman, J. (1997) Model assisted Survey Sampling, Springer.

[4] BFS (2000) Eidg. Obstbaumzählung 2001. Konzept Version 3.0.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

Publikationsprogramm BFS

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

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

Programme des publications de l'OFS

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

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

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

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das alle 2 Jahre nachgeführte **Publikationsverzeichnis.** Es ist auf dem Internet abzurufen oder kann gratis über Tel. 032 713 60 60 bezogen werden.

La **Liste des publications**, qui est mise à jour tous les deux ans, donne davantage de détails sur les divers moyens de diffusion. Elle est proposée sur Internet et peut être obtenue gratuitement en téléphonant au n° 032 713 60 60.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodological reports of 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 Hochschule und 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.

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.

Bis anhin wurden die Obstbäume anhand einer Vollerhebung erhoben. Die Resultate wurden jeweils nach Kantonen und Fruchtart aufgeteilt publiziert. Für die Obstbaumzählung 2001 wurde entschieden, nur noch in den landwirtschaftlichen Betrieben eine Vollerhebung durchzuführen. Für die nicht landwirtschaftlichen Betriebe war eine Stichprobenerhebung vorgesehen. Die Unterscheidung zwischen landwirtschaftlichen und nicht landwirtschaftlichen Betrieben wurde in den vorhergehenden Erhebungen nicht gemacht. Darum wurden zuerst verschiedene Modelle für die Prognose der Anzahl Bäume in nicht landwirtschaftlichen Betrieben entwickelt. Die Prognose wurde verwendet, um die Güte eines Quotientenschätzers der Anzahl Bäume zu beurteilen. Der Quotientenschätzer stützt sich auf die Korrelation der Anzahl Bäume auf nicht landwirtschaftlichen und landwirtschaftlichen Betrieben. Die Stichprobe der Gemeinden wurde nach Kantonen oder Gruppierungen von Kantonen (z.B. Grossregionen NUTS2) geschichtet. Das Ziel war, in sämtlichen Schichten einen Variationskoeffizient von weniger als 5% zu erreichen. Die Analyse der Schätzungen zeigt, dass eine Stichprobengrösse von 200 Gemeinden nicht ausreicht, um die Anzahl Bäume in nicht landwirtschaftlichen Betrieben nach Kantonen aufgeteilt mit genügender Genauigkeit zu schätzen. Sofern die Schichtung beibehalten wird, muss die Stichprobengrösse auf rund 750 Gemeinden erhöht werden. Ein gangbarer Mittelweg wäre vermutlich eine moderate Erhöhung der Stichprobe zusammen mit einer Reduktion der Ansprüche an die Genauigkeit in einigen Kantonen.

Bestellnummer:
338-0008

Bestellungen:
032 713 60 60 Fax: 032 713 60 61

Preis:
gratis

ISBN 3-303-00241-X