Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Revision des Schweizerischen Lohnindex

Schätzmethoden der Lohnindices und deren Varianzschätzer

Autor **Daniel Kilchmann**
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
Swiss Federal Statistical Office

Neuchâtel, 2004

Vorwort

Innerhalb der Revision des Schweizerischen Lohnindex (SLI) wurde auch die Schätzmethode auf einen allfälligen Handlungsbedarf untersucht. Dieser Methodenbericht beschreibt die Methode der Schätzung und der Genauigkeit der Schätzer und zeigt auf, mit welcher Schätzgenauigkeit gerechnet werden kann. Zudem wurden die Schätzer und deren Varianzschätzer eines neuen Lohnindex, der die Veränderung der Lohnmasse verfolgen soll, formalisiert und geschätzt. Die methodischen Arbeiten wurden von Daniel Kilchmann (METH) durchgeführt.

Der Autor dankt Thérèse Jeanloz und André Farine der Sektion LOHN für die fachliche Unterstützung und die gute Zusammenarbeit. Beat Hulliger (METH) sei für die methodische Unterstützung und die Durchsicht des Berichts herzlich gedankt und ebenso Philippe Eichenberger (METH) für die hilfreichen Diskussionen.

Zusammenfassung

Aufgrund der Anforderungen der BFS-Sektion Löhne und Arbeitsbedingungen an die Population, die Periodizität und die Kreuzungsvariablen werden Schätzer für den bestehenden Schweizerischen Lohnindex und für einen neu einzuführenden Lohnmassenindex geschätzt. Für den bestehenden Lohnindex wird dabei darauf geachtet, dass es sich um einen *Laspeyres*-Preisindex handelt. Die Daten der Sammelstelle für die Statistik der Unfallversicherung, SSUV, der Beschäftigtenstatistik (BESTA) und der Schweizerischen Arbeitskräfteerhebung (SAKE) werden für die revidierten Schätzungen verwendet. Vertrauensintervalle der Schätzer der Lohnindices, der Veränderung der Lohnindices und des Vergleichs zweier Lohnindices werden auf derzeit verwendeten Publikationsniveaus geschätzt.

Schlüsselwörter

Methodenbericht; Lohnindex; SLI; Lohnmassenindex, SLMI, *Laspeyres*-Preisindex; Unfallwahrscheinlichkeit; Selektions Bias; Homogene Gruppen, Klassifikationsbäume; Varianzschätzer; Bias; Veränderung der Lohnindices; *Taylor*-Linearisierung; *Jackknife*.

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: 338-0026

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Grafik/Layout: BFS

Copyright: BFS, Neuchâtel 2004
Abdruck – ausser für kommerzielle Nutzung–
unter Angabe der Quelle gestattet.

ISBN: 3-303-00302-5

Inhaltsverzeichnis

1 Einführung **5**

2 Population **6**

3 Theoretische Lohnindices **6**
3.1 Definitionen 6
3.2 Zielgrösse Durchschnittslohn und Durchschnittslohnindex 8
3.3 Zielgrösse Lohnmasse und Lohnmassenindex 10

4 Schätz-Methoden **11**
4.1 Derzeitige Methode 11
4.2 Stichproben-Methode 14
4.2.1 Grundgesamtheit 14
4.2.2 Schätzung der Einschlusswahrscheinlichkeit 14
4.2.3 Schätzung der Lohnmasse anhand einer ”Stichprobe” . . . 21
4.2.4 Schätzung des Durchschnittslohns anhand einer ”Stichprobe” 24
4.2.5 Schätzung des Lohnmassenindex 25
4.2.6 Schätzung des Durchschnittslohnindex 31
4.3 Einfluss der Einschlusswahrscheinlichkeit auf die Schätzer 35

5 Modell-Methode **38**
5.1 Regressions-Methode 38
5.1.1 Definitionen 38
5.1.2 Selektions Bias 39
5.2 Kalibrierungs-Methode 40

6 Diskussion der Schätz- und Modell-Methoden **42**

7 Homogene Gruppen **42**
7.1 Warum Homogenität? 43
7.2 Methode LOHN 43
7.2.1 Definitionen 44
7.2.2 Verifizierung der Homogenitäts-Hypothese 45
7.2.3 Diskussion der Methode LOHN 45
7.3 Klassifikationsbäume 46
7.4 HG bei der Schätz-Methode 47
7.5 HG bei der Modell-Methode 47
7.5.1 Daten und Variablen 48
7.5.2 Resultate 48

8 Gewählte Methode **50**
8.1 Entscheid LOHN 50
8.2 Vorteile und ungelöste Probleme dieser Methode 51
8.2.1 Strukturwandel in der Population 51
8.2.2 Homogene Gruppen und deren Gliederung 51
8.3 Revidierte Methode LOHN 52
8.4 BESTA und SAKE 52
8.5 Unterschiedliche Definitionen der SLI, BESTA und der SAKE . . 53
8.6 Notationen und Schätzer der Populationsgrösse 56
8.7 Varianzschätzer der Populationsgrösse 60

9 Beispiel anhand der BESTA99 und der SAKE99 **63**
9.1 SAKE99 und BESTA99 63
9.2 Schätzung der Population 63
9.3 Bias der Populations-Schätzer 67
9.4 Genauigkeit der Lohnindices-Schätzer 68
9.4.1 Lohnmassenindex 69
9.4.2 Durchschnittslohnindex 79

10 Zusammenfassung **93**

11 Schlussfolgerung **98**

Anhang **102**

A Variablen und Wirtschaftszweige (u.a. NOGA28) **102**

B Klassifikationsbäume **105**

C Einsetzung der NOGA in der SAKE99 **107**

D Schätzung der Populationsgrösse **108**

E Jackknife-Kovarianzschätzer Methode **114**

F Simulation des Lohnmassenindex **115**

G Vergleich zweier Durchschnittslohnindices **117**

H SAS Programme **122**

Literatur **123**

1 Einführung

Die BFS-Sektion Statistische Methoden (METH) wurde von der Sektion Löhne und Arbeitsbedingungen (LOHN) angefragt die methodische Unterstützung bei der Revision des jährlichen Lohnindex zu liefern, der bis anhin berechnet wurde. Zusätzlich wurde ein Modell für einen vierteljährlichen Lohnindex der Lohnmasse gefragt. Die beiden Indices unterscheiden sich zusätzlich noch bei der Definition des Lohns der Lohnempfänger, auf welchem der Index beruht. Die Lohnindices sollen auf homogenen Lohngruppen beruhen, so dass der Total Lohnindex die gewichtete Summe der Lohnindices der Gruppen ist. Zudem wurde von LOHN die Auflage gemacht, dass Indexberechnungen ”à la carte” möglich sein sollen. Dies bedeutet, dass die Gewichtung an jede einzelne Person in den Daten gekoppelt sein muss, oder dass die Gewichtung für jede ”à la carte” Indexberechnung neu berechnet werden muss. Für LOHN wäre die erste Variante wünschenswert. Nach wie vor sollen die Lohndaten von den Unfalldeklarationen der ”Sammelstelle für die Statistik der Unfallversicherung” (SSUV) stammen. Der Einfachheit halber werden die beiden Indices im Folgenden kurz ’SLI’, schweizerische Lohnindices, genannt. Eine differenziertere Definition folgt im Abschnitt 3.
Die Indexberechnung hat zum Ziel die Veränderung einer Grösse zu evaluieren, wobei diese Grösse zum Zeitpunkt t_0 als Referenzgrösse fixiert wird. In unserem Fall werden die Veränderungen des Durchschnittslohns und der Lohnmasse für homogene Gruppen berechnet. Die homogenen Lohngruppen werden derzeit anhand von sozio-ökonomischen Charakteristiken der Personen definiert. Die Bildung der revidierten homogenen Lohngruppen wird in diesem Bericht nicht näher untersucht. Dies soll zu einem späteren Zeitpunkt untersucht werden.
Wir werden, nachdem wir die Population der SLI definiert haben, die theoretische Basis für die SLI beschreiben, indem wir zwei Zielgrössen für die Indexberechnung definieren. Danach werden wir den Berechnungsvorschlag von LOHN formalisieren und die Auswirkungen dieser Indexberechnungen auf die Methode aufzeichnen. Darauf folgt ein Abschnitt in welchem wir die statistischen Grundlagen zur Indexberechnung liefern und verschiedene Schätzmethoden beschreiben. In einem weiteren Abschnitt werden wir die homogenen Lohngruppen auf ihre Homogenität untersuchen und deren Bildung anhand von Klassifikationsbäumen analysieren. Darauf folgt ein Abschnitt, der sich mit der von LOHN gewählten Methode auseinandersetzt und die Schätzer der Lohnindices und deren Varianzschätzer beschreibt. Vor einer Zusammenfassung des Berichts und dem Abschnitt mit den Schlussfolgerungen wenden wir in einem Abschnitt die gewählte Methode anhand der Daten der Beschäftigungsstatistik (BESTA) und der Schweizerischen Arbeitskräfteerhebung (SAKE) an. Bei der gewählten Methode handelt es sich um eine Anpassung der derzeit verwendeten Methode. Wir untersuchen ebenfalls die Qualität der Schätzer und stellen sie den derzeit verwendeten Qualitätsansprüchen

in den Publikationen gegenüber.

2 Population

Die Grundgesamtheit für die Berechnung der Lohnindices beschränkt sich auf die Lohnempfänger, welche in einer Produktionsstätte in der Schweiz arbeiten. Unabhängige, Mitarbeiter im Familienbetrieb, oberes Kader, Lehrlinge und Personen mit unbekannter Berufssituation werden ausgeschlossen.

3 Theoretische Lohnindices

Wie bereits in der Einführung erwähnt, werden wir zwei Lohndices schätzen. Aus mathematisch-statistischer Sicht, unterscheiden sich die beiden dadurch, das mit dem einen die Veränderungen des Durchschnittslohns beobachtet werden sollen, und mit dem anderen sollen die Veränderungen der Lohnmasse beobachtet werden. Wir werden im folgenden die Berechnung dieser Lohnindices formal beschreiben. Dabei wird auch die bisherige Methode miteinbezogen, sowie Vorschläge von LOHN bezüglich der Grösse, welche die Veränderung der Lohnmasse beobachten soll.
Wir gehen zuerst davon aus, dass sämtliche Daten von sämtlichen Lohnempfängern erhältlich sind. Wir sind uns bewusst, dass dies nicht der Fall sein wird, aber wir stellen diese Hypothese auf, um nachher Überlegungen anstellen zu können, wie die entsprechenden Grössen geschätzt werden können, oder aber welche Abweichungen die vorgeschlagene Methode nach sich zieht.

3.1 Definitionen

Es sei
t = Zeitindex der Erhebungsperiode. Dabei handelt es sich bei der Zielgrösse Durchschnittslohn um Jahre und bei der Zielgrösse Lohnmasse (c.f. Abschnitt 3.3) um vierteljährliche Ereignisse innerhalb der Jahre. Im folgenden werden wir von dieser Unterscheidung ausgehen, sie aber nicht explizit erwähnen.
$U_{j,t}$ = Lohngruppe j.
N_u = Anzahl Lohngruppen. Wir gehen davon aus, dass die Anzahl Lohngruppen sich nicht ändert.
$E_{d,t} = \bigcup_{j \in J_d} U_{j,t}$ ist der Untersuchungsbereich (Domain), welcher für die Publikation verwendet wird. Der Untersuchungsbereich $E_{d,t}$ ist durch die J_d Lohngruppen definiert. Wir gehen davon aus, dass die Definition der Untersuchungsberei-

che sich nicht ändern wird. Somit werden wir im folgenden nur noch E_d für $E_{d,t}$ schreiben.
$N_{j,t}$ = Anzahl Lohnempfänger in der Gruppe $U_{j,t}$.
$N_t = \sum_{j \in N_u} N_{j,t}$ gesamte Anzahl Lohnempfänger.
$N_{d,t} = \sum_{j \in J_d} N_{j,t}$ Anzahl Lohnempfänger im Untersuchungsbereich E_d.
$P_{i,j,t}$ = standardisierter Lohn des Lohnempfängers i, welcher der Gruppe $U_{j,t}$ angehört. Die Standardisierung des Lohns wurde von LOHN vorgenommen.
$M_{j,t} = \sum_{i \in U_{j,t}} P_{i,j,t}$ = Lohnmasse der Gruppe $U_{j,t}$ in der Periode t.
$M_t = \sum_j M_{j,t} = \sum_j \sum_{i \in U_{j,t}} P_{i,j,t}$ = Lohnmasse der Periode t.
$M_{d,t} = \sum_{j \in J_d} M_{j,t}$ = Lohnmasse im Untersuchungsbereich E_d der Periode t.
$\mu_{j,t} = \frac{1}{N_{j,t}} M_{j,t} = \frac{1}{N_{j,t}} \sum_{i \in U_{j,t}} P_{i,j,t}$ = Durchschnittslohn in der Gruppe $U_{j,t}$.
$w_{j,t}$ = Gewicht der Gruppe $U_{j,t}$.
I^D_{j,t,t_0} = theoretischer Durchschnittslohnindex der Gruppe $U_{j,t}$, für die Periode t bezüglich der Referenzperiode t_0.
I^D_{t,t_0} = theoretischer Durchschnittslohnindex über alle Gruppen.
I^D_{d,t,t_0} = theoretischer Durchschnittslohnindex im Untersuchungsbereich E_d, für die Periode t bezüglich der Referenzperiode t_0.
I^M_{j,t,t_0} = theoretischer Lohnmassenindex der Gruppe $U_{j,t}$, für die Periode t bezüglich der Referenzperiode t_0.
I^M_{d,t,t_0} = theoretischer Lohnmassenindex im Untersuchungsbereich E_d, für die Periode t bezüglich der Referenzperiode t_0.
I^M_{t,t_0} = theoretischer Lohnmassenindex über alle Gruppen.

Bevor wir die Lohnindices beschreiben wollen wir zuerst einen Typus eines Index vorstellen, den wir in diesem Bericht immer wieder antreffen werden, der ***Laspeyres*-Preisindex**. Wir haben uns bei der Definition des *Laspeyres*-Index auf jene in [11] abgestützt. Bei diesem Index wird zu einer Referenzperiode t_0 ein Warenkorb festgelegt und berechnet, wieviel man dafür bezahlen muss. Dann wir zum Zeitpunkt $t > t_0$ berechnet, wieviel für den gleichen Warenkorb bezahlt werden muss. Der Index besteht aus dem Quotienten zwischen dem 'Einkauf' zum Zeitpunkt t und dem 'Einkauf' zum Zeitpunkt t_0. Hiermit wird die Preisentwicklung festgestellt, die unabhängig von der Menge der Güter ist, da diese durch den Warenkorb festgelegt wurden. So können wir zum Beispiel den Warenkorb in der Referenzperiode mit N_t festlegen, der aus den einzelnen 'Produkten' N_{j,t_0}, den Lohnempfängern in der Lohngruppe j besteht. Als 'Preis' den wir für die 'Produkte' bezahlen müssen, können wir den Durchschnittslohn der entsprechenden Lohngruppe definieren. Somit haben wir einen Index für die Entwicklung des Durchschnitts-

lohns, wenn zum Zeitpunkt t gleich viele Lohnempfänger in jeder Lohnklasse sind wie zur Referenzperiode. Man kann dies auch als Index der Entwicklung der Lohnmasse betrachten, wenn davon ausgegangen wird, dass in jeder Lohnklasse gleich viele Lohnempfänger sind wie zur Referenzperiode. Anders gesagt, wieviel Lohnmasse müsste aufgewendet werden, wenn zum Zeitpunkt t gleich viele Lohnempfänger wären, wie zum Zeitpunkt t_0 waren und sich somit nur der 'Preis' der Lohnempfänger verändert hat. Die Entwicklung der realen Lohnmasse kann aber mit diesem Index nicht beobachtet werden, da die 'Struktur' (der Warenkorb) konstant ist.

3.2 Zielgrösse Durchschnittslohn und Durchschnittslohnindex

Anhand dieser Zielgrösse wird die Entwicklung des Durchschnittslohns untersucht. Der Durchschnittslohn wird dabei in sogenannten Homogenen (Lohn-) Gruppen berechnet. Die Homogenität der Löhne innerhalb der Gruppen hat dabei einen grossen Einfluss auf die zeitliche Entwicklung des Index.
Die Periodizität für die Berechnung eines Durchschnittslohnindex ist jährlich.
Die Berechnung des theoretischen Durchschnittslohnindex kann folgendermassen beschrieben werden: Der theoretische Durchschnittslohnindex pro Lohngruppe ist

$$I^D_{j,t,t_0} = \frac{\mu_{j,t}}{\mu_{j,t_0}} \tag{1}$$

$$I^D_{t,t_0} = \sum_{j \in N_u} w_{j,t_0} I^D_{j,t,t_0} \tag{2}$$

Die Berechnung des theoretischen Durchschnittslohnindex des Untersuchungsbereichs E_d ist:

$$I^D_{d,t,t_0} = \sum_{j \in J_d} w_{j,t_0} I^D_{j,t,t_0} \tag{3}$$

Die derzeitige Berechnungsmethode des Durchschnittslohnindex ist ein *Laspeyres*-Preisindex, bei welchem das Gewicht der Gruppe $U_{j,t}$ proportional zu ihrem Lohnmassenanteil der Periode t_0 ist:

$$w_{j,t_0} = \frac{M_{j,t_0}}{\sum\limits_j M_{j,t_0}} \tag{4}$$

Die Gewichtung (4) ist somit für jede Gruppe und unter der jeweiligen Lohndefiniton konstant in der Zeit, d.h. $w_{j,t_r} = w_{j,t_s}$, $t_r \geq t_0,\ t_s \geq t_0$.
Man könnte sich auch eine Gewichtung vorstellen, welche z.B. auf der Anzahl Arbeitnehmern basiert: $w_{j,t_0} = \frac{N_{j,t_0}}{\sum\limits_j N_{j,t_0}}$. Der entsprechende Durchschnittslohnindex

wäre jedoch kein *Laspeyres*-Preisindex und somit entspricht er nicht den Vorgaben von LOHN.

Mit der Gewichtung (4) ist der theoretische Durchschnittslohnindex über alle Gruppen

$$\begin{aligned} I^D_{t,t_0} &= \sum_j w_{j,t_0} I^D_{j,t,t_0} = \sum_j \frac{M_{j,t_0}}{\sum\limits_j M_{j,t_0}} \frac{\mu_{j,t}}{\mu_{j,t_0}} \\ &= \frac{1}{\sum\limits_j M_{j,t_0}} \sum_j N_{j,t_0} \mu_{j,t_0} \frac{\mu_{j,t}}{\mu_{j,t_0}} \\ &= \frac{\sum\limits_{j \in N_u} N_{j,t_0} \mu_{j,t}}{\sum\limits_{j \in N_u} N_{j,t_0} \mu_{j,t_0}} \end{aligned} \tag{5}$$

Der Durchschnittslohnindex (5) ist also ein *Laspeyres*-Preisindex mit der Anzahl Lohnempfänger pro Gruppe während der Referenzperiode als konstanter Warenkorb.
Dies bedeutet, dass die Lohnmasse der Referenzperiode jeder Lohngruppe bekannt sein muss, ebenso wie die Anzahl Lohnempfänger dieser Lohngruppe.
Der Durchschnittslohnindex wird in den Untersuchungsbereichen E_d publiziert, welche zusammengefassten homogenen Gruppen gleichkommen. Der anhand von (4) gewichtete theoretische Durchschnittslohnindex pro Untersuchungsbereich ist

$$I^D_{d,t,t_0} = \sum_{j \in J_d} w_{j,t_0} I^D_{j,t,t_0} = \frac{\sum\limits_{j \in J_d} N_{j,t_0} \mu_{j,t}}{\sum\limits_{j \in N_u} N_{j,t_0} \mu_{j,t_0}} \tag{6}$$

Der Durchschnittslohnindex (6) ist kein *Laspeyres*-Preisindex, da der Warenkorb der Referenzperiode grösser ist, als der Warenkorb der Periode t ist dieser Index auch in der Referenzperiode < 1. Mit anderen Worten, die Vorgabe, einen Durchschnittslohnindex zu berechnen, der ein *Laspeyres*-Preisindex ist, kann mit der vorgeschlagenen Gewichtung nur für den Durchschnittslohnindex über alle Lohngruppen erreicht werden. Nur eine Anpassung der Gewichtung kann garantieren, dass die Lohnindices in den Untersuchungsbereichen ebenfalls *Laspeyres*-Preisindices sind. Um dieses Ziel zu erreichen, müssen wir die Gewichtung folgendermassen definieren:

$$w_{j,d,t_0} = \frac{M_{j,t_0}}{\sum\limits_{j \in J_d} M_{j,t_0}} \qquad \forall j \in J_d, \forall d \tag{7}$$

Mit dieser Gewichtung wird ein Durchschnittslohnindex berechnet, der die Lohnveränderung im Untersuchungsbereich E_d beschreibt [1].

$$I^D_{d,t,t_0} = \sum_{j \in J_d} w_{j,d,t_0} I^D_{j,t,t_0} = \frac{\sum\limits_{j \in J_d} N_{j,t_0} \mu_{j,t}}{\sum\limits_{j \in J_d} N_{j,t_0} \mu_{j,t_0}} \tag{8}$$

Das bedeutet aber, dass für den gleichen Lohnempfänger mehrere Gewichte existieren. Zum Beispiel, wenn ein Durchschnittslohnindex für den 2. Sektor berechnet werden soll, dann hat der Lohnempfänger im Wirtschaftszweig NOGA=20 nicht das gleiche Gewicht, wie wenn ein Durchschnittslohnindex für die NOGA=20 berechnet werden soll. Somit kann die Vorgabe, einen *Laspeyres*-Preisindex zu berechnen nicht gleichzeitig mit der Vorgabe, ein einziges Gewicht pro Lohnempfänger zu berechnen, erfüllt werden.

3.3 Zielgrösse Lohnmasse und Lohnmassenindex

Bei der Lohnindexberechnung mit der Lohnmasse als Zielgrösse wird die Entwicklung der Lohnmasse in den homogenen Gruppen untersuchen. Durch eine Aggregation der Resultate der homogenen Gruppen auf ein höheres Niveau soll der Index für die Variablen NOGA und Geschlecht und falls möglich Beschäftigungsgrad veröffentlicht werden. Die Periodizität dieses neuen Lohnindex sollte vierteljährlich sein. Es handelt sich dabei um einen neuen Index, der bisher nicht berechnet wurde.
Wir stellen im folgend den Vorschlag von LOHN vor.
Der theoretische Lohnmassenindex kann folgendermassen definiert werden:

$$I^M_{j,t,t_0} = \frac{M_{j,t}}{M_{j,t_0}} = \frac{N_{j,t}\mu_{j,t}}{N_{j,t_0}\mu_{j,t_0}} \tag{9}$$

$$\begin{aligned} I^M_{t,t_0} &= \sum_j w_{j,t_0} \frac{M_{j,t}}{M_{j,t_0}} = \sum_j \frac{M_{j,t_0}}{\sum\limits_j M_{j,t_0}} \frac{M_{j,t}}{M_{j,t_0}} \\ &= \frac{\sum\limits_j M_{j,t}}{\sum\limits_j M_{j,t_0}} = \frac{\sum\limits_j N_{j,t}\mu_{j,t}}{\sum\limits_j N_{j,t_0}\mu_{j,t_0}} = \frac{M_t}{M_{t_0}} \end{aligned} \tag{10}$$

Der theoretische Lohnmassenindex für den Untersuchungsbereich E_d ist folgendermassen definiert:

$$I^M_{d,t,t_0} = \frac{M_{d,t}}{M_{d,t_0}} = \frac{N_{d,t}\mu_{d,t}}{N_{d,t_0}\mu_{d,t_0}} \tag{11}$$

[1] $w_{j,t_0} = w_{j,d,t_0}$ wenn $J_d = \bigcup\limits_{j \in N_u} U_j$.

Diese Form des Lohnmassenindex für E_d kann mit der derzeit verwendeten Gewichtung (4) nicht erreicht werden. Wenn hingegen die Gewichtung (7) verwendet wird, dann finden wir den erwünschten Index:

$$I^M_{d,t,t_0} = \sum_{j \in J_d} w_{j,d,t_0} I^M_{j,t,t_0} = \frac{M_{d,t}}{M_{d,t_0}} \tag{12}$$

Die Lohnmassenindices (9), (10) und (11) sind Wertindices und folglich keine *Laspeyres*-Indices. Wie aus (9) bis (11) ersichtlich ist, hat die Homogenität der Lohnmasse in den Homogenen Lohngruppen keinen Einfluss auf die Lohnmassenindices. Im weiteren fällt auf, dass die Gewichtung w_{j,t_0} dem Konzept eines Lohnmassenindex entspricht.

4 Schätz-Methoden

In den folgenden Abschnitten werden wir die statistischen Grundlagen liefern, welche zur Schätzung der Lohnindices dienen. Zuerst werden wir die bis anhin verwendete Methode beschreiben, bevor wir alternative Schätz-Methoden untersuchen werden.
Es stehen uns zwei Methoden für die Schätzung der Lohnmasse und des Durchschnittslohns zur Verfügung. Einerseits eine Modell basierte Methode und andererseits eine Methode, welche von einer Zufallsstichprobe ausgeht.

4.1 Derzeitige Methode

Wie bereits erwähnt stammen die Lohndaten nicht aus einer Erhebung im eigentlichen Sinne (weder Stichprobe noch Vollerhebung), sondern von den Unfalldeklarationen der SSUV. Wenn wir im folgenden von Verunfallten sprechen, so meinen wir damit jene Verunfallten, deren Daten wir verwenden. I^D_{t,t_0} und I^D_{d,t,t_0} müssen aufgrund der Daten der SSUV geschätzt werden. Wir definieren:
$V_{j,t} =$ Verunfalltengruppe j. Wenn die Charakteristiken der Lohngruppe $U_{j,t}$ mit jenen der Verunfalltengruppe $V_{j,t}$ übereinstimmen, so haben wir $V_{j,t} \subset U_{j,t}$. Dies ist die implizit postulierte Hypothese des derzeitigen Lohnindex, da dies die Voraussetzung für eine korrekte Schätzung mit den homogenen Gruppen als Schichtung ist.
$N_v =$ Anzahl Verunfalltengruppen. Wir gehen davon aus, dass N_v sich nicht ändert, nachdem sie festgelegt wurden.
$n_{j,t} =$ Anzahl Lohnempfänger in der Gruppe $V_{j,t}$.
$p_{i,j,t} =$ normalisierter Monatslohn des Lohnempfängers i, welcher der Gruppe $V_{j,t}$ angehört.

$m_{j,t} = \sum_{i \in V_j} p_{i,j,t} =$ Lohnmasse der Gruppe $V_{j,t}$ in der Periode t.
$m_t = \sum_j m_{j,t} = \sum_j \sum_{i \in V_{j,t}} p_{i,j,t} =$ Lohnmasse der Verunfallten der Periode t.
$\nu_{j,t} = \frac{1}{n_{j,t}} m_{j,t} = \frac{1}{n_{j,t}} \sum_{i \in V_{j,t}} p_{i,j,t} =$ Durchschnittslohn in der Gruppe $V_{j,t}$.
$\tilde{I}^D_{j,t,t_0} =$ Durchschnittslohnindex der Gruppe $U_{j,t}$, für die Periode t bezüglich der Referenzperiode t_0, der anhand der Daten von $V_{j,t}$ geschätzt wurde.
$\tilde{I}^D_{t,t_0} =$ Schätzer für den Durchschnittslohnindex über alle Gruppen, auf den Daten der SSUV basierend.
$\tilde{I}^D_{d,t,t_0} =$ Schätzer für den Durchschnittslohnindex für den Untersuchungsbereich E_d, auf den Daten der SSUV basierend.
Die Schätzung aufgrund der Daten der SSUV des Durchschnittslohnindex innerhalb der Gruppen kann folgendermassen beschrieben werden

$$\tilde{I}^D_{j,t,t_0} = \frac{\nu_{j,t}}{\nu_{j,t_0}} \tag{13}$$

und die entsprechende Schätzung des Durchschnittslohnindex über alle Gruppen ist somit

$$\begin{aligned} \tilde{I}^D_{t,t_0} &= \sum_{j \in N_v} \tilde{w}_{j,t_0} \tilde{I}^D_{j,t,t_0} = \sum_j \frac{\tilde{M}_{j,t_0}}{\sum_j \tilde{M}_{j,t_0}} \frac{\nu_{j,t}}{\nu_{j,t_0}} \\ &= \frac{\sum_{j \in N_v} \tilde{N}_{j,t_0} \frac{\tilde{\mu}_{j,t_0}}{\tilde{\nu}_{j,t_0}} \tilde{\nu}_{j,t}}{\sum_{j \in N_v} \tilde{N}_{j,t_0} \tilde{\mu}_{j,t_0}} \end{aligned} \tag{14}$$

wobei die Schätzung der Anzahl Lohnempfänger in der Gruppe j für die Referenzperiode mit $\tilde{N}_{j,t_0}$ bezeichnet wurde. Ebenso haben wir den Lohnmassenschätzer mit $\tilde{M}_{j,t_0}$ bezeichnet und die daraus folgende Gewichtung mit

$$\tilde{w}_{j,t_0} = \frac{\tilde{M}_{j,t_0}}{\sum_j \tilde{M}_{j,t_0}}.$$

Der Durchschnittslohn in der Lohnempfängergruppe U_{j,t_0} wurde durch den Durchschnittslohnschätzer in der Verunfalltengruppe V_{j,t_0} geschätzt:

$$\tilde{\mu}_{j,t_0} = \tilde{\nu}_{j,t_0}.$$

Bemerkung: Der Quotient der Durchschnittslöhne (13) wurde im Detailkonzept mit RSM_j bezeichnet.

Der Schätzer des Durchschnittslohnindex für den Untersuchungsbereich E_d nimmt analog zu (14) und aufgrund von 8 folgende Form an

$$\begin{aligned} \tilde{I}^D_{d,t,t_0} &= \sum_{j \in J_d} \tilde{w}_{j,d,t_0} \tilde{I}^D_{j,t,t_0} \\ &= \frac{\sum_{j \in J_d} \tilde{N}_{j,t_0} \frac{\tilde{\mu}_{j,t_0}}{\tilde{\nu}_{j,t_0}} \tilde{\nu}_{j,t}}{\sum_{j \in J_d} \tilde{N}_{j,t_0} \tilde{\mu}_{j,t_0}} \end{aligned} \tag{15}$$

Diskussion des Durchschnittslohnindex

Aus (14) ist ersichtlich, dass die Schätzung des Durchschnittslohnindex nur dann ein *Laspeyres*-Index ist, wenn

$$\frac{\mu_{j,t_0}}{\nu_{j,t_0}} = 1 \quad \forall j, \tag{16}$$

das heisst, wenn die Schätzung des Durchschnittslohns aus der SSUV der Referenzperiode für jede Gruppe dem wahren Wert μ_{j,t_0} entspricht. Bis anhin ist μ_{j,t_0} nur als Schätzer vorhanden (z.B. Lohnstrukturerhebung (LSE) oder SAKE). Dieser hat wie jeder Schätzer eine Varianz und somit kann das Kriterium (16) nur mit einem statistischen Fehler erfüllt werden. Es ist weiter fraglich, ob der Durchschnittslohn der Verunfalltengruppen $V_{j,t}$ dem Durchschnittslohn der Lohngruppen $U_{j,t}$ entspricht, oder dass letzterer zumindest angenähert wird. In der derzeitigen Berechnungsmethode wurde diese Hypothese angenommen. Eine Überprüfung dieser Hypothese kann nur aufgrund von Daten einer anderen Erhebung gemacht werden, z.B. LSE oder SAKE. Eine Überprüfung über mehrere Jahre hinweg ist unabdingbar.
Die Daten der LSE stammen von den Betrieben und beziehen sich somit auf die Beschäftigten und nicht auf die Arbeitnehmer (resp. Lohnempfänger). Somit muss mit einer Abweichung gerechnet werden. Hinzu kommt, dass der Lohn in der LSE nicht gleich definiert ist wie für die Lohnindex Berechnungen. Die Lohngruppen können anhand der LSE mit den Variablen NOGA, Geschlecht und Beschäftigungsgrad gebildet werden. Die Variable Anforderungsniveau der LSE ist der Variablen Qualifikation der SSUV nicht gleich zu setzen. Da die LSE nur alle zwei Jahre erhoben wird, ist diese für unser Vorhaben nicht geeignet.
Es stellt sich die Frage, ob anstatt des Durchschnittslohns nicht der Median des Lohns benutzt werden sollte. Dieser wäre robust gegenüber aussergewöhnlichen Löhnen innerhalb der homogenen Gruppen, welche die Homogenität in Frage stellen könnten. Es scheint, dass der Median in den meisten anderen Lohnstatistiken des BFS benutzt wird. **Diskussion des Lohnmassenindex**

Um die Lohnmassenindices (9) - (11) schätzen zu können, müssen wir die gesamte

Lohnmasse in jeder Gruppe für jede Periode schätzen. Es gibt zur Zeit keine Erhebung, welche die Lohnmasse vierteljährlich für die vorgesehenen Lohngruppen liefern könnte. Deshalb müssen wir ein Modell erstellen, welches die Lohnmasse aus beobachteten Grössen vorhersagt. Zum Beispiel

$$M_{j,t} = f(N_{j,t}, m_{j,t}, n_{j,t}, t) + \epsilon_{j,t} \tag{17}$$

wobei $\epsilon_{j,t}$ der statistische Fehler des Modells ist. Hierbei haben wir die Definitionen des Abschnitts 3.1 verwendet.
Das einfachste Modell wäre ein Totalschätzer anhand des Durchschnittslohns.

$$\hat{M}_{j,t} = \hat{N}_{j,t}\frac{m_{j,t}}{n_{j,t}} \tag{18}$$

Die Gültigkeit des Modells (18) könnte zum Beispiel durch den Lohnmassenschätzer der LSE überprüft werden. Die Erarbeitung eines solchen Modelles ist sehr zeitaufwändig und der genaue Ablauf der Berechnung müsste durchdacht werden (z.B. Verzögerung der Datenerhebung und unterschiedliche Definitionen und Grundgesamtheiten).
Dieser Ansatz kann die vierteljährlichen Veränderungen der Lohndaten nicht mit einbeziehen, da keine Erhebung vorhanden ist, welche die Lohnmasse vierteljährlich erhebt.

4.2 Stichproben-Methode

Wir stellen die Hypothese auf, dass sich die Unfälle innerhalb der homogenen Gruppen zufällig und unabhängig voneinander ereignen. Diese Methode hat zum Ziel, eine allfällige Verzerrung durch die Unfallhäufigkeit zu berücksichtigen. Bei der Stichproben-Methode bauen die Schätzer somit auf einem postulierten Stichprobenplan für das Unfallgeschehen auf.

4.2.1 Grundgesamtheit

Die Grundgesamtheit U_t besteht aus den Lohnempfängern, das heisst, $U_t = \bigcup_j U_{j,t}$.

Somit ist die Grundgesamtheit von der Zeit abhängig, da die Menge der Lohnempfänger sich ändert.

4.2.2 Schätzung der Einschlusswahrscheinlichkeit

Die Einschlusswahrscheinlichkeit oder die Wahrscheinlichkeit, dass die Daten eines Verunfallten uns geliefert werden, muss geschätzt werden. Wir gehen davon aus, dass die Einschlusswahrscheinlichkeit innerhalb der homogenen Gruppen für

alle Verunfallten dieser Gruppe gleich ist. Das heisst, dass die Unfallwahrscheinlichkeit $\pi_{i,j,t}$ in der Gruppe $V_{j,t}$ für die Periode t konstant ist: $\pi_{i,j,t} = \pi_{j,t} = \frac{n_{j,t}}{N_{j,t}} \quad \forall i \in V_{j,t}, \forall j, t.$
Wir versuchen ein Modell für die Einschlusswahrscheinlichkeit zu schätzen, welches auf beobachteten Grössen basiert.

$$\pi_{j,t} = f\left(\frac{n_{j,t}}{N_{j,t}}\right) + \epsilon_{j,t} \qquad \forall t \tag{19}$$

wobei $\epsilon_{j,t}$ der statistische Fehler des Modells ist.
Wir bezeichnen den Schätzer der Einschlusswahrscheinlichkeit für die Periode t mit $\hat{\pi}_{j,t}$. Anhand eines ajustierten Modells ist $\hat{\pi}_{j,t} = E(\pi_{j,t}) \quad \forall j, t$, '$E()$' steht für den Erwartungswert.
Wir werden ein einfaches lineares Regressionsmodell anwenden, um das folgende Modell zu überprüfen:

$$\hat{\pi}_{j,t} = \frac{\hat{n}_{j,t}^{SAKE}}{\hat{N}_{j,t}^{SAKE}} \tag{20}$$

Da 1999 eine Frage bezüglich der Verunfallung in der SAKE gestellt wurde, können wir die Unfallwahrscheinlichkeit anhand dieser Daten schätzen.

$$\hat{\pi}_{j,99} = \frac{\hat{n}_{j,99}^{SAKE}}{\hat{N}_{j,99}^{SAKE}} \tag{21}$$

Die Anzahl Lohnempfänger in jeder Gruppe wird ebenfalls anhand der SAKE geschätzt. Somit ajustieren wir das folgende Modell

$$n_{j,99}^{SSUV} = f\left(\hat{n}_{j,99}^{SAKE}\right) + \epsilon_{j,99}. \tag{22}$$

Wir haben aus der SAKE99 nur die vollzeitbeschäftigten Arbeitnehmer[2], deren Alter bekannt war, berücksichtigt, da in den Daten der SSUV99 nur die Vollzeitbeschäftigten erhoben wurden. Somit verblieben 5766 Arbeitnehmer in den Daten der SAKE99, wovon 754 verunfallt waren, für diese Analyse. In der SSUV99 verblieben nach der Datenaufarbeitung 144'260 verunfallte Arbeitnehmer, wovon 91 einen fehlenden NOGA Code oder einen NOGA Code über 93 hatten. Effektiv verblieben somit 144'169 Personen in den Daten der SSUV99. Mehrfachangaben wurden in der Datenaufarbeitung nicht bearbeitet, somit können durchaus Doubletten in den verwendeten Daten der SSUV vorhanden sein. In den SSUV Daten wird noch die alte Branchenkodierung (ASWZ85) verwendet. Die NOGA95 wird durch einen Umrechnungsschlüssel definiert. Dies ist in Zukunft nicht

[2]Das heisst, Erwerbsstatus = Arbeitnehmer und Beschäftigungsgrad $\geq$ 90%.

mehr nötig, da die Erhebung der NOGA in die Revision des Lohnindex miteinbezogen worden ist (siehe [4]). Wir definieren die Lohngruppen anhand der Variablen Branche, Alter und Geschlecht. Wobei wir den Vorschlag von LOHN übernommen haben, zwei Altersklassen zu kreieren: 19-39 Jahre und 40-64 Jahren.
Bereits diese Kombination von Variablen stellt jedoch ein Problem dar, da nicht für alle 112 Kreuzungen Daten in der SAKE99 vorhanden sind. Aus der Tabelle 1 geht hervor, dass in der SAKE99 in 20 der 112 Zellen keine Daten vorhanden waren. In 87 Zellen waren weniger als 10 Observationen vorhanden. Rund 6% der verunfallten Arbeitnehmer in der SAKE sind ohne NOGA Code. Diese fehlenden Angaben wurden proportional auf die vorhandenen Angaben verteilt, siehe Anhang C. Die Anzahl Verunfallter werden in der SAKE mit 262'226 gegenüber den 144'169 aus der SSUV klar überschätzt. Oder aber wir haben in der SAKE auch Bagatellen, welche in der SSUV nicht mehr vorhanden sind. Diese können wir leider nicht eruieren.
Wir haben verschiedene Regressions-Modelle für die Modellierung von (22) getestet. Schlussendlich haben wir ein einfaches lineares Modell beibehalten (c.f. Abbildung 1):

$$n_{j,99}^{SSUV} = 0.5637 * \hat{n}_{j,99}^{SAKE} \tag{23}$$

Der Pearson-Korrelationskoeffizient von $n_{j,99}^{SSUV}$ und $\hat{n}_{j,99}^{SAKE}$ beträgt 0.8477. Der Anteil der erklärten Varianz anhand von (23) beträgt 81.24%. Diese Werte sind recht gut. Die Regressionsgerade und die Daten der SSUV99 und der SAKE99 sind in der Graphik 2 aufgezeichnet. Wir haben eine einfache Methode zur Ausreissererkennung angewandt, um zu analysieren, ob sich die obige lineare Regression massgeblich verbessern lässt. Dabei haben wir alle Gruppen, welche prozentual die grössten absoluten Unterschiede zwischen den SSUV Daten und den SAKE Daten aufwiesen ausgeschlossen. Das heisst, wir haben die Daten ab dem 90. Percentil der Verteilung von

$$\frac{\left|n_{j,99}^{SSUV} - \hat{n}_{j,99}^{SAKE}\right|}{n_{j,99}^{SSUV}}$$

eliminiert. Auf diese Weise wurden weitere 10 Gruppen aus den Daten ausgeschlossen. Wir haben somit nur jene Daten beibehalten haben, welche folgende Bedingung erfüllt haben:

$$\frac{\left|n_{j,99}^{SSUV} - \hat{n}_{j,99}^{SAKE}\right|}{n_{j,99}^{SSUV}} < 5.85. \tag{24}$$

Der Pearson-Korrelationskoeffizient stieg auf 0.8654, die erklärende Varianz stieg auf 84.18%. Dieser Gewinn ist vernachlässigbar. Zudem hätte der Ausschluss der Ausreisser aus den Daten dazu geführt, dass für diese Gruppen eine andere Methode für die Schätzung der Einschlusswahrscheinlichkeit verwendet werden muss.

Wir haben (23) an die SAKE98 angewandt um das Modell zu testen. Wir gehen davon aus, dass die Unfallwahrscheinlichkeit innerhalb der Lohngruppen zwischen 1998 und 1999 konstant ist. Zuerst schätzen wir die Anzahl Verunfallter in der SAKE98 anhand von (21) und der SAKE98: $\hat{n}_{j,98}^{SAKE} = \hat{\pi}_{j,99} * \hat{N}_{j,98}^{SAKE}$. Danach schätzen wir die Anzahl Verunfallter, die wir in den Daten der SSUV98 aufgrund von (23) vorfinden müssten: $\hat{n}_{j,98}^{SSUV} = 0.5637 * \hat{n}_{j,98}^{SAKE}$. Danach testen wir, ob die Verteilung der Geschätzten Verunfallten der SSUV98 verschieden ist von der Verteilung der erfassten Verunfallten der SSUV98. Das heisst, $H_0 : n_{j,98}^{SSUV}$ und $\hat{n}_{j,98}^{SSUV}$ haben die gleiche Verteilung. Für diesen Test verwenden wir einen Pearson χ^2-Test.

$$d = \sum_j \frac{\left(n_{j,98}^{SSUV} - \hat{n}_{j,98}^{SSUV}\right)^2}{\hat{n}_{j,98}^{SSUV}} = 66977.49 > \chi^2_{91,0.95} = 114.27 \qquad (25)$$

Der Test ist klar signifikant, somit verwerfen wir H_0, das heisst, $n_{j,98}^{SSUV}$ und $\hat{n}_{j,98}^{SSUV}$ haben verschiedene Verteilungen.
Anhand einer linearen Regression haben wir die durchschnittliche Abweichung geschätzt:

$$\hat{n}_{j,98}^{SSUV} = 1.0352 * \hat{n}_{j,98}^{SSUV} \qquad (26)$$

Der Schätzer der Anzahl der Verunfallten aus der SAKE98 kombiniert mit (23) unterschätzt die Anzahl Verunfallter der SSUV98 im Durchschnitt um 3.52%. Die Abweichungen in den einzelnen Lohngruppen sind jedoch zu gross, um (23) für andere Jahre anzuwenden.
Eine Schätzung der Zeitkomponente der Einschlusswahrscheinlichkeit ist nicht möglich, da hierfür eine Zeitreihenanalyse über mehrere Jahre durchgeführt werden müsste. Dazu fehlen uns jedoch genügend Daten. Wir werden diese Zeitkomponente in den folgenden Abschnitten jedoch weiterhin verwenden, da die Unfallwahrscheinlichkeit sich mit der technologischen Entwicklung und der Sozialstruktur verändert.

The SAS System

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
i_nsake	1	0.5637	0.0284	19.85	<.0001	1.0000	1.0000

Abbildung 1: Lineare Regression mit dem Schätzer der Anzahl Verunfallter aus der SAKE99, nach der Imputation der fehlenden NOGA Codes, 'i_nsake'. Die Anzahl Verunfallter in den Daten der SSUV99 diente als unabhängige Variable.

The SAS System

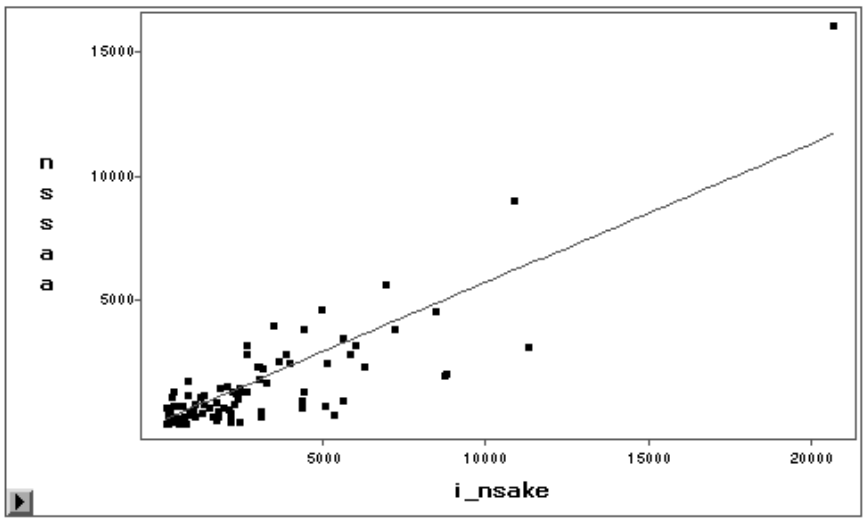


Abbildung 2: Graphik der Anzahl Verunfallter der SSUV99, 'nssaa', gegen die geschätzte Anzahl Verunfallter aus der SAKE99, nach der Imputation der fehlenden NOGA Codes, 'i_nsake'. Die Gerade entspricht der Regressionsgeraden aus (23).

Tabelle 1: Anzahl verunfallte Personen in der SAKE in den 112 Schichten, welche durch die NOGA, das Geschlecht und die Altersklassen gebildet wurden. '% der Anzahl Schichten' bezieht sich ausschliesslich auf die Schichten in welchen mindestens eine Person in der SAKE99 vorhanden ist.

Anzahl Personen in der Schicht	**Anzahl Schichten Schichten**	**% der Anzahl**
0	20	
1	11	11.96
2	13	14.13
3	10	10.87
4	4	4.35
5	3	3.26
6	8	8.70
7	6	6.52
8	4	4.35
9	8	8.70
10	3	3.26
11	4	4.35
12	1	1.09
13	1	1.09
14	2	2.17
15	2	2.17
16	2	2.17
18	1	1.09
22	1	1.09
23	2	2.17
24	2	2.17
26	1	1.09
27	1	1.09
32	1	1.09
54	1	1.09
Total 754	112	100.00

Tabelle 2: Vergleich der geschätzten Anzahl Verunfallter aus der SAKE mit der Anzahl Verunfallter in den SSUV Daten. 'NOGA_d2' bezeichnet die 28 zusammengefassten NOGA Codes. 'n_SSUV' sind die Verunfallten in der SSUV. Die fehlenden NOGA Werte in der SAKE wurden proportional verteilt, die geschätzte Anzahl Verunfallter wurde mit '$\widehat{ni}$_SAKE' bezeichnet. Ohne Imputation wurde die geschätzte Anzahl Verunfallter mit '$\hat{n}$_SAKE' bezeichnet. Die antwortenden verunfallten Arbeitnehmer sind in der Spalte 'ra_SAKE' und die antwortenden Arbeitnehmer sind in der Spalte 'r_SAKE' aufgeführt. Der Schätzer der Anzahl Arbeitnehmer aus der SAKE findet sich unter '$\hat{N}$_SAKE'. Die geschätzte Unfallwahrscheinlichkeit aufgrund der Anzahl Verunfallter in der SSUV ist mit $\hat{\pi}_{99}^{\textbf{SSUV}}$ bezeichnet, und $\hat{\pi}_{99}^{\textbf{SAKE}}$ steht für die geschätzte Unfallwahrscheinlichkeit aufgrund der geschätzten Anzahl Verunfallter aus der SAKE. $\Delta = \hat{\pi}_{99}^{\textbf{SSUV}} - \hat{\pi}_{99}^{\textbf{SAKE}}$.

NOGA_d2	n_SSUV	$\widehat{ni}$_SAKE	$\hat{n}$_SAKE	ra_SAKE	r_SAKE	$\hat{N}$_SAKE	$\hat{\pi}_{99}^{\textbf{SSUV}}$	$\hat{\pi}_{99}^{\textbf{SAKE}}$	Δ
1	1614	388	368	1	8	3516	0.46	0.11	0.35
2	641	275	261	1	5	1773	0.36	0.16	0.21
10	1567	4510	4279	13	62	22563	0.07	0.20	-0.13
15	5942	5055	4796	10	109	39165	0.15	0.13	0.02
17	1377	787	746	2	31	10740	0.13	0.07	0.05
20	4323	3129	2968	10	48	17408	0.25	0.18	0.07
21	3370	6243	5923	20	122	42672	0.08	0.15	-0.07
23	6015	8677	8231	22	199	70153	0.09	0.12	-0.04
26	1120	1959	1858	5	23	7362	0.15	0.27	-0.11
27	8482	9051	8586	22	154	65312	0.13	0.14	-0.01
29	6475	17835	16919	40	250	92494	0.07	0.19	-0.12
30	6184	10738	10187	33	264	92813	0.07	0.12	-0.05
36	316	4226	4009	7	42	17612	0.02	0.24	-0.22
45	25725	32455	30788	84	419	153432	0.17	0.21	-0.04
50	3317	4991	4735	15	93	32827	0.10	0.15	-0.05
51	8910	14113	13389	44	400	132160	0.07	0.11	-0.04
52	5253	14183	13455	41	382	125059	0.04	0.11	-0.07
55	8246	8579	8139	16	120	45803	0.18	0.19	-0.01
60	9192	15221	14439	52	334	107288	0.09	0.14	-0.06
64	3913	7665	7271	23	208	63920	0.06	0.12	-0.06
65	2721	11864	11254	40	297	87842	0.03	0.14	-0.10
66	1186	6226	5907	22	147	40953	0.03	0.15	-0.12
67	89	1675	1589	5	39	13722	0.01	0.12	-0.12
70	10058	15736	14928	56	519	172545	0.06	0.09	-0.03
75	9824	14249	13517	49	454	137997	0.07	0.10	-0.03
80	1084	11468	10879	35	322	101784	0.01	0.11	-0.10
85	5287	20591	19533	58	499	163692	0.03	0.13	-0.09
90	1938	10337	9806	28	216	75183	0.03	0.14	-0.11
Total1	144169	262226	248761	754	5766	1937791	0.07	0.14	-0.06
fehlt	91	0	13465	0	0	106696			
Total	144260	262226	262226	754	5766	2044487	0.07	0.13	-0.06

4.2.3 Schätzung der Lohnmasse anhand einer "Stichprobe"

Die Annahme, dass die Einschlusswahrscheinlichkeit ($\pi_{i,j,t}$) für alle Lohnempfänger innerhalb derselben homogenen Gruppe gleich ist, entspricht einer Bernoulli Stichprobe innerhalb der homogenen Gruppen. Die Gruppierung in homogene Gruppen entspricht einer Schichtung in N_v Schichten mit gleicher Einschlusswahrscheinlichkeit in den Schichten.
Formal heisst dies:

$$P\left(i \in V_{j,t} \mid i \in U_{j,t}\right) = \pi_{i,j,t} = \pi_{j,t} = \frac{n_{j,t}}{N_{j,t}} \qquad \forall i \in U_{j,t}, \forall j, t. \tag{27}$$

Unter der Hypothese der Unabhängigkeit der Unfälle innerhalb der homogenen Gruppen ist die Einschlusswahrscheinlichkeit zweiter Ordnung:

$$P\left(i \in V_{j,t} \wedge k \in V_{j,t} \mid i, k \in U_{j,t}\right) = \pi_{j,t}^2 \tag{28}$$

Die Stichprobengrösse $n_{j,t}$ ist in der Bernoulli Stichprobe ein Zufallszahl, deren Erwartungswert folgende Form hat:

$$E\left(n_{j,t}\right) = N_{j,t}\pi_{j,t} \tag{29}$$

Unter der Annahme, dass $\pi_{j,t} = \frac{n_{j,t}}{N_{j,t}}$ ist, kann die Varianz von $n_{j,t}$ folgendermassen ausgedrückt werden

$$Var(n_{j,t}) = N_{j,t}\frac{n_{j,t}}{N_{j,t}}\left(1 - \frac{n_{j,t}}{N_{j,t}}\right) = n_{j,t}\left(1 - \frac{n_{j,t}}{N_{j,t}}\right) = n_{j,t}(1 - f_{j,t}) \tag{30}$$

wobei $1 - f_{j,t}$ die Korrektur für die endliche Population (*finite population correction, fpc*) ist.
Der Schätzer für die Lohnmasse der Verunfalltengruppe $V_{j,t}$ für die Periode t ist unter der Voraussetzung der Homogenität der Einschlusswahrscheinlichkeit, d.h. $\pi_{i,j,t} = \pi_{j,t}$, $\forall i \in V_j$, und wenn $\pi_{j,t} = \frac{n_{j,t}}{N_{j,t}}$ ist:

$$\begin{aligned}
\hat{M}_{j,t,\pi} &= N_{j,t}\hat{\mu}_{j,t,\pi} = N_{j,t}\frac{\sum\limits_{i \in V_{j,t}} \frac{p_{i,j,t}}{\pi_{i,j,t}}}{\sum\limits_{i \in V_{j,t}} \frac{1}{\pi_{i,j,t}}} \\
&= N_{j,t}\frac{\sum\limits_{i \in V_{j,t}} \frac{p_{i,j,t}}{\pi_{j,t}}}{\frac{n_{j,t}}{\pi_{j,t}}} = N_{j,t}\frac{\sum\limits_{i \in V_{j,t}} p_{i,j,t}}{n_{j,t}} \\
&= \frac{1}{\pi_{j,t}}\sum_{i \in V_{j,t}} p_{i,j,t}
\end{aligned} \tag{31}$$

und der Totalschätzer der Lohnmasse aller Lohnempfänger ist

$$\hat{M}_{t,\pi} = \sum_j \hat{M}_{j,t} \tag{32}$$

Der Lohnmassenschätzer im Untersuchungsbereich E_d ist

$$\hat{M}_{d,t,\pi} = \sum_{j \in J_d} \hat{M}_{j,t} \tag{33}$$

Die Schätzer aus (31), (32) und (33) sind erwartungstreu, sofern die Einschlusswahrscheinlichkeit bekannt ist. Ein erwartungstreuer Varianzschätzer für den Lohnmassenschätzer (31) ist

$$\hat{V}ar\left(\hat{M}_{j,t,\pi}\right) = \frac{1}{\pi_{j,t}}\left(\frac{1}{\pi_{j,t}} - 1\right) \sum_{i \in V_{j,t}} p_{i,j,t}^2 \tag{34}$$

Unter der Hypothese, dass die Unfälle unabhängig voneinander geschehen, ist der Varianzschätzer für den Totalschätzer der Lohnmasse (32) ebenfalls erwartungstreu:

$$\hat{V}ar\left(\hat{M}_{t,\pi}\right) = \sum_j \hat{V}ar\left(\hat{M}_{j,t,\pi}\right) \tag{35}$$

Da die Untersuchungsbereiche E_d aus disjunkten und unabhängigen Gruppen bestehen ist ein erwartungstreuer Varianzschätzer für den Lohnmassenschätzer in E_d

$$\hat{V}ar\left(\hat{M}_{d,t,\pi}\right) = \sum_{j \in J_d} \hat{V}ar\left(\hat{M}_{j,t,\pi}\right) \tag{36}$$

In den folgenden Formeln (37)-(42) ziehen wir die Schätzung von $\pi_{j,t}$ in die Schätzer der Formeln (31)-(36) mit ein. Der Schätzer für die Lohnmasse der Verunfallt-engruppe $V_{j,t}$ für die Periode t, mit geschätzter Einschlusswahrscheinlichkeit ist

$$\hat{M}_{j,t,\hat{\pi}} = \frac{1}{\hat{\pi}_{j,t}} \sum_{i \in V_{j,t}} p_{i,j,t} \tag{37}$$

Vorausgesetzt, dass die Einschlusswahrscheinlichkeit $\pi_{j,t}$ durch $\hat{\pi}_{j,t} = \frac{n_{j,t}}{\hat{N}_{j,t}}$ geschätzt wird, ist der Lohnmassenschätzer für die homogenen Gruppen

$$\hat{M}_{j,t,\hat{\pi}} = \frac{\hat{N}_{j,t}}{n_{j,t}} \sum_{i \in V_{j,t}} p_{i,j,t} = \hat{N}_{j,t}\nu_{j,t} \tag{38}$$

und der Lohnmassenschätzer über alle Gruppen ist

$$\hat{M}_{t,\hat{\pi}} = \sum_j \hat{M}_{j,t,\hat{\pi}} = \sum_j \hat{N}_{j,t}\nu_{j,t} \tag{39}$$

Die Lohnmassenschätzung (38) entspricht dem Vorschlag von LOHN.
Da wir die Einschlusswahrscheinlichkeit in Wirklichkeit schätzen, ist die Erwartungstreue jedoch nicht mehr gesichert. Die beiden Schätzer sind jedoch konsistent, falls $\hat{\pi}_{j,t} \longrightarrow \pi_{j,t}$.

Unter der Bedingung, dass $\hat{\pi}_{j,t} \longrightarrow \pi_{j,t}$ sind die Varianzschätzer der Schätzer für die Lohnmasse konsistent:
Anhand von (31) und unter der Bedingung, dass $\hat{\pi}_{j,t} \longrightarrow \pi_{j,t}$ und $\hat{\pi}_{j,t} = \frac{n_{j,t}}{\hat{N}_{j,t}} \doteq \frac{1}{\hat{g}_{j,t}}$ ist der Varianzschätzer des Schätzers für die Lohnmasse der homogenen Gruppen anhand einer *Taylor*-Linearisierung

$$
\begin{aligned}
\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right) &= \hat{V}ar\left(\hat{N}_{j,t}\nu_{j,t}\right) \\
&= \left(\sum_{i\in V_{j,t}} p_{i,j,t}\right)^2 \hat{V}ar\left(\frac{\hat{N}_{j,t}}{n_{j,t}}\right) = m_{j,t}^2 \hat{V}ar(\hat{g}_{j,t}) \\
&\approx \left(\frac{m_{j,t}}{n_{j,t}}\right)^2 \Big[\hat{V}ar(\hat{N}_{j,t}) + \hat{g}_{j,t}^2 \hat{V}ar(n_{j,t}) \\
&\quad -2\hat{g}_{j,t}\hat{C}ov(n_{j,t}, \hat{N}_{j,t})\Big] \\
&\longleftarrow \nu_{j,t}^2 \Big[\hat{V}ar(\hat{N}_{j,t}) + (1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t} \\
&\quad -2\hat{g}_{j,t}\hat{C}ov(n_{j,t}, \hat{N}_{j,t})\Big] \\
&= \nu_{j,t}^2 \Big[\hat{V}ar(\hat{N}_{j,t}) + (1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t} \\
&\quad -2\hat{g}_{j,t}\hat{\rho}_s(n_{j,t}, \hat{N}_{j,t})\sqrt{\hat{V}ar(n_{j,t})\hat{V}ar(\hat{N}_{j,t})}\Big] \\
&= \nu_{j,t}^2 \Big[\hat{V}ar(\hat{N}_{j,t}) + (1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t} \\
&\quad -2\hat{\rho}_s(n_{j,t}, \hat{N}_{j,t})\sqrt{(1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t}\hat{V}ar(\hat{N}_{j,t})}\Big] \qquad (40)
\end{aligned}
$$

wobei $1-\hat{f}_{j,t} = 1 - \frac{n_{j,t}}{\hat{N}_{j,t}}$ die geschätzte Korrektur für die endliche Population (*fpc*[3]) ist und $\hat{\rho}_s(n_{j,t}, \hat{N}_{j,t})$ der Schätzer der Stichprobenkorrelation zwischen $n_{j,t}$ und $\hat{N}_{j,t}$ ist.

Analog ist der Varianzschätzer des Schätzers für die gesamte Lohnmasse unter der Hypothese, dass die Unfälle unabhängig voneinander geschehen und die Popula-

[3] finite population correction

tionen der homogenen Gruppen untereinander unabhängig sind

$$\begin{aligned}\hat{V}ar\left(\hat{M}_{t,\hat{\pi}}\right) &= \sum_j \hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right) \\ &\longleftarrow \sum_j \nu_{j,t}^2 \Big[\hat{V}ar(\hat{N}_{j,t}) + (1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t} \\ &\quad -2\hat{\rho}_s(n_{j,t},\hat{N}_{j,t})\sqrt{(1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t}\hat{V}ar(\hat{N}_{j,t})}\Big] \quad (41)\end{aligned}$$

Analog zu (41) wird auch der Varianzschätzer des Lohnmassenschätzers im Untersuchungsbereich E_d geschätzt.

$$\begin{aligned}\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) &= \sum_{j\in J_d} \hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right) \\ &\longleftarrow \sum_{j\in J_d} \nu_{j,t}^2 \Big[\hat{V}ar(\hat{N}_{j,t}) + (1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t} \\ &\quad -2\hat{\rho}_s(n_{j,t},\hat{N}_{j,t})\sqrt{(1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t}\hat{V}ar(\hat{N}_{j,t})}\Big] \quad (42)\end{aligned}$$

4.2.4 Schätzung des Durchschnittslohns anhand einer "Stichprobe"

Analog zum obigen Abschnitt werden wir auch die Schätzer des Durchschnittslohns beschreiben. Dabei werden wir die Schätzung von $\pi_{j,t}$ von Beginn an mit einbeziehen.
Der Schätzer für den Durchschnittslohn in der Verunfalltengruppe $V_{j,t}$ ist anhand von (38)

$$\hat{\mu}_{j,t,\hat{\pi}} = \frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{N}_{j,t}} = \nu_{j,t} \quad (43)$$

gleichermassen ist der Schätzer für den Durchschnittslohn aller Lohnempfänger

$$\hat{\mu}_{t,\hat{\pi}} = \frac{\sum_j \hat{M}_{j,t,\hat{\pi}}}{\sum_j \hat{N}_{j,t}} = \frac{\hat{M}_{t,\hat{\pi}}}{\hat{N}_t} \quad (44)$$

Der Varianzschätzer für den Schätzer des Durchschnittslohns in der Verunfallteengruppe $V_{j,t}$ kann folgendermassen anhand einer *Taylor*-Linearisierung ausge-

drückt werden:

$$\begin{aligned}\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right) = \hat{V}ar(\nu_{j,t}) &= \hat{V}ar\left(\frac{m_{j,t}}{n_{j,t}}\right) = m_{j,t}^2\hat{V}ar\left(\frac{1}{n_{j,t}}\right) \\ &\approx \left(\frac{m_{j,t}}{n_{j,t}}\right)^2 \frac{1}{n_{j,t}^2}\hat{V}ar(n_{j,t}) \\ &\longleftarrow \frac{\nu_{j,t}^2}{n_{j,t}}(1-\hat{f}_{j,t}) \end{aligned} \tag{45}$$

Der Varianzschätzer für den Schätzer des Durchschnittslohns über alle Gruppen hinweg wird analog zu (45) entwickelt

$$\hat{V}ar\left(\hat{\mu}_{t,\hat{\pi}}\right) = \hat{V}ar(\nu_t) \longleftarrow \frac{\nu_t^2}{n_t}(1-\hat{f}_t) \tag{46}$$

wobei $m_t = \sum_j m_{j,t}$, $n_t = \sum_j n_{j,t}$, $\nu_t = \frac{m_t}{n_t}$, $\hat{N}_t = \sum_j \hat{N}_{j,t}$, $\hat{f}_t = \frac{n_t}{\hat{N}_t}$.

4.2.5 Schätzung des Lohnmassenindex

Wir definieren folgenden Wertindex für die Veränderung der Lohnmasse in der Verunfalltengruppe $V_{j,t}$ bezüglich der Referenzperiode t_0:

$$\hat{I}_{j,t,t_0}^M = \frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}} = \frac{\hat{N}_{j,t}\nu_{j,t}}{\hat{N}_{j,t_0}\nu_{j,t_0}} \tag{47}$$

Anhand von (10) ist der gewichtete Lohnmassenindex

$$\begin{aligned}\hat{I}_{t,t_0}^M &= \sum_j w_{j,t_o}\hat{I}_{j,t,t_0}^M = \sum_j w_{j,t_o}\frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}} \\ &= \sum_j \frac{\hat{M}_{j,t_0,\hat{\pi}}}{\sum_j \hat{M}_{j,t_0,\hat{\pi}}}\frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}} = \frac{\hat{M}_{t,\hat{\pi}}}{\hat{M}_{t_0,\hat{\pi}}} \end{aligned} \tag{48}$$

Mit der Gewichtung (7) ist der Schätzer des Lohnmassenindex für den Untersuchungsbereich E_d analog zu (12)

$$\hat{I}_{d,t,t_0}^M = \frac{\hat{M}_{d,t,\hat{\pi}}}{\hat{M}_{d,t_0,\hat{\pi}}} \tag{49}$$

Ein Qualitätsmass für den Lohnmassenindex (47) kann anhand einer Approximation durch eine *Taylor*-Linearisierung berechnet werden. Dabei wurde der Quotient $\frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}}$ entwickelt.

$$\begin{aligned}\hat{V}ar\left(\hat{I}^M_{j,t,t_0}\right) &\approx \frac{1}{\left(\hat{M}_{j,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)+\left(\frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}}\right)^2\hat{V}ar\left(\hat{M}_{j,t_0,\hat{\pi}}\right)\right.\\ &\quad\left.-2\frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}}\hat{C}ov\left(\hat{M}_{j,t_0,\hat{\pi}},\hat{M}_{j,t,\hat{\pi}}\right)\right] \qquad (50)\end{aligned}$$

Wir müssen davon ausgehen, dass $V_{j,t} \neq V_{j,t_0}$, das heisst, dass die Verunfalltengruppe j der Periode t nicht die gleiche ist, wie die Verunfalltengruppe der Referenzperiode. Daher kann die Kovarianz in (50) nicht geschätzt werden. Unter der Hypothese, dass die Unfallhäufigkeit und die Populationsgrösse unabhängig von der Erhebungsperiode sind, sind die Lohnmassenschätzer $\hat{M}_{j,t_0,\hat{\pi}}$ und $\hat{M}_{j,t,\hat{\pi}}$ unabhängig voneinander, und somit ist $\hat{C}ov\left(\hat{M}_{j,t_0,\hat{\pi}},\hat{M}_{j,t,\hat{\pi}}\right)=0$. Unter dieser Hypothese ist der Varianzschätzer für den Schätzer des Lohnmassenindex pro Gruppe folgendermassen definiert

$$\breve{V}ar\left(\hat{I}^M_{j,t,t_0}\right) \doteq \frac{1}{\left(\hat{M}_{j,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)+\left(\hat{I}^M_{j,t,t_0}\right)^2\hat{V}ar\left(\hat{M}_{j,t_0,\hat{\pi}}\right)\right] \qquad (51)$$

Das Vertrauensintervall für den Lohnmassenindex pro homogene Gruppe ist anhand von (51)

$$\hat{I}^M_{j,t,t_0} \in L^M_{j,t,\hat{\pi}} = \left[\hat{I}^M_{j,t,t_0} - z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^M_{j,t,t_0}\right)}, \hat{I}^M_{j,t,t_0} + z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^M_{j,t,t_0}\right)}\right] \qquad (52)$$

Das Vertrauensintervall liefert uns eine Information, in welchem Bereich sich $\hat{I}^M_{j,t,t_0}$ befindet.
Oft wird die Veränderung in bezug auf die vorhergehende Periode publiziert, d.h. $\Delta^M_{j,t} = \hat{I}^M_{j,t,t_0} - \hat{I}^M_{j,t-1,t_0}$. Der Varianzschätzer der Veränderung $\Delta^M_{j,t}$ ist

$$\begin{aligned}\hat{V}ar(\Delta^M_{j,t}) &= \hat{V}ar\left(\hat{I}^M_{j,t,t_0} - \hat{I}^M_{j,t-1,t_0}\right)\\ &= \hat{V}ar\left(\hat{I}^M_{j,t,t_0}\right)+\hat{V}ar\left(\hat{I}^M_{j,t-1,t_0}\right)\\ &\quad -2\hat{C}ov\left(\hat{I}^M_{j,t,t_0},\hat{I}^M_{j,t-1,t_0}\right) \qquad (53)\end{aligned}$$

Die Kovarianz aus (53) kann folgendermassen ausgedrückt werden

$$\hat{C}ov\left(\hat{I}^M_{j,t,t_0}, \hat{I}^M_{j,t-1,t_0}\right) = \hat{\rho}_s\left(\hat{I}^M_{j,t,t_0}, \hat{I}^M_{j,t-1,t_0}\right)\sqrt{\hat{V}ar\left(\hat{I}^M_{j,t,t_0}\right)\hat{V}ar\left(\hat{I}^M_{j,t-1,t_0}\right)} \tag{54}$$

wobei $\hat{\rho}_s$ die geschätzte Stichprobenkorrelation ist. Der Schätzer der Veränderung zieht somit auch die Information der Gemeinsamkeiten der 'Stichprobenpläne' der Lohnmassenindices zweier aufeinanderfolgender Perioden mit ein.
Unter der Annahme, dass $\hat{\sigma}_{j,t} = \hat{\sigma}_{j,t-1}$ ist, kann gezeigt werden, dass

$$\check{I}^M_{j,t,t_0} \in L^M_{j,t,t-1,\hat{\pi}} \quad = \quad \left[\hat{I}^M_{j,t,t_0} - 2z_{\alpha/2}\hat{\sigma}_{j,t}(1-\hat{\rho}_s), \hat{I}^M_{j,t,t_0} + 2z_{\alpha/2}\hat{\sigma}_{j,t}(1-\hat{\rho}_s)\right] \tag{55}$$

Wenn $\hat{\rho}_s \in [0,1]$ und ohne Bedingung für $\hat{\sigma}_{j,t}$ und $\hat{\sigma}_{j,t-1}$, ist

$$\hat{V}ar(\Delta^M_{j,t}) \in \left[\left(\hat{\sigma}_{j,t} - \hat{\sigma}_{j,t-1}\right)^2, \hat{\sigma}^2_{j,t} + \hat{\sigma}^2_{j,t-1}\right] \tag{56}$$

wobei $\hat{\sigma}^2_{j,t} = \hat{V}ar\left(\hat{I}^M_{j,t,t_0}\right)$ und $\hat{\sigma}^2_{j,t-1} = \hat{V}ar\left(\hat{I}^M_{j,t-1,t_0}\right)$ ist.
Wenn $\hat{\rho}_s < 0$ dann ist

$$\hat{V}ar(\Delta^M_{j,t}) \in \left(\hat{\sigma}^2_{j,t} + \hat{\sigma}^2_{j,t-1}, \left(\hat{\sigma}_{j,t} + \hat{\sigma}_{j,t-1}\right)^2\right] \tag{57}$$

Es sei $\Delta^M_{j,t} \geq 0$, $\hat{\rho}_s = 1$, $\hat{\sigma}_{j,t} - \hat{\sigma}_{j,t-1} \geq 0$, dann definieren wir

$$\begin{aligned}\Delta_{max} &= \Delta^M_{j,t} + z_{\alpha/2}\left(\hat{\sigma}_{j,t} - \hat{\sigma}_{j,t-1}\right)\\ \Delta_{min} &= \Delta^M_{j,t} - z_{\alpha/2}\left(\hat{\sigma}_{j,t} - \hat{\sigma}_{j,t-1}\right)\end{aligned}$$

Mit dieser Notation können wir ein Vertrauensintervall für den Schätzer des Lohnmassenindex, den wir $\check{I}^M_{j,t,t_0}$ nennen, beschreiben, ohne $\check{I}^M_{j,t,t_0}$ zu schätzen. Das Vertrauensintervall basiert auf dem Lohnmassenindex der Vorperiode und der geschätzten Veränderung.

$$\begin{aligned}\check{I}^M_{j,t,t_0} \in L^M_{j,t,t-1,\hat{\pi}} &= \left[\hat{I}^M_{j,t-1,t_0} + \Delta_{min}, \hat{I}^M_{j,t-1,t_0} + \Delta_{max}\right]\\ &= \left[\hat{I}^M_{j,t,t_0} - z_{\alpha/2}\left(\hat{\sigma}_{j,t} - \hat{\sigma}_{j,t-1}\right), \hat{I}^M_{j,t,t_0} + z_{\alpha/2}\left(\hat{\sigma}_{j,t} - \hat{\sigma}_{j,t-1}\right)\right]\end{aligned} \tag{58}$$

Mit der gleichen Notation ist das Vertrauensintervall (52)

$$\hat{I}^M_{j,t,t_0} \in L^M_{j,t,\hat{\pi}} = \left[\hat{I}^M_{j,t,t_0} - z_{\alpha/2}\hat{\sigma}_{j,t}, \hat{I}^M_{j,t,t_0} + z_{\alpha/2}\hat{\sigma}_{j,t}\right] \tag{59}$$

Somit ist unter den oben genannten Voraussetzungen $L^M_{j,t,t-1,\hat{\pi}} \subseteq L^M_{j,t,\hat{\pi}}$, womit gesagt ist, dass ein Signifikanztest der Veränderung in diesem Falle konservativer ist, wenn wir uns auf $L^M_{j,t,\hat{\pi}}$ abstützen, als wenn wir uns auf $L^M_{j,t,t-1,\hat{\pi}}$ abstützen. Dadurch, dass wir die Information der Stichprobenkorrelation in die Schätzung mit einbeziehen, ist die daraus resultierende Schätzung des Vertrauensintervalls genauer als wenn wir diese Information nicht verwenden würden. [4].

Unter den gleichen Voraussetzungen wie oben, aber mit der Annahme, dass $\hat{\rho}_s = -1$ ist, kann durch die gleiche Herleitung gezeigt werden, dass

$$\check{I}^M_{j,t,t_0} \in L^M_{j,t,t-1,\hat{\pi}} \quad = \quad \left[\hat{I}^M_{j,t,t_0} - z_{\alpha/2}\left(\hat{\sigma}_{j,t} + \hat{\sigma}_{j,t-1}\right), \hat{I}^M_{j,t,t_0} + z_{\alpha/2}\left(\hat{\sigma}_{j,t} + \hat{\sigma}_{j,t-1}\right)\right] \tag{60}$$

Das Vertrauensintervall für den Schätzer des Lohnmassenindex $\hat{I}^M_{j,t,t_0}$ hat nach wie vor die Form in (59). Somit ist unter diesen Voraussetzungen $L^M_{j,t,\hat{\pi}} \subseteq L^M_{j,t,t-1,\hat{\pi}}$ und ein Signifikanztest der Veränderung darf nicht auf $L^M_{j,t,\hat{\pi}}$ basieren, da dies zu einem falschen Schluss führen kann. Wir empfehlen daher immer $L^M_{j,t,t-1,\hat{\pi}}$ zu verwenden, um eine Veränderung auf ihre Signifikanz zu testen.
Die Beobachtungen bezüglich der Veränderung von $\hat{I}^M_{j,t,t_0}$ kann ohne weiteres auf $\hat{I}^M_{d,t,t_0}$ und $\hat{I}^M_{t,t_0}$ angewandt werden, da sie unabhängig vom Aggragationsniveau der Lohngruppen sind. Ebenso können diese Beobachtungen an nicht aufeinanderfolgende Perioden angewandt werden. Für den Vergleich zweier Lohnmassenindices in der gleichen Periode verwenden wir in den obigen Formeln $\Lambda^M_{j_1,j_2,t} = \hat{I}^M_{j_1,t,t_0} - \hat{I}^M_{j_2,t,t_0}$ an der Stelle von $\Delta^M_{j,t}$ und für die Varianzen der Lohnmassenindices schreiben wir $\hat{\sigma}^2_{j_1,t}$ und $\hat{\sigma}^2_{j_2,t}$.
Der Lohnmassenindex der Lohngruppen erlaubt es, die Referenzperiode beliebig zu ändern. Dabei wird der Quotient zweier Lohnmassenindices der Lohngruppen geschätzt. Die Erfassungsperiode des Nenners wird somit zur Referenzperiode. Zur Veranschaulichung werden wir dies an zwei Lohnmassenindices aufzeigen,

[4]Korrelieren zwei aufeinanderfolgende Lohnmassenindices positiv, so können wir einen konservativen Signifikanztest für die Indexveränderung folgendermassen formulieren: eine Indexveränderung innerhalb der Gruppe j von der Periode $t-1$ zur Periode t ist mit einem Fehler erster Art von 5% signifikant ($z_{\alpha/2} = 1.96$), wenn $L^M_{j,t-1,\hat{\pi}} \cap L^M_{j,t,\hat{\pi}} = \emptyset$. Bei negativer Korrelation ist dieser Test jedoch falsch, da in diesem Falle die Vertrauensintervalle, deren Schätzungen sämtliche vorhandene Information mit einbeziehen grösser als $L^M_{j,t-1,\hat{\pi}}$ und $L^M_{j,t,\hat{\pi}}$ sind.

die zeitlich aufeinanderfolgen, siehe (61). Es sei

$$\hat{Q}^M_{j,t,\hat{\pi}} \doteq \frac{\hat{I}^M_{j,t,t_0}}{\hat{I}^M_{j,t-1,t_0}} = \frac{\hat{M}_{j,t_0,\hat{\pi}}\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}\hat{M}_{j,t-1,\hat{\pi}}} = \frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t-1,\hat{\pi}}} = \hat{I}^M_{j,t,t-1} \tag{61}$$

$$\begin{aligned}\hat{Var}\left(\hat{Q}^M_{j,t,\hat{\pi}}\right) \approx & \left(\frac{1}{\hat{M}_{j,t-1,\hat{\pi}}}\right)^2 \left[\hat{Var}(\hat{M}_{j,t,\hat{\pi}}) + \left(\hat{Q}^M_{j,t,\hat{\pi}}\right)^2 \hat{Var}(\hat{M}_{j,t-1,\hat{\pi}}) \right. \\ & \left. -2\hat{Q}^M_{j,t,\hat{\pi}}\hat{Cov}\left(\hat{M}_{j,t,\hat{\pi}}, \hat{M}_{j,t-1,\hat{\pi}}\right)\right] \end{aligned} \tag{62}$$

Der Varianzschätzer in (62) wurde dabei mit einer *Taylor*-Linearisierung geschätzt. Unter der Hypothese, dass die Lohnmassenschätzer $\hat{M}_{j,t,\hat{\pi}}$ und $\hat{M}_{j,t-1,\hat{\pi}}$ unabhängig voneinander sind, ist $\hat{Cov}\left(\hat{M}_{j,t,\hat{\pi}}, \hat{M}_{j,t-1,\hat{\pi}}\right) = 0$. Somit kann die Qualität des Schätzers zum Beispiel durch den Variationskoeffizienten von $\hat{Q}^M_{j,t,\hat{\pi}}$ ausgedrückt werden:

$$cv(\hat{Q}^M_{j,t,\hat{\pi}}) = \frac{\sqrt{\hat{Var}\left(\hat{Q}^M_{j,t,\hat{\pi}}\right)}}{\hat{Q}^M_{j,t,\hat{\pi}}}.$$

Der Varianzschätzer für den Schätzer des Lohnmassenindex über alle Gruppen wird unter den gleichen Hypothesen und anhand der gleichen Schätzmethode wie (51) folgendermassen definiert

$$\breve{Var}\left(\hat{I}^M_{t,t_0}\right) \doteq \frac{1}{\left(\hat{M}_{t_0,\hat{\pi}}\right)^2}\left[\hat{Var}\left(\hat{M}_{t,\hat{\pi}}\right) + \left(\hat{I}^M_{t,t_0}\right)^2 \hat{Var}\left(\hat{M}_{t_0,\hat{\pi}}\right)\right] \tag{63}$$

Das entsprechende Vertrauensintervall ist somit

$$\hat{I}^M_{t,t_0} \in L^M_{t,\hat{\pi}} = \left[\hat{I}^M_{t,t_0} - z_{\alpha/2}\sqrt{\hat{Var}\left(\hat{I}^M_{t,t_0}\right)}, \hat{I}^M_{t,t_0} + z_{\alpha/2}\sqrt{\hat{Var}\left(\hat{I}^M_{t,t_0}\right)}\right] \tag{64}$$

Die Veränderung des Lohnmassenindex bezüglich der vorhergegangenen Periode wird wie in (53) bis (60) geschätzt und die gleichen Schlussfolgerungen sind auch für $\hat{I}^M_{t,t_0}$ gültig. Das heisst, es sollte immer das Vertrauensintervall $L^M_{t,t-1,\hat{\pi}}$ verwendet werden, anstatt dem Vertrauensintervall $L^M_{t,\hat{\pi}}$, sobald eine Veränderung des Lohnmassenindex bezüglich der vorhergegangenen Periode interpretiert werden soll.

Der entsprechende Lohnmassenindex bezüglich der vorhergegangenen Periode ist

$$\hat{Q}^M_{t,\hat{\pi}} = \frac{\hat{I}^M_{t,t_0}}{\hat{I}^M_{t-1,t_0}} = \frac{\hat{M}_{t,\hat{\pi}}}{\hat{M}_{t-1,\hat{\pi}}} \tag{65}$$

$$\begin{aligned}\hat{V}ar\left(\hat{Q}^M_{t,\hat{\pi}}\right) \approx & \left(\frac{1}{\hat{M}_{t-1,\hat{\pi}}}\right)^2 \left[\hat{V}ar(\hat{M}_{t,\hat{\pi}}) + \left(\hat{Q}^M_{t,\hat{\pi}}\right)^2 \hat{V}ar(\hat{M}_{t-1,\hat{\pi}})\right. \\ & \left. -2\hat{Q}^M_{t,\hat{\pi}}\hat{C}ov\left(\hat{M}_{t,\hat{\pi}}, \hat{M}_{t-1,\hat{\pi}}\right)\right]\end{aligned} \tag{66}$$

Der Varianzschätzer in (66) wurde dabei mit einer *Taylor*-Linearisierung geschätzt. Unter der Hypothese, dass die Lohnmassenschätzer $\hat{M}_{t,\hat{\pi}}$ und $\hat{M}_{t-1,\hat{\pi}}$ unabhängig voneinander sind, ist $\hat{C}ov\left(\hat{M}_{t,\hat{\pi}}, \hat{M}_{t-1,\hat{\pi}}\right) = 0$. Somit kann die Qualität der Schätzers der Veränderung durch den Variationskoeffizienten von $\hat{Q}^M_{t,\hat{\pi}}$ ausgedrückt werden:

$$cv(\hat{Q}^M_{t,\hat{\pi}}) = \frac{\sqrt{\hat{V}ar\left(\hat{Q}^M_{t,\hat{\pi}}\right)}}{\hat{Q}^M_{t,\hat{\pi}}}.$$

Der Varianzschätzer des Schätzers des Lohnmassenindex des Untersuchungsbereichs E_d wird mit einer *Taylor*-Linearisierung angenähert.

$$\begin{aligned}\hat{V}ar\left(\hat{I}^M_{d,t,t_0}\right) \approx & \frac{1}{\left(\hat{M}_{d,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) + \left(\hat{I}^M_{d,t,t_0}\right)^2 \hat{V}ar\left(\hat{M}_{d,t_0,\hat{\pi}}\right)\right. \\ & \left. -2\hat{I}^M_{d,t,t_0}\hat{C}ov\left(\hat{M}_{d,t,\hat{\pi}}, \hat{M}_{d,t_0,\hat{\pi}}\right)\right] \\ = & \frac{1}{\left(\hat{M}_{d,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) + \left(\hat{I}^M_{d,t,t_0}\right)^2 \hat{V}ar\left(\hat{M}_{d,t_0,\hat{\pi}}\right)\right. \\ & \left. -2\hat{I}^M_{d,t,t_0}\hat{\rho}_s\left(\hat{M}_{d,t,\hat{\pi}}, \hat{M}_{d,t_0,\hat{\pi}}\right)\sqrt{\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right)\hat{V}ar\left(\hat{M}_{d,t_0,\hat{\pi}}\right)}\right]\end{aligned} \tag{67}$$

$$\breve{V}ar\left(\hat{I}^M_{d,t,t_0}\right) \doteq \frac{1}{\left(\hat{M}_{d,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) + \left(\hat{I}^M_{d,t,t_0}\right)^2 \hat{V}ar\left(\hat{M}_{d,t_0,\hat{\pi}}\right)\right] \tag{68}$$

Wenn davon ausgegangen werden kann, dass die Lohnmasse zum Zeitpunkt t unabhängig von der Lohnmasse zu Zeitpunkt t_0 erhoben wird, dann ist die Kovarianz

in (67) aus dem stichproben-theoretischen Ansatz 0. Dies heisst aber nicht, dass unter dem klassischen statistischen Ansatz keine Korrelation zwischen den beiden Grössen vorhanden ist.

4.2.6 Schätzung des Durchschnittslohnindex

Aufgrund der Definitionen im Abschnitt 3.1, der Definition des theoretischen Durchschnittslohnindex (5) und den Schätzungen des Durchschnittslohn (43) und (44) ist der Schätzer für den Durchschnittslohnindex über alle Gruppen

$$\hat{I}^D_{t,t_0,\hat{\pi}} = \frac{\sum\limits_{j \in N_v} \hat{N}_{j,t_0} \hat{\mu}_{j,t,\hat{\pi}}}{\sum\limits_{j \in N_v} \hat{N}_{j,t_0} \hat{\mu}_{j,t_0,\hat{\pi}}} \tag{69}$$

wobei der Schätzer des Durchschnittslohnindex pro Gruppe folgende Form annimmt

$$\hat{I}^D_{j,t,t_0,\hat{\pi}} = \frac{\hat{\mu}_{j,t,\hat{\pi}}}{\hat{\mu}_{j,t_0,\hat{\pi}}} \tag{70}$$

und der Durchschnittslohnindex im Untersuchungsbereich E_d mit der Gewichtung (7) nimmt folgende Form an

$$\hat{I}^D_{d,t,t_0,\hat{\pi}} = \frac{\sum\limits_{j \in J_d} \hat{N}_{j,t_0} \hat{\mu}_{j,t,\hat{\pi}}}{\sum\limits_{j \in J_d} \hat{N}_{j,t_0} \hat{\mu}_{j,t_0,\hat{\pi}}} \tag{71}$$

Unter der Voraussetzung dass $\hat{\pi}_{j,t} = \frac{n_{j,t}}{N_{j,t}} \quad \forall j,t$, ist die Schätzung des globalen Durchschnittslohnindex

$$\hat{I}^D_{t,t_0,\hat{\pi}} = \frac{\sum\limits_{j \in N_v} \hat{N}_{j,t_0} \nu_{j,t}}{\sum\limits_{j \in N_v} \hat{N}_{j,t_0} \nu_{j,t_0}} \tag{72}$$

Der Schätzer (72) entspricht dem *Laspeyres*-Index (5), welcher von LOHN bereits verwendet wird. Ebenso entspricht der Schätzer des Durchschnittslohnindex für die homogenen Gruppen dem bereits verwendeten Index (13)

$$\hat{I}^D_{j,t,t_0,\hat{\pi}} = \frac{\nu_{j,t}}{\nu_{j,t_0}} \tag{73}$$

Wie bereits im Abschnitt 4.2.5 verwenden wir auch hier eine *Taylor*-Linearisierung für die Schätzung der Varianzschätzer. Aufgrund von (45) und unter der Hypothese, dass die Unfallhäufigkeit unabhängig von der Erhebungsperiode ist, ergibt dies

für den Varianzschätzer des Schätzers für den Durchschnittslohnindex der homogenen Gruppen

$$\begin{aligned}\breve{V}ar\left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right) &= \breve{V}ar\left(\frac{\hat{\mu}_{j,t,\hat{\pi}}}{\hat{\mu}_{j,t_0,\hat{\pi}}}\right) \\ &\approx \frac{1}{\nu^2_{j,t_0}}\left[\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right) + \left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right)^2 \hat{V}ar\left(\hat{\mu}_{j,t_0,\hat{\pi}}\right)\right] \end{aligned} \quad (74)$$

Aufgrund von (45) kann der Varianzschätzer des Schätzers des Durchschnittslohns in (74) ersetzt werden. Somit wird der Varianzschätzer des Schätzers des Durchschnittslohnindex der Gruppe $V_{j,t}$ bezüglich der Referenzperiode t_0 folgendermassen geschätzt

$$\breve{V}ar\left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right) \approx \frac{1}{\nu^2_{j,t_0}}\left[(1-\hat{f}_{j,t})\frac{\nu^2_{j,t}}{n_{j,t}} + (1-\hat{f}_{j,t_0})\left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right)^2 \frac{\nu^2_{j,t_0}}{n_{j,t_0}}\right] \quad (75)$$

Das Vertrauensintervall für den Schätzer des Durchschnittslohnindex pro Gruppe ist

$$\hat{I}^D_{j,t,t_0} \in L^D_{j,t,\hat{\pi}} = \left[\hat{I}^D_{j,t,t_0} - z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^D_{j,t,t_0}\right)}, \hat{I}^d_{j,t,t_0} + z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^D_{j,t,t_0}\right)}\right] \quad (76)$$

Für den Schätzer des globalen Durchschnittslohnindex ist der Varianzschätzer anhand einer *Taylor*-Linearisierung und unter Unabhängigkeit der Populationsgrössen und Durchschnittslohnschätzer der Gruppen untereinander und in der Zeit

$$\begin{aligned}\breve{V}ar\left(\hat{I}^D_{t,t_0,\hat{\pi}}\right) &= \breve{V}ar\left(\frac{\sum_{j\in N_v}\hat{N}_{j,t_0}\nu_{j,t}}{\sum_{j\in N_v}\hat{N}_{j,t_0}\nu_{j,t_0}}\right) \\ &\approx \frac{1}{\left(\sum_{j\in N_v}\hat{N}_{j,t_0}\nu_{j,t_0}\right)^2}\left[\sum_{j\in N_v}\hat{V}ar\left(\hat{N}_{j,t_0}\nu_{j,t}\right)\right. \\ &\quad \left. + \left(\hat{I}^D_{t,t_0,\hat{\pi}}\right)^2 \sum_{j\in N_v}\hat{V}ar\left(\hat{N}_{j,t_0}\nu_{j,t_0}\right)\right] \end{aligned} \quad (77)$$

wobei, unter der Voraussetzung, dass die Daten der SSUV einer Bernoulli-Stichprobe gleichgesetzt werden können und aufgrund von (45) und einer *Taylor*-Linearisierung

$$\begin{aligned}\hat{V}ar\left(\hat{N}_{j,t_0}\nu_{j,t}\right) &\approx \nu_{j,t}^2\hat{V}ar\left(\hat{N}_{j,t_0}\right)+\hat{N}_{j,t_0}^2\hat{V}ar\left(\nu_{j,t}\right)\\ &= \nu_{j,t}^2\hat{V}ar\left(\hat{N}_{j,t_0}\right)+(1-\hat{f}_{j,t})\frac{\left(\hat{N}_{j,t_0}\nu_{j,t}\right)^2}{n_{j,t}} \qquad (78)\end{aligned}$$

und

$$\hat{V}ar\left(\hat{N}_{j,t_0}\nu_{j,t_0}\right) \approx \nu_{j,t_0}^2\hat{V}ar\left(\hat{N}_{j,t_0}\right)+(1-\hat{f}_{j,t_0})\frac{\left(\hat{N}_{j,t_0}\nu_{j,t_0}\right)^2}{n_{j,t_0}} \qquad (79)$$

Aus (77), (78) und (79) folgt sodann

$$\begin{aligned}\breve{V}ar\left(\hat{I}^D_{t,t_0,\hat{\pi}}\right) \approx \frac{1}{\hat{M}^2_{t_0,\hat{\pi}}}\Bigg[&\sum_{j\in N_v}\left(\nu_{j,t}^2+\left(\hat{I}^D_{t,t_0,\hat{\pi}}\nu_{j,t_0}\right)^2\right)\hat{V}ar\left(\hat{N}_{j,t_0}\right)+\\ &\sum_{j\in N_v}\left\{\frac{(1-\hat{f}_{j,t})}{n_{j,t}}\left(\hat{N}_{j,t_0}\nu_{j,t}\right)^2+\frac{(1-\hat{f}_{j,t_0})}{n_{j,t_0}}\left(\hat{I}^D_{t,t_0,\hat{\pi}}\hat{M}_{j,t_0,\hat{\pi}}\right)^2\right\}\Bigg] \qquad (80)\end{aligned}$$

Das Vertrauensintervall des Schätzers des Durchschnittslohnindex über alle Gruppen ist

$$\hat{I}^D_{t,t_0}\in L^D_{t,\hat{\pi}}=\left[\hat{I}^D_{t,t_0}-z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^D_{t,t_0}\right)},\hat{I}^D_{t,t_0}+z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^D_{t,t_0}\right)}\right] \qquad (81)$$

Der Varianzschätzer des Durchschnittslohnindex im Untersuchungsbereich E_d ist analog zu (80)

$$\begin{aligned}\breve{V}ar\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\right) \approx \frac{1}{\hat{M}^2_{d,t_0,\hat{\pi}}}\Bigg[&\sum_{j\in J_d}\left(\nu_{j,t}^2+\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\nu_{j,t_0}\right)^2\right)\hat{V}ar\left(\hat{N}_{j,t_0}\right)+\\ &\sum_{j\in J_d}\left\{\frac{(1-\hat{f}_{j,t})}{n_{j,t}}\left(\hat{N}_{j,t_0}\nu_{j,t}\right)^2+\frac{(1-\hat{f}_{j,t_0})}{n_{j,t_0}}\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\hat{M}_{j,t_0,\hat{\pi}}\right)^2\right\}\Bigg] \qquad (82)\end{aligned}$$

$$\hat{I}^D_{d,t,t_0}\in L^D_{d,t,\hat{\pi}}=\left[\hat{I}^D_{d,t,t_0}-z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^D_{d,t,t_0}\right)},\hat{I}^d_{d,t,t_0}+z_{\alpha/2}\sqrt{\breve{V}ar\left(\hat{I}^D_{d,t,t_0}\right)}\right] \qquad (83)$$

Die Resultate (53) bis (59) der Schätzung des Lohnmassenindex und die zugehörigen Kommentare können ohne weiteres an den Durchschnittslohnindex angewandt werden. Auf dem Niveau des Untersuchungsbereichs E_d ist somit das Vertrauensintervall folgendermassen definiert:
Unter der Voraussetzung, dass $\rho_s\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}, \hat{I}^D_{d,t-1,t_0,\hat{\pi}}\right) \in [0,1]$, ist

$$\hat{V}ar(\Delta^D_{d,t}) \in \left[(\hat{\sigma}^D_{d,t} - \hat{\sigma}^D_{d,t-1})^2, (\hat{\sigma}^D_{d,t})^2 - (\hat{\sigma}^D_{d,t-1})^2\right] \tag{84}$$

wobei $\hat{\sigma}^D_{d,t} = \breve{V}ar\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\right)$.
Ist $\hat{\sigma}^D_{d,t} = \hat{\sigma}^D_{d,t-1}$, so ist mit obiger Notation

$$\hat{V}ar(\Delta^D_{d,t}) \in \left[0, 4(\hat{\sigma}^D_{d,t})^2\right] \tag{85}$$

Der Varianzschätzer für den Vergleich von zwei Durchschnittslohnindices der gleichen Periode der Bereiche d_1 und d_2, z.B. $d_1 =$ (NOGA=65, Geschlecht=M, Beschäftigungsgrad=Vollzeit) und $d_2 =$ (NOGA=66, Geschlecht=F, Beschäftigungsgrad=Vollzeit), nimmt folgende Form an, wenn davon ausgegangen wird, dass die Schätzung von zwei Durchschnittslohnindices in disjunkten Bereichen unabhängig von einander ist

$$\begin{aligned}
\hat{V}ar(\Lambda^D_{d_1,d_2,t}) &\doteq \hat{V}ar\left(\hat{I}^D_{d_1,t,t_0,\hat{\pi}} - \hat{I}^D_{d_2,t,t_0,\hat{\pi}}\right) \\
&= \breve{V}ar\left(\hat{I}^D_{d_1,t,t_0,\hat{\pi}}\right) + \breve{V}ar\left(\hat{I}^D_{d_2,t,t_0,\hat{\pi}}\right) \\
&\quad \underbrace{-2\,\hat{C}ov\left(\hat{I}^D_{d_1,t,t_0,\hat{\pi}}, \hat{I}^D_{d_2,t,t_0,\hat{\pi}}\right)}_{=0} \\
&\doteq (\hat{\sigma}^D_{d_1,t})^2 + (\hat{\sigma}^D_{d_2,t})^2
\end{aligned} \tag{86}$$

Stützen wir den Vergleich auf $L^D_{d_1,t,\hat{\pi}}$ und $L^D_{d_2,t,\hat{\pi}}$ ab, so ist der Unterschied signifikant, wenn

$$\begin{aligned}
\Lambda^D_{d_1,d_2,t} &> z_{\alpha/2}(\hat{\sigma}^D_{d_1,t} + \hat{\sigma}^D_{d_2,t}) \\
&\overset{\substack{\text{Minkowski's}\\ \text{Ungleichheit}}}{\geq} z_{\alpha/2}\sqrt{(\hat{\sigma}^D_{d_1,t})^2 + (\hat{\sigma}^D_{d_2,t})^2}
\end{aligned} \tag{87}$$

Somit ist der Signifikanztest konservativer wenn wir ihn auf die Vertrauensintervalle $L^D_{d_1,t,\pi}$ und $L^D_{d_2,t,\pi}$ abstützen, als wenn wir ihn auf das Vertrauensintervall des Unterschieds basieren.
Für den Quotienten zweier aufeinanderfolgender Durchschnittslohnindices sehen wir nur eine Anwendung, wenn wir uns auf das Niveau der Lohngruppen beschränken. Dann entspricht dieser Quotient einer Neuindexierung auf der Periode

$t-1$:

$$\hat{Q}^D_{j,t,\hat{\pi}} \doteq \frac{\hat{I}^D_{j,t,t_0}}{\hat{I}^D_{j,t-1,t_0}} = \frac{\nu_{j,t}}{\nu_{j,t-1}} = \hat{I}^D_{j,t,t-1}$$

Auf dem Niveau der Untersuchungsbereiche oder des totalen Durchschnittslohnindex verliert dieser Quotient die Eigenschaft des *Laspeyres*-Index.

4.3 Einfluss der Einschlusswahrscheinlichkeit auf die Schätzer

Wie bereits in 4.2.2 erwähnt wurde, muss die Einschlusswahrscheinlichkeit geschätzt werden. Wir werden in diesem Abschnitt untersuchen, welchen Einfluss eine Abweichung in der Schätzung hat. Wir gehen dabei von der Hypothese aus, dass die Anzahl Verunfallter durch die Daten der SSUV bekannt und keine Zufallsvariable ist. Sofern eine Schichtung anhand der Variablen in Tabelle 3 gemacht werden soll, fehlen uns hingegen die Daten der Grundgesamtheit für die einzelnen Schichten. Mit anderen Worten, gehen wir von einem Bias für den Schätzer der Populationsgrösse in den einzelnen Schichten aus.

Durchschnittslohn- und Lohnmassenschätzer
Wir gehen von einer Abweichung der Anzahl Lohnempfänger der Grundgesamtheit in den Schichten aus. Diese Abweichung definieren wir folgendermassen:
$\xi_{j,t} =$ Abweichung der Anzahl Lohnempfänger in der Gruppe (Schicht) j für die Periode t, $\xi_{j,t} \in \mathbb{N}$. Wir gehen davon aus, dass $\sum_j \xi_{j,t} = 0$. Das heisst, dass wir davon ausgehen, dass die gesamte Anzahl Lohnempfänger keine Abweichung aufweist.
Die geschätzte Anzahl Lohnempfänger ist somit $\hat{N}_{j,t} = N_{j,t} + \xi_{j,t}$, wobei $\sum_j \hat{N}_{j,t} = N_t$. Diese Anzahl Lohnempfänger entspricht der Anzahl Lohnempfänger, welche derzeit verwendet wird. Die geschätzte Einschlusswahrscheinlichkeit aufgrund der geschätzten Anzahl Lohnempfänger ist

$$\tilde{\pi}_{j,t} = \frac{n_{j,t}}{N_{j,t} + \xi_{j,t}} \tag{88}$$

Der Lohnmassenschätzer für die homogenen Gruppen ist dann
$\tilde{M}_{j,t,\tilde{\pi}} = (N_{j,t} + \xi_{j,t})\nu_{j,t}$, und der Lohnmassenschätzer über alle Gruppen ist

$$\tilde{M}_{t,\tilde{\pi}} = \sum_j \tilde{M}_{j,t,\tilde{\pi}} = \hat{M}_{t,\hat{\pi}} + \sum_j \xi_{j,t}\nu_{j,t} \tag{89}$$

Somit weist der Lohnmassenschätzer einen Bias von $\sum_j \xi_{j,t}\nu_{j,t}$ auf, obwohl die gesamte Anzahl Lohnempfänger keine Abweichung aufweist. Der Varianzschätzer

des Schätzers für die Lohnmasse in den Gruppen ist

$$\hat{Var}\left(\tilde{M}_{j,t,\tilde{\pi}}\right) \longleftarrow \left(\frac{N_{j,t}+\xi_{j,t}}{n_{j,t}} - 1\right) \sum_{i \in V_{j,t}} p_{i,j,t}^2 \tag{90}$$

Wenn $\xi_{j,t} \ll n_{j,t}$, dann ist der Bias gering. Wenn zum Beispiel $\xi_{j,t} = n_{j,t}$, dann ist der Bias $\frac{2N_{j,t}}{n_{j,t}} \sum_{i \in V_{j,t}} p_{i,j,t}^2$, was ein enormes Ausmass annehmen könnte.

$$\hat{Var}\left(\tilde{M}_{t,\tilde{\pi}}\right) \longleftarrow \sum_j \hat{Var}\left(\tilde{M}_{j,t,\tilde{\pi}}\right) \tag{91}$$

Der Schätzer des Durchschnittslohns in den Gruppen ist

$$\tilde{\mu}_{j,t,\tilde{\pi}} = \frac{\tilde{M}_{j,t,\tilde{\pi}}}{N_{j,t}+\xi_{j,t}} = \nu_{j,t} \tag{92}$$

Der Schätzer des Durchschnittslohns in den Verunfalltengruppen $V_{j,t}$ ist somit robust gegenüber einer Abweichung der Anzahl Lohnempfänger.
Der Schätzer des Durchschnittslohns über alle Gruppen ist

$$\begin{aligned} \tilde{\mu}_{t,\tilde{\pi}} &= \frac{\tilde{M}_{t,\tilde{\pi}}}{\sum_j N_{j,t}+\xi_{j,t}} = \frac{\hat{M}_{t,\hat{\pi}} + \sum_j \xi_{j,t}\nu_{j,t}}{\sum_j N_{j,t}+\xi_{j,t}} \\ &= \frac{\hat{M}_{t,\hat{\pi}} + \sum_j \xi_{j,t}\nu_{j,t}}{N_t} = \nu_t + \frac{\sum_j \xi_{j,t}\nu_{j,t}}{N_t} \end{aligned} \tag{93}$$

Die Varianzen der Schätzer des Durchschnittslohns werden folgendermassen geschätzt

$$\hat{Var}\left(\tilde{\mu}_{j,t,\tilde{\pi}}\right) \longleftarrow \frac{1}{N_{j,t}+\xi_{j,t}^2} \left(\frac{N_{j,t}+\xi_{j,t}}{n_{j,t}} - 1\right) \sum_{i \in V_{j,t}} p_{i,j,t}^2 \tag{94}$$

$$\hat{Var}\left(\tilde{\mu}_{t,\tilde{\pi}}\right) \longleftarrow \sum_j \frac{1}{N_{j,t}+\xi_{j,t}^2} \left(\frac{N_{j,t}+\xi_{j,t}}{n_{j,t}} - 1\right) \sum_{i \in V_{j,t}} p_{i,j,t}^2 \tag{95}$$

Die Varianzschätzer sind nicht robust gegenüber einer Abweichung der Anzahl Lohnempfänger.

Schätzung der Lohnindices

Aufgrund der Schätzer im obigen Abschnitt, werden wir den Einfluss einer Abweichung auf die Anzahl der Lohnempfänger für die Lohnindices formalisieren. Wir bezeichnen den Lohnmassenindex für die Verunfalltengruppe $V_{j,t}$ mit $\tilde{I}_{j,t,t_0}^M$

und den Lohnmassenindex über sämtliche Gruppen mit $\tilde{I}^M_{t,t_0}$. Die Schätzung des Lohnmassenindex für die Gruppen, unter Berücksichtigung einer Abweichung der Anzahl der Lohnempfänger ist

$$\tilde{I}^M_{j,t,t_0} = \frac{(N_{j,t} + \xi_{j,t})\nu_{j,t}}{(N_{j,t_0} + \xi_{j,t_0})\nu_{j,t_0}} \tag{96}$$

Wir bilden den Quotienten von (96) und (47), um den Einfluss der Abweichung auf den Lohnmassenindex zu evaluieren:

$$\begin{aligned} \frac{\tilde{I}^M_{j,t,t_0}}{\hat{I}^M_{j,t,t_0}} &= \frac{(N_{j,t} + \xi_{j,t})\nu_{j,t}}{(N_{j,t_0} + \xi_{j,t_0})\nu_{j,t_0}} \frac{N_{j,t_0}\nu_{j,t_0}}{N_{j,t}\nu_{j,t}} \\ &= \frac{N_{j,t_0}}{N_{j,t}} \frac{(N_{j,t} + \xi_{j,t})}{(N_{j,t_0} + \xi_{j,t_0})} \end{aligned} \tag{97}$$

Aus (97) geht hervor, dass die Abweichung keinen Einfluss auf die Schätzung des Lohnmassenindex für die Gruppen hat, wenn $N_{j,t} = N_{j,t_0}, \forall j,t$ und wenn $N_{j,t} + \xi_{j,t} = N_{j,t_0} + \xi_{j,t_0}, \forall j,t$. Die zweite Bedingung ist in der derzeitigen Schätzmethode des Durchschnittslohns immer erfüllt, da für sämtliche Perioden t die geschätzte Anzahl Lohnempfänger der Referenzperiode verwendet werden. Die erste Bedingung kann aufgrund des strukturellen Wandels der Anzahl Lohnempfänger in den einzelnen Gruppen nicht erfüllt werden.
Die Schätzung des Lohnmassenindex über alle Gruppen ist

$$\begin{aligned} \tilde{I}^M_{t,t_0} &= \frac{\tilde{M}_{t,\tilde{\pi}}}{\tilde{M}_{t_0,\tilde{\pi}}} \\ &= \frac{\hat{M}_{t,\hat{\pi}} + \sum_j \xi_{j,t}\nu_{j,t}}{\hat{M}_{t_0,\hat{\pi}} + \sum_j \xi_{j,t_0}\nu_{j,t_0}} \end{aligned} \tag{98}$$

Der Quotient von (98) und (48) ist

$$\frac{\tilde{I}^M_{t,t_0}}{\hat{I}^M_{t,t_0}} = \frac{\hat{M}_{t_0,\hat{\pi}}\tilde{M}_{t,\tilde{\pi}}}{\hat{M}_{t,\hat{\pi}}\tilde{M}_{t_0,\tilde{\pi}}} \tag{99}$$

Aus (99) geht hervor, dass die beiden Schätzer des Lohnmassenindex nur dann dieselben sind, wenn die Abweichung $\xi_{j,t} = 0, \forall j,t$, ist. Dies ist jedoch nicht gegeben.

Abweichungen der Anzahl der Lohnempfänger in den Gruppen haben keinen Einfluss auf die Schätzung des Durchschnittslohnindex für die Gruppen.

$$\tilde{I}^D_{j,t,t_0,\tilde{\pi}} = \hat{I}^D_{j,t,t_0,\hat{\pi}} = \frac{\nu_{j,t}}{\nu_{j,t_0}} \tag{100}$$

Der Schätzer des gewichteten Durchschnittslohnindex über alle Gruppen mit Berücksichtigung der Abweichung der Anzahl Lohnempfänger ist nach wie vor ein *Laspeyres*-Index

$$\tilde{I}^D_{t,t_0,\tilde{\pi}} = \frac{\sum\limits_j (N_{j,t_0} + \xi_{j,t_0})\nu_{j,t}}{\sum\limits_j (N_{j,t_0} + \xi_{j,t_0})\nu_{j,t_0}} \tag{101}$$

Der Quotient von (72) und (101) ist

$$\begin{aligned} \frac{\tilde{I}^D_{t,t_0,\tilde{\pi}}}{\hat{I}^D_{t,t_0,\hat{\pi}}} &= \frac{\sum\limits_j (N_{j,t_0} + \xi_{j,t_0})\nu_{j,t}}{\sum\limits_j (N_{j,t_0} + \xi_{j,t_0})\nu_{j,t_0}} \frac{\sum\limits_j N_{j,t_0}\nu_{j,t_0}}{\sum\limits_j N_{j,t_0}\nu_{j,t}} \\ &= \frac{\hat{M}_{t_0\hat{\pi}} \sum\limits_j N_{j,t_0}\nu_{j,t} + \sum\limits_j \xi_{j,t_0}\nu_{j,t} \sum\limits_j N_{j,t_0}\nu_{j,t_0}}{\hat{M}_{t_0\hat{\pi}} \sum\limits_j N_{j,t_0}\nu_{j,t} + \sum\limits_j \xi_{j,t_0}\nu_{j,t_0} \sum\limits_j N_{j,t_0}\nu_{j,t}} \end{aligned} \tag{102}$$

Aus (102) ist ersichtlich, dass die beiden Schätzer nur dann gleich sind, wenn $\nu_{j,t} = \nu_{j,t_0}, \forall j,t$ oder $\xi_{j,t} = 0, \forall j,t$. Wir gehen aber davon aus, dass sich der Durchschnittslohn verändert.

5 Modell-Methode

Bei der Modell-Methode werden die Schätzer aus einem postulierten stochastischen Modell für den Lohn abgeleitet. Das Ziel ist, die Lohnmasse anhand von Hilfsvariablen zum Zeitpunkt t zu modellisieren. Zu diesem Zeitpunkt muss die Lohnmasse vorhanden sein. Für die Perioden $t+h$, $h \in \mathbb{N}$ wird das geschätzte Modell verwendet, um die Lohnmasse aufgrund der Hilfsvariablen zu schätzen.
Im folgenden Abschnitt stellen wir eine Modell-Methode vor, welche auf einer Regression basiert. Im darauf folgenden Abschnitt stellen wir eine Methode vor, die auf einer Kalibrierung basiert.

5.1 Regressions-Methode

5.1.1 Definitionen

Es sei
p : die Anzahl Hilfsvariablen, welche zur Verfügung stehen.
$\mathbf{x}_{j,t}$: der Vektor der p Hilfsvariablen, welche für sämtliche Personen $i \in V_{j,t}$ und für sämtliche Perioden t bekannt sind, $\mathbf{x}_{j,t} \in \mathbb{R}^p$.
θ_j : der Vektor der Parameter des Modells für die Gruppe j, $\theta_j \in \mathbb{R}^{(q)}$, wobei $q \geq p$.

$f_{\theta_j}()$: die Modellfunktion, $f : \mathbb{R}^p \to \mathbb{R}_+$.
$\varepsilon_{j,t}$: statistischer Fehler, so dass $E(\varepsilon_{j,t}) = 0,\ E(\varepsilon_{j,t}\varepsilon_{k,t}) = 0$ wenn $j \neq k$ und $E(\varepsilon_{j,t}\varepsilon_{j,t}) = \sigma_{j,t}$.
Die Lohnmasse für die Lohngruppe $U_{j,t}$ ist somit

$$M_{j,t} = f_{\theta_j}(\mathbf{x}_{j,t}) + \varepsilon_{j,t} \tag{103}$$

Beispiel 1: Ein lineares Modell, das auf der Populationsgrösse in den Lohngruppen basiert ist

$$\hat{M}_{j,t} = \hat{\theta}_0 + \hat{\theta}_1 N_{j,t}$$

In diesem Modell sind die entsprechenden Vektoren der Hilfsvariablen und der Parameter $\mathbf{x}'_{j,t} = (1, N_{j,t})$, und $\hat{\theta}_j = \left(\hat{\theta}_0, \hat{\theta}_1\right)$, $\forall j$. Mit $\hat{\theta}_0 = 0$ schätzen wir mit diesem Modell den Durchschnittslohn $\hat{\theta}_1 = \hat{\mu}_{j,t}$.
Beispiel 2: Ein Modell, welches auf dem Durchschnittslohn in der homogenen Gruppe $V_{j,t} \subset U_{j,t}$ basiert, kann folgendermassen beschrieben werden

$$\hat{M}_{j,t} = N_{j,t}\nu_{j,t} = \frac{N_{j,t}}{n_{j,t}} \sum_{i \in V_j} p_{i,j,t}$$

Die entsprechenden Vektoren der Hilfsvariablen und der Parameter sind somit:
$\mathbf{x}'_{j,t} = (1, p_{i,j,t}, n_{j,t}, N_{j,t}) \quad , \forall j$
$\theta_j = (0, 1, 1, 1)$

5.1.2 Selektions Bias

Da die Daten nicht anhand einer Stichprobe gezogen wurden entsteht ein Bias, der so genannte Selektions Bias. *Heckman* [14] hat ein Modell entwickelt, welches erlaubt, dem Selektions Bias Rechnung zu tragen (siehe auch *Greene* [12]). Dieses Modell geht von einem einfachen bivariaten Regressionsmodell aus:

$$\begin{array}{rcll} Y_{1i} & = & X_{1i}\beta_1 + \varepsilon_{1i} & \\ Y_{2i} & = & X_{2i}\beta_2 + \varepsilon_{2i} & (i = 1, ..., I) \end{array} \tag{104}$$

wobei $E(\varepsilon_{ki}) = 0, \qquad E(\varepsilon_{ki}\varepsilon_{lr}) = \sigma_{kl}, \quad i = r,$
$= 0, \quad i \neq r.$
Zudem wird davon ausgegangen, dass die Daten von Y_{1i} nur zur Verfügung stehen, wenn eine Bedingung in Abhängigkeit von Y_{2i} erfüllt ist. Das heisst, dass Y_{1i} und Y_{2i} nicht unabhängig sind. Die Variable Y_{2i} muss für die ganze Population berechnet werden können, um das bivariate Regressionsmodell schätzen zu können. Oft wird in dieser Methode Y_{2i} dazu verwendet, um das Vorhandensein der Daten in der ersten Gleichung anzugeben. In unserem Falle wäre somit

$$Y_{2i} = \begin{cases} 1 & \text{wenn } i \in \bigcup V_{j,t} \\ 0 & \text{wenn } i \notin \bigcup V_{j,t} \end{cases}$$

Anhand dieses Ansatzes wird die Einschlusswahrscheinlichkeit durch die Indikatorvariable Y_{2i} nachträglich in das Modell miteinbezogen.
Diese Methode ist nur anwendbar, wenn die Hilfsvariable X_{2i} für die gesamte Population bekannt ist. Die Volkszählung ist zurzeit die einzige Datengrundlage, welche sämtliche Lohnempfänger aufweist. Somit müssten die Verunfallten aus den Daten der SSUV in den Daten der Volkszählung identifiziert werden können. Dies ist aus Gründen des Datenschutz nicht möglich. Da diese Methode nicht anwendbar ist, werden wir darauf verzichten, sie weiter zu beschreiben.

5.2 Kalibrierungs-Methode

Wir werden in diesem Abschnitt die SLI anhand der Lohnmasse aus der LSE kalibrieren.

$$\ddot{M}_{j,t} = \frac{\hat{M}_{j,t}}{\hat{M}_{j,t_0}} \check{M}_{j,t_0} \tag{105}$$

wobei $\check{M}_{j,t_0}$ der Lohnmassenschätzer der Gruppe $U_{j,t}$ für die Referenzperiode aus der LSE ist.
$\hat{M}_{j,t}$ könnte anhand der SSUV-Daten geschätzt werden, siehe (38). Für t_0 ist $\ddot{M}_{j,t_0} = \check{M}_{j,t_0}$.
Der Lohnmassenindex pro Gruppe ist in diesem Falle

$$\ddot{I}^{M}_{j,t,t_0} = \frac{\ddot{M}_{j,t}}{\ddot{M}_{j,t_0}} \frac{\ddot{M}_{j,t}}{\check{M}_{j,t_0}} = \frac{\hat{M}_{j,t}}{\hat{M}_{j,t_0}} \tag{106}$$

Für die gesamte Population ist der Lohnmassenschätzer

$$\ddot{M}_{t} = \frac{\hat{M}_{t}}{\hat{M}_{t_0}} \check{M}_{t_0} \tag{107}$$

wobei $\check{M}_{t_0}$ der Lohnmassenschätzer über alle Gruppen für die Referenzperiode aus der LSE ist.
Die Gewichtung (4) bezieht sich in diesem Modell ausschliesslich auf die Lohnmasse der LSE in der Referenzperiode:

$$\ddot{w}_{j,t_0} = \frac{\ddot{M}_{j,t_0}}{\sum\limits_{j} \ddot{M}_{j,t_0}} = \frac{\check{M}_{j,t_0}}{\check{M}_{t_0}} \qquad \forall j \tag{108}$$

Der Lohnmassenindex über alle Gruppen ist

$$\begin{aligned} \ddot{I}^{M}_{t,t_0} &= \sum_{j\in N_v} \ddot{w}_{j,t_0} \ddot{I}^{M}_{j,t,t_0} = \sum_{j\in N_v} \frac{\check{M}_{j,t_0}}{\check{M}_{t_0}} \frac{\hat{M}_{j,t}}{\hat{M}_{j,t_0}} \\ &= \frac{1}{\check{M}_{t_0}} \sum_{j\in N_v} \frac{\hat{M}_{j,t}}{\hat{M}_{j,t_0}} \check{M}_{j,t_0} = \frac{1}{\check{M}_{t_0}} \sum_{j\in N_v} \ddot{M}_{j,t} \end{aligned} \tag{109}$$

Für die Schätzung im Untersuchungsbereich E_d passen wir die Gewichtung (108) folgendermassen an

$$\ddot{w}_{j,d,t_0} = \frac{\ddot{M}_{j,t_0}}{\sum\limits_{j\in J_d} \ddot{M}_{j,t_0}} = \frac{\check{M}_{j,t_0}}{\check{M}_{d,t_0}} \qquad \forall j \in J_d \tag{110}$$

Der Schätzer des Lohnmassenindex im Untersuchungsbereich E_d ist somit

$$\begin{aligned} \ddot{I}^{M}_{d,t,t_0} &= \sum_{j\in J_d} \ddot{w}_{j,d,t_0} \ddot{I}^{M}_{j,t,t_0} \\ &= \sum_{j\in J_d} \frac{\check{M}_{j,t_0}}{\check{M}_{d,t_0}} \frac{\hat{M}_{j,t}}{\hat{M}_{j,t_0}} \\ &= \frac{1}{\check{M}_{d,t_0}} \sum_{j\in J_d} \ddot{M}_{j,t} \end{aligned} \tag{111}$$

Der Schätzer für den Durchschnittslohnindex pro Gruppe entspricht der derzeit verwendeten Methode:

$$\ddot{I}^{D}_{j,t,t_0} = \frac{\ddot{M}_{j,t}\hat{N}_{j,t_0}}{\ddot{M}_{j,t_0}\hat{N}_{j,t}} = \frac{\hat{M}_{j,t}\hat{N}_{j,t_0}}{\hat{M}_{j,t_0}\hat{N}_{j,t}} = \frac{\hat{\mu}_{j,t}}{\hat{\mu}_{j,t_0}} \tag{112}$$

Wenn die kalibrierte Lohnmasse und die Gewichtung (108) für den Schätzer des Durchschnittslohnindex über alle Gruppen angewandt wird, heisst das

$$\ddot{I}^{D}_{t,t_0} = \sum_{j\in N_v} \ddot{w}_{j,t_0} \ddot{I}^{D}_{j,t,t_0} = \sum_{j\in N_v} \frac{\check{M}_{j,t_0}\hat{\mu}_{j,t}}{\check{M}_{t_0}\hat{\mu}_{j,t_0}} \tag{113}$$

Für den Schätzer des Durchschnittslohnindex im Untersuchungsbereich E_d verwenden wir die Gewichtung (110):

$$\ddot{I}^{D}_{d,t,t_0} = \sum_{j\in J_d} \ddot{w}_{j,d,t_0} \ddot{I}^{D}_{j,t,t_0} = \sum_{j\in J_d} \frac{\check{M}_{j,t_0}\hat{\mu}_{j,t}}{\check{M}_{d,t_0}\hat{\mu}_{j,t_0}} \tag{114}$$

Keiner der in diesem Abschnitt aufgeführten Indices ist ein ein *Laspeyres*-Preisindex, wie wir ihn im Abschnitt 3.1 beschrieben haben, da die Gewichte kalibriert wurden.Die Kalibrierung muss bei dieser Methode nur für die Referenzperiode angewandt werden. Die Populationsgrösse, respektive die Einschlusswahrscheinlichkeit muss aber dennoch vierteljährlich geschätzt werden, um den Lohnmassenschätzer $\hat{M}_{j,t}$ schätzen zu können. Der Vorteil dieser Methode gegenüber der bisherigen Methode ist, dass sie erlaubt, die Schätzer der Lohnindices mit der Lohnmasse der LSE zu verknüpfen und somit kommt eine Harmonisierung der SLI mit der LSE zustande. Dies setzt jedoch eine Harmonisierung der Population der LSE und der Population der SLI voraus.

6 Diskussion der Schätz- und Modell-Methoden

Anhand der Stichproben-Methode konnten wir unter gewissen Hypothesen die beiden Lohnindices nach der Vorgabe von LOHN schätzen. Ebenso konnten Vertrauensintervalle geschätzt werden. Dabei spielt die Unfallhäufigkeit eine zentrale Rolle. Die Lohnstrukturen können bei dieser Methode bei der Bildung der homogenen Gruppen eine Rolle spielen (siehe entsprechender Abschnitt weiter unten). Die Regressions-Methode versucht zuerst die Charakteristiken des Lohns zu modellisieren und danach wird anhand der Selektions Bias Methode die Unfallhäufigkeit miteinbezogen. Wir haben jedoch Zweifel daran, ob die Regressions-Methode realisiert werden kann. Denn sie basiert auf einer Identifizierung der Verunfallten aus den SSUV Daten in der Volkszählung oder einer Vollerhebung aller Lohnempfänger. Abgesehen von den Problemen die Lohnempfänger aus den Daten der SSUV in der Volkszählung zu identifizieren muss bedenkt werden, dass diese nur alle 10 Jahre durchgeführt wird. Die entsprechenden Daten müssten jedoch vierteljährlich für den Lohnmassenindex und jährlich für den Durchschnittslohnindex vorhanden sein. Die Kalibrierungs-Methode führt zu einer Harmonisierung der LSE mit den SLI. Die Harmonisierung basiert vor allem auf der Gewichtung der Lohnindices der einzelnen Gruppen. Voraussetzung dafür ist jedoch, dass die Population der SLI in der LSE bestimmt werden kann, bis zu den einzelnen Lohngruppen der SLI. Der Verlust der *Laspeyres*-Preisindex Charakteristik durch die Kalibrierung, ist mit Blick auf die Lohnindices der EU-Länder nicht wünschenswert.

7 Homogene Gruppen

Die homogenen Gruppen (HG) werden von Variablen gebildet, welche die Charakteristiken der Zielvariablen weitgehend erklären können. Die Homogenität der

Gruppen wird durch Niveaus der Zielvariablen zum Zeitpunkt t_0 bestimmt und sollte über die folgenden Perioden bestehen bleiben. Nachfolgend werden wir die Gründe für die Homogenität erläutern. Daraufhin werden wir zuerst die bisherige Methode zur Bildung von homogenen Gruppen vorstellen und dann einen statistischen Ansatz vorstellen, die so genannten Klassifikationsbäume (*Classification trees*), die verwendet werden um homogene Gruppen zu bilden. Die Anwendung der Klassifikationsbäume wird nach Zielvariable einzeln behandelt.

7.1 Warum Homogenität?

Aus drei Gründen müssen die Gruppen homogen sein.

1. Der Schätzer des Durchschnittslohns der Verunfallten ($m_{j,t}$) muss ein guter Schätzer des Durchschnittslohns der Population ($\mu_{j,t}$) sein. Das heisst, $m_{j,t} \approx \mu_{j,t}, \forall j, t$, wobei die Varianz des Schätzers möglichst gering sein soll.

2. Die Population der HG j muss zum Zeitpunkt $t + 1$ vergleichbar sein mit der Population derselben HG zum Zeitpunkt t. Das heisst es muss sich um das "gleiche Produkt"handeln.

3. Damit die Aussagekraft, respektive die Akzeptanz, der Lohnindices verstärkt werden kann, müssen erkennbare Beschäftigungstypen gebildet werden können, welche als solche homogen sind.

7.2 Methode LOHN

Die bisherige Methode hatte zum Ziel, die Lohnvariabilität innerhalb der HG zu minimieren. Die homogenen Gruppen wurden anhand der 6 Variablen der Tabelle 3 gebildet. Das Vorgehen kann in 5 Schritte aufgeteilt werden:

1. Aus der Betriebszählung 1991 (BZ91) wird die Anzahl Vollzeitbeschäftigter Lohnempfänger nach Geschlecht und ASWZ bestimmt.

2. Aus der Eidgenössischen Volkszählung 1990 (VZ90) wird anhand eines Umrechnugsschlüssels der Arbeitsbereich nach Geschlecht und ASWZ aus dem Beruf ermittelt.

3. Aus der VZ90 wird anhand eines Umrechnugsschlüssels die Qualifikation nach Geschlecht und ASWZ aus den Sozio-professionellen Kategorien ermittelt.

4. Aus der Lohn- und Gehaltserhebung vom Oktober 1991 (LOK91) wird die Lohnart nach Geschlecht und ASWZ bestimmt.

5. Aus der SSUV91 wird die Unfallart nach Geschlecht und ASWZ ermittelt.

7.2.1 Definitionen

Wir vernachlässigen den Zeitindex im folgenden Abschnitt, da die Bestimmung der homogenen Gruppen nur zur Referenzperiode t_0 gemacht wird. Dies, obwohl die Datenquellen zu verschiedenen Zeitpunkten erhoben wurden. Es sei f die Häufigkeitsfunktion, welche die Anzahl Ereignisse pro Modalität angibt. Gleichermassen definieren wir die folgenden relativen Häufigkeitsfunktionen
f_A : des Arbeitsbereichs
f_Q : der Qualifikation
f_L : der Lohnart
f_U : der Unfallart.
Die jeweiligen Schätzer werden mit einem Hut ("ˆ") bezeichnet, z.B. $\hat{f}_L =$ Schätzer von f_L.
Sämtliche Charakteristiken können durch einen Vektor X ausgedrückt werden: $X = (X_N = x_N, X_G = x_G, X_A = x_A, X_Q = x_Q, X_L = x_L, X_U = x_U)$.
Die Schätzung der Anzahl Lohnempfänger in den Gruppen U_j ist durch die folgende Häufigkeitsfunktion definiert:

$$\begin{aligned} \hat{f}(X) &= \hat{f}_{N,G}(X_N, X_G) \cdot f_A(X_A|X_N, X_G) \cdot \\ &\quad f_Q(X_Q|X_N, X_G) \cdot \hat{f}_L(X_L|X_N, X_G) \cdot f_U(X_U|X_N, X_G) \end{aligned} \quad (115)$$

wobei $\hat{f}_{N,G}(X_N, X_G)$ die Schätzung aus der BZ91 der gemeinsamen Häufigkeitsfunktion von X_N und X_G ist.

Tabelle 3: Variablen für die Bildung der homogenen Gruppen. Für einzelne Variablen wurde ein Umrechnungsschlüssel angewandt, um die gewünschte Variable zu erhalten. Die entsprechenden Datenquellen wurden mit einem * versehen.

Variablenname	Variable	# Mod.	Modalitäten	Datenquelle
ASWZ	X_N	28	1..93	BZ91
Geschlecht	X_G	2	Mann/Frau	BZ91
Arbeitsbereich	X_A	3	Betrieb/Büro/Verkauf	VZ90*
Qualifikation	X_Q	2	Qualifiziert/nicht oder semi qualif.	VZ90*
Lohnart	X_L	2	Stundenlohn/Monatslohn	LOK91
Unfallart	X_U	2	Berufs-/Nicht Berufsunfall	SSUV91

7.2.2 Verifizierung der Homogenitäts-Hypothese

Wir haben zuerst die Lohnverteilung innerhalb der HG untersucht. Dabei haben wir uns auf die Daten des Jahres 1998 gestützt. Wir werden im folgenden die ersten beiden Gründe, weshalb wir HG brauchen, aus 7.1 aufgreifen.

Ist $m_{j,t} \approx \mu_{j,t}$?
Obwohl die Population der LSE sowie die Definition des Lohns nicht dieselben sind wie bei der Schätzung des Durchschnittslohnindex, haben wir einen Vergleich der Entwicklung des Durchschnittslohnindex angestellt. Wir haben anhand der LSE einen Durchschnittslohnindex geschätzt, wobei die Referenzperiode 1996 war. Wir werden im folgenden keine Resultate zeigen, da LOHN sich zurzeit mit einer solchen Vergleichsstudie beschäftigt. Dabei sollen die unterschiedlichen Definitionen so gut wie möglich in die Schätzung miteinbezogen werden.
Wir haben in unserem Vergleich jedoch derart grosse Unterschiede festgestellt, dass wir daran zweifeln, dass dies einzig eine Frage der Definition der Population und des Lohns ist. Auf die Streuung des Lohns gehen wir im Abschnitt 7.5 ein.

Vergleichbarkeit der Population
Wenn ein grosses Flugunternehmen die jungen Mitarbeiter mit tiefen Löhnen entlässt, steigt der Durchschnittslohn in bestimmten HG und somit der Durchschnittslohnindex. Da die Altersstruktur nicht in die Bildung der HG miteinbezogen wurde, ist die strikte Homogenität sicherlich nicht gewährleistet.

7.2.3 Diskussion der Methode LOHN

Die bisherige Methode geht implizit davon aus, dass der Arbeitsbereich, die Qualifikation, die Lohnart und die Unfallart unabhängig voneinander einen Einfluss auf den Lohn innerhalb der Gruppierungen nach ASWZ und Geschlecht haben. Ein Grund für diese Hypothese könnte der Mangel an Daten gewesen sein, welche die Kreuzung dieser Charakteristiken ermöglichen. Die Charakteristik Unfallart X_U muss noch für eine andere Charakteristik stehen, denn sonst macht die Schätzung der Anzahl Lohnempfänger für diese Gruppierung keinen Sinn, da die Daten der Verunfallten von der SSUV geliefert werden. Die Homogenität innerhalb der gebildeten Gruppen wird zusätzlich für die Daten der SSUV vorausgesetzt.
Bis anhin wurde die Schätzung der Anzahl Lohnempfänger in den HG nicht als Schätzproblem behandelt.

7.3 Klassifikationsbäume

Die Bildung der homogenen Gruppen hängt davon ab, welche Zielvariable untersucht wird. Die Methode ist unabhängig von der Zielvariablen jedoch die gleiche. Wir werden versuchen homogene Gruppen anhand von Klassifikationsbäumen zu bilden.
Diese Methode bildet rekursiv Knoten und Äste. Wir gehen davon aus, dass wir nur kategorische Hilfsvariablen für die Berechnung der Klassifikationsbäume zur Verfügung haben. Es sei:
$p =$ Anzahl Hilfsvariablen
$x_l =$ die l-te Hilfsvariable, $l \in \{1, ..., p\}$.
$x_{i,l} =$ der Wert der Hilfsvariablen l für den Verunfallten i.
$a_l =$ Anzahl Modalitäten der Hilfsvariablen l.
Somit kann der Wert der Hilfsvariablen l für den Verunfallten i als Vektor beschrieben werden:
$(x_{i,l,1}, ..., x_{i,l,a_l})$, wobei $x_{i,l,k} = 1$ für $k = x_{i,l}$ und $x_{i,l,k} = 0$ für $k \neq x_{i,l}$.
$\tau_{i,l} = (\tau_{l,1}, ..., \tau_{l,a_l}) =$ Wahrscheinlichkeitsvektor der Hilfsvariablen x_l, dass der Verunfallte i das jeweilige Merkmal besitzt. Somit ist $\sum_{k=1}^{a_l} \tau_{i,l,k} = 1$.
Bei der Berechnung der Klassifikationsbäume wird ein Schätzer $\hat{\tau}_l$ für τ_l anhand einer maximum-likelihood Methode geschätzt.
Als Devianzfunktion verwenden wir die Entropie

$$D(\tau_{i,l}; x_{i,l}) = -2 \sum_{k=1}^{a_l} x_{i,l,k} \log(\tau_{i,l,k}) \tag{116}$$

Anhand dieser Methode versuchen wir die Population so aufzuteilen, dass die Verteilung der Zielvariablen innerhalb der gebildeten Gruppen unabhängig von den Hilfsvariablen ist: $f(P_{j,t}|\mathbf{x})$ unabhängig von $\mathbf{x}$.
Es werden Knoten gebildet, in welchen die Population in zwei Gruppen aufgeteilt wird. In den Knoten ist die Devianz $D(\hat{\tau}; \mathbf{x}) = \sum_i \sum_l D(\hat{\tau}_l; x_{i,l})$. Bei der Aufteilung in zwei Gruppen werden zwei Wahrscheinlichkeitsvektoren $\hat{\tau}_{R,l}$ und $\hat{\tau}_{L,l}$ geschätzt. Hiermit wird eine Teilungsdevianz berechnet:

$$D(\hat{\tau}_L, \hat{\tau}_R; \mathbf{x}) = \sum_L \sum_l D(\hat{\tau}_L; x_{i,l}) + \sum_R \sum_l D(\hat{\tau}_R; x_{i,l})$$

Eine Teilung kann anhand des Devianzunterschieds zwischen dem ungeteilten Knoten und dem geteilten Knoten charakterisiert werden:

$$\Delta D = D(\hat{\tau}; \mathbf{x}) - D(\hat{\tau}_L, \hat{\tau}_R; \mathbf{x})$$

Die Teilung, welche den Devianzunterschied maximiert, wird gewählt, um dann in den beiden neu entstandenen Knoten weitere Teilungen in Betracht zu ziehen. Hiermit wird der Gewinn an Homogenität der Teilung maximiert. Die Teilung wird gestoppt, wenn gewisse Kriterien erfüllt sind. Diese müssen vom Benutzer definiert werden. Standardmässig können drei Kriterien Einfluss auf die Beendung der Teilung haben:
- die Anzahl Knoten
- die Anzahl Elemente in den Knoten
- eine untere Grenze für den Devianzunterschied ΔD.

Diese Methode bildet somit homogene Gruppen bezüglich der Zielvariablen, unter Berücksichtigung der Hilfsvariablen. Somit haben wir eine statistische Homogenität, aber noch keine allgemein gültige Definition für die Homogenität.

7.4 HG bei der Schätz-Methode

Bei de Schätzmethode ist die Einschlusswahrscheinlichkeit die Zielvariable. Wir müssen versuchen die Population so aufzuteilen, dass in den HG die Einschlusswahrscheinlichkeit für alle Verunfallten, welche dieser Gruppe angehören, die gleiche ist:

$$\pi_{i,j,t} = \pi_{k,j,t} \quad \forall i,k \in V_{j,t} \tag{117}$$

Die homogenen Gruppen könne nur über eine längere Zeitspanne beibehalten werden, wenn das Kriterium (117) während dieser Zeitspanne erfüllt ist.
Die Einschlusswahrscheinlichkeit ist zurzeit nicht verfügbar. Deshalb ist es nicht möglich homogene Gruppen anhand dieses Kriteriums zu bilden. Bei der bisherigen Methode wird die Einschlusswahrscheinlichkeit implizit für alle Verunfallten der gleichen HG gleich gesetzt. Die Modellisierung der Einschlusswahrscheinlichkeit in 4.2.2 hat jedoch gezeigt, dass diese Hypothese eingehenden Tests bedarf.

7.5 HG bei der Modell-Methode

Die HG bei der Modell-Methode sollten ein homogenes Lohnniveau innerhalb der HG aufweisen. Das Lohnniveau sollte nicht nur während der Referenzperiode homogen sein, sondern auch über die folgenden Perioden. Während dieser Perioden könnten somit dieselben homogenen Gruppen beibehalten werden. Wie in [3] aufgezeigt wurde, hat eine Inhomogenität der Löhne einer Gruppe Auswirkungen auf den Durchschnittslohnindex. Wie gross diese Auswirkungen sind, wurde nicht untersucht, da eine Definition der Homogenität nicht gegeben wurde (werden kann). In der bisherigen Methode wird davon ausgegangen, dass diese Homogenität der Gruppen gegeben ist.

7.5.1 Daten und Variablen

Zurzeit stehen keine Daten zur Verfügung, welche es erlauben würden, die HG mit denselben Variablen zu bilden, wie dies in der bisherigen Methode gemacht wurde. In unseren Tests haben wir die Daten der SSUV von 1998 verwendet. Für das Klassifikationsbäume-Modell haben wir die SSUV Daten von 1999 verwendet.
Die homogenen Gruppen werden anhand der Variablen in Tabelle 10 gebildet. Die untersuchten Wirtschaftszweige sind in Tabelle 11 im Anhang aufgelistet.

7.5.2 Resultate

Wir haben die Homogenität in den bestehenden HG untersucht. Dabei haben wir den Durchschnittslohn und das entsprechende 95% Vertrauensintervall jeder Gruppe berechnet. Das Vertrauensintervall des monatlichen Durchschnittslohns hat folgende Form:

$$L^{D}_{j,98}(\nu_{j,98}) = \left[\nu_{j,98} \pm \frac{1.96}{\sqrt{(n_{j,98})}} cv(\nu_{j,98})\nu_{j,98} \right] \tag{118}$$

Wir haben die Grösse des Vertrauensintervalls für jede Gruppe in der Abbildung 3 aufgezeichnet. In der Abbildung 4 haben wir nur die HG aufgezeichnet, bei welchen die Grösse des Vertrauensintervalls kleiner als 1500 Fr. beträgt. Aus den beiden Grafiken ist ersichtlich, dass in einigen HG das Vertrauensintervall ziemlich gross ist. Wir zweifeln deshalb an der Hypothese der Homogenität des Durchschnittslohns in den bestehenden HG, obwohl, wie bereits gesagt, keine Definition der Homogenität der Löhne besteht.

Anhand der Klassifikationsbäume-Methode haben wir untersucht, welche Variablen für die Homogenität der Löhne vor allem verantwortlich sind. Wir haben die Variablen verwendet, welche zur Zeit die Bildung der homogenen Gruppen charakterisieren. Dabei haben wir den optimalen Klassifikationsbaum mit der Funktion 'tree' in S-Plus berechnet. Die entsprechenden Abbildungen sind im Anhang B. Aus der Abbildung 21 ist ersichtlich, dass die Knotenbildung anhand der Variablen Qualifikation und ASWZ eine grosse Devianzreduktion zur Folge haben. Die restlichen Variablen spielen eine kleinere Rolle. Wir haben diesen Klassifikationsbaum seines Umfangs wegen nicht im Textformat in den Bericht eingefügt. Daraus wäre jedoch ersichtlich gewesen, dass die Devianz in den meisten Ästen des optimalen Baumes immer noch beträchtlich ist. Dies kann als weiterer Hinweis aufgefasst werden, dass die derzeitige Bildung der homogenen Gruppen nicht zu

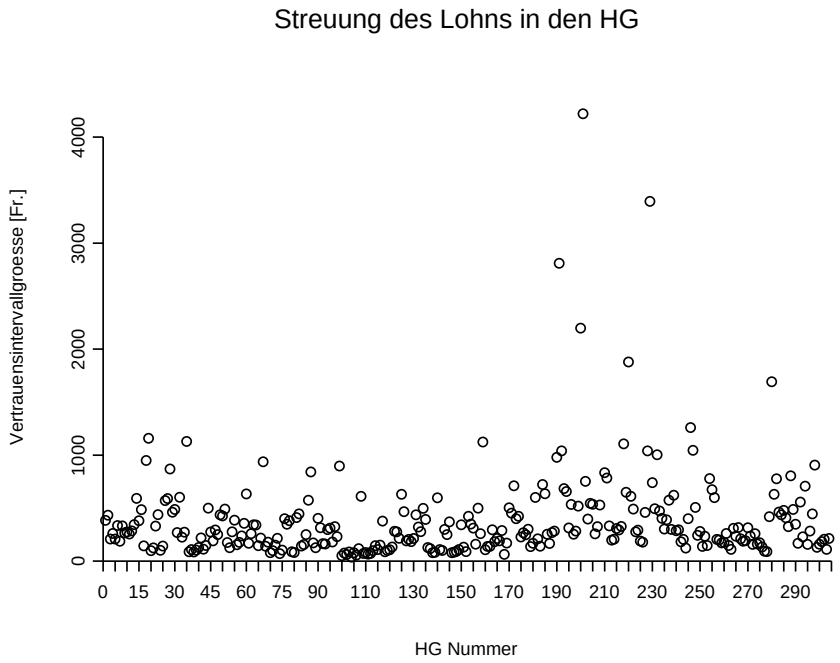


Abbildung 3: Grösse des Vertrauensintervalls der Schätzung des Durchschnittslohns in den 304 HG.

homogenen Lohngruppen führt. In zwei weiteren Modellen haben wir die Löhne in Lohnklassen zu Fr. 500.- respektive Fr. 1000.- eingeteilt. In der Abbildung 22 haben wir den Baum abgebildet, bei welchem wir Fr. 1000.- Lohnklassen gebildet haben, da dieser eine kleinere Devianz hat und der Anteil der Falschklassierungen geringer ist im Unterschied zum Baum mit Fr. 500.- Lohnklassen. Die Knotenbildung anhand der Variablen Qualifikation, ASWZ und des Geschlechts führt zur grössten Devianzreduktion. Im Vergleich mit der Abbildung 21 wurden die Knoten zumindest im oberen Teil des Baums mit übereinstimmenden oder ähnlichen Kriterien gebildet. Die Devianz ist in den meisten Ästen immer noch beträchtlich, wobei sie im Vergleich mit 21 rund um einen Faktor 10000 kleiner ist. Mit einem Anteil von 58.87% an Falschklassierungen genügt dieses Modell jedoch nicht.
Aufgrund dieser Resultate scheint es uns sehr schwierig zu sein, anhand der verwendeten Variablen homogene Lohngruppen zu bilden.

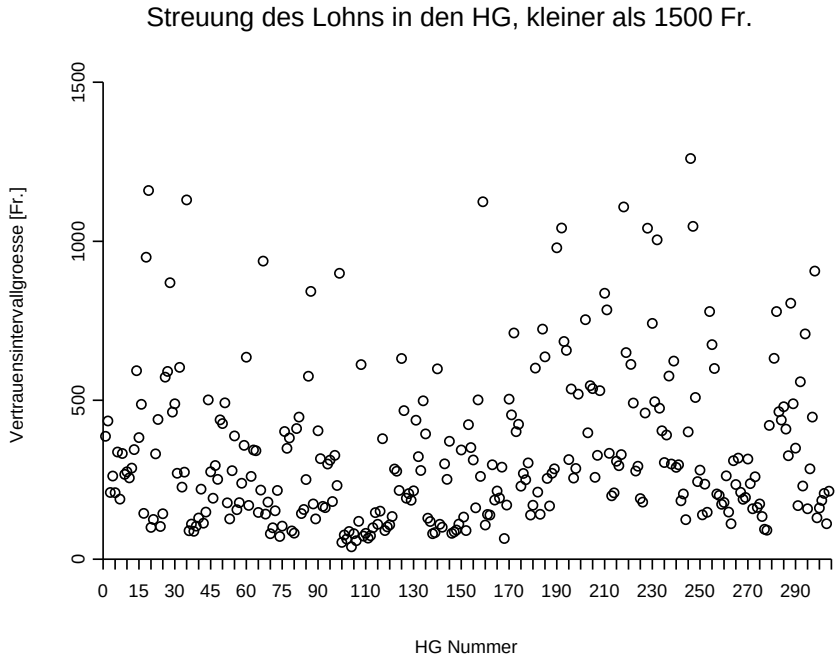


Abbildung 4: Grösse des Vertrauensintervalls der Schätzung des Durchschnittslohns in den HG, deren Vertrauensintervall kleiner als 1500 Fr. ist.

8 Gewählte Methode

8.1 Entscheid LOHN

Aus pragmatischen Gründen und da keine zusätzlichen Informationen zur Verfügung gestellt werden können, hat LOHN entschieden, die folgenden Hypothesen zu akzeptieren:

1. Die Daten der SSUV können als einfache Stichprobe der Population angesehen werden, wobei die Einschlusswahrscheinlichkeit (Unfallwahrscheinlichkeit) innerhalb der homogenen Gruppen die gleiche ist.

2. Die Schätzung der Anzahl Lohnempfänger für die Referenzperiode ist erwartungstreu und beinhaltet weder Bias noch eine Varianz.

3. Die Homogenität der Löhne in den HG ist gewährleistet.

4. Die homogenen Gruppen weisen immer genügend Fälle auf, damit die Lohnindices berechnet werden können.

Aus 1. und 2. folgt, dass vorausgesetzt wird, dass der Durchschnittslohn in den HG der wahre Durchschnittslohn in den entsprechenden Lohngruppen ist

$$\tilde{\nu}_{j,t_0} = \tilde{\mu}_{j,t_0} = \mu_{j,t_0}, \forall j.$$

Die Bildung der homogenen Gruppen wird auf weniger Variablen als bisher basieren. Die Schätzung der Anzahl Lohnempfänger wird jedoch auf einem detaillierteren Niveau anhand der BESTA und der SAKE durchgeführt. Zudem wurde in der SAKE02 die Frage bezüglich der Verunfallung wieder eingeführt, womit LOHN auf einem aggregierten Niveau die Unfallwahrscheinlichkeit mit den SSUV Daten vergleichen kann. Die Anzahl Lohnempfänger im 1. Sektor müssen aufgrund der Daten der AGR-BZ geschätzt werden. Diese Erhebung kann keine vierteljährlichen Schätzer liefern. Somit kann der vierteljährliche Lohnmassenindex nur für den 2. und 3. Sektor, nicht aber für den 1. Sektor vierteljährlich geschätzt werden.

8.2 Vorteile und ungelöste Probleme dieser Methode

Die Daten der Lohnempfänger werden von der SSUV geliefert und wir können davon ausgehen, dass wir keine Antwortausfälle für die Lohnkomponenten haben. Dies im Gegensatz zu einer Erhebung bei den Lohnempfängern, bei welcher wir mit einem grossen Anteil Antwortausfälle für diese Variable rechnen müssen.

8.2.1 Strukturwandel in der Population

Ein Strukturwandel in der Population kann Auswirkungen auf die Daten der SSUV haben (z.B. Populationsverschiebungen zwischen Branchen). Anhand einer kontrollierten Einschlusswahrscheinlichkeit könnte diesem Problem Rechnung getragen werden.
Für die Schätzung des Lohnmassenindex kann die Population vierteljährlich auf dem Niveau NOGA $\times$ Geschlecht $\times$ Beschäftigungsgrad anhand der BESTA geschätzt werden. Eine Veröffentlichung von Resultaten auf einem feineren Niveau ist nicht ratsam, da keine entsprechende Datengrundlage vorhanden ist.

8.2.2 Homogene Gruppen und deren Gliederung

Aus der Tabelle 4 geht hervor, dass nur eine feine Gliederung der HG und eine genügende Homogenität zu einer guten Schätzung für die Lohnindices führen können. Hierfür müsste aber die Einschlusswahrscheinlichkeit auf einem sehr feinem Niveau geschätzt werden, was nicht möglich ist.

Tabelle 4: Gliederung und Qualität von homogenen Gruppen sowie deren Auswirkung auf die Schätzung der Lohnindices. 'Index i.o.' bedeutet, dass die Qualität der Lohnindices gut ist. 'nicht' wurde mit 'n.' abgekürzt.

	HG grob gegliedert	HG fein gegliedert
Homogenität genügend	existiert n., Index i.o.	kann n. geschätzt werden
Homogenität ungenügend	Index nicht i.o.	bestehender I^D, nicht i.o.

8.3 Revidierte Methode LOHN

Für die Bildung der Lohngruppen werden die Variablen NOGA2[5], Beschäftigungsgrad, Altersklassen, Geschlecht und ausgeübter Beruf (ISCO[6]) verwendet.
Für den Sektor 1 muss ein spezielles Vorgehen erarbeitet werden. Für die Sektoren 2 und 3 kann anhand der BESTA die Anzahl Beschäftigte nach NOGA2, Beschäftigungsgrad und Geschlecht geschätzt werden. Anhand der SAKE wird die Anzahl Lohnempfänger nach den Variablen ausgeübter Beruf und Altersklassen geschätzt. Durch die Verbindung der BESTA und der SAKE wird somit die Population für die Lohngruppen geschätzt.
Bevor wir die revidierte Methode mathematisch beschreiben, wie wir dies in 7.2 für die bestehende Methode gemacht haben, werden wir in den folgenden Abschnitten die BESTA und die SAKE kurz vorstellen und deren Unterschiede aufzeigen.

8.4 BESTA und SAKE

BESTA
Die BESTA (Beschäftigten Statistik) ist eine vierteljährliche Beschäftigten Umfrage bei den Betrieben. Dabei wird die Anzahl Beschäftigte nach dem Geschlecht und dem Beschäftigungsgrad gefragt. Die Beschäftigten umfassen alle Angestellten, Inhaber, Lehrlinge, Aushilfen und im Aussendiest tätigen Personen ohne die Heimarbeiter. Diese Grundgesamtheit unterscheidet sich in zwei Punkten von der Definition der Grundgesamtheit der Lohnindices. Zum einen handelt es sich bei den Lohnindices um Lohnempfänger und zum anderen werden Inhaber, Lehrlinge und Aushilfen im Aussendienst sowie Heimarbeiter von der Grundgesamtheit der Lohnindices ausgeschlossen. Die Stichprobe der BESTA wurde für die Schätzung der Vollzeitäquivalenten auf der Ebene der Wirtschaftsabteilungen und der Wirtschaftssektoren der Grossregionen optimiert. Die Grundgesamtheit wurde anhand der NOGA (2. Position), den 7 Grossregionen der Schweiz und der Betriebsgrösse

[5] 28 zum Teil zusammengefasste NOGA-Codes (2 Positionen), siehe Anhang A Tabelle 12.
[6] ISCO=International Standard Classification of Occupations

gemessen in Vollzeitäquivalenten geschichtet.
Die Varianz der Schätzer in der BESTA wird mit einem separierten Quotientenschätzer geschätzt.

SAKE
Die SAKE (Schweizerische Arbeitskräfte Erhebung) ist eine jährlich durchgeführte Haushaltsbefragung, bei der alle Haushaltsmitglieder ab 15 Jahren miteinbezogen werden. Dabei wird eine Person zufällig aus dem Haushalt ausgewählt, welche während 5 Jahren alljährlich befragt wird. Sie gibt ebenso einzelne Angaben zu den anderen Arbeitskräften im Haushalt. Die Population der Erwerbstätigen der SAKE besteht aus sämtlichen Lohnempfängern, Heimarbeitern, Arbeitern in Marine, Konsulaten und Botschaften, sowie Unabhängigen, Inhabern, Lehrlingen und so genannten 'free lancer'. Die Population der Lohnindices beschränkt sich dagegen auf die Lohnempfänger.
Die SAKE wurde nicht für die Schätzung der Anzahl Lohnempfänger konzipiert. Somit ist der Schätzer der Anzahl Lohnempfänger aus der SAKE nicht optimal. Durch eine Kalibrierung anhand der BESTA kann die Anzahl Lohnempfänger vierteljährlich aus der SAKE geschätzt werden.
Die Varianz der Schätzer wird in der SAKE anhand einer *Jackknife*-Methode geschätzt (siehe [9]).

8.5 Unterschiedliche Definitionen der SLI, BESTA und der SAKE

In der Tabelle 5 haben wir die Unterschiede der Definition der Lohnempfänger der SSUV, der Erwerbstätigen der SAKE, sowie jene der Beschäftigten der BESTA und der AGR-BZ aufgezeigt. Die Unterschiede müssen in die Schätzungen der Population miteinbezogen werden. Probleme ergeben sich bei der Harmonisierung vor allem bei den Charakteristiken, die in der jeweiligen Erhebung nicht gesondert ausgewiesen werden können.
In der BESTA sind sämtliche Erwerbstätige, die in Betrieben, in welchen weniger als 20 Stunden pro Woche gearbeitet wird, ausgeschlossen. Aus der SAKE kann diese Anzahl aus der BESTA ausgeschlossener Erwerbstätiger nicht geschätzt werden, da die Gesamtarbeitszeit der Betriebe in der SAKE nicht erhoben wird. In der SAKE kann hingegen die Anzahl selbständig Erwerbender, die weniger als 20 Stunden pro Woche arbeiten, geschätzt werden. Wir gehen davon aus, dass der Unterschied zu den aus der BESTA ausgeschlossenen Erwerbstätigen gering ist. Die Grenzgänger können auf dem Niveau NOGA ab 2004 aus der BESTA geschätzt werden. Das ZAR (Zentrales Ausländer Register) kann ab 2004 keine Daten mehr über die Grenzgänger liefern.

Bei der folgenden Entwicklung der Schätzer sind wir davon ausgegangen, dass die Anzahl Grenzgänger auf dem Niveau NOGA x Geschlecht x Beschäftigungsgrad aus dem ZAR geschätzt werden könnten. Diese Annahme hat sich im Nachhinein als falsch herausgestellt. Ein entsprechendes Vorgehen muss noch entwickelt werden, um die Anzahl Grenzgänger auf diesem Niveau zu schätzen.

Tabelle 5: Unterschiede in den Charakteristiken der SLI, SAKE, BESTA und der AGR-BZ. 'N_(-20h)' sind die Lohnempfänger/Erwerbstätigen, die in Betrieben arbeiten, in welchen insgesamt weniger als 20h pro Woche gearbeitet wird. 'Grenzgänger' bezeichnen die Zupendler aus dem Ausland, d.h. Personen, die in keinem Haushalt der Schweiz erreichbar sind, aber in der Schweiz arbeiten. 'Mehrfachbeschäftigte' bezeichnet die Erwerbstätigen, welche mehr als eine Anstellung haben. Wenn diese Zelle angekreuzt ist, dann heisst das, dass diese Erwerbstätigen mehrfach gezählt werden. 'Mindest-Arbeitszeit' bezeichnet die Anzahl Arbeitsstunden die verrichtet werden muss, damit der Erwerbstätige/Beschäftigte zur Grundgesamtheit der Erhebung gezählt wird. In den SLI und der SAKE bezieht sich dieses Kriterium auf die gesamte Arbeitszeit, evtl. in mehreren Betrieben. In der BESTA und der AGR-BZ bezieht sich dieses Kriterium auf den Arbeitsplatz in einem Betrieb. 'X' bedeutet allgemein, dass die entsprechende Personen zur Grundgesamtheit gehören.

	Charakteristiken								
Erhebung	**Inhaber**	**N_(-20h)**	**Lehrlinge**	**Heim-arbeiter**	**Grenz-gänger**	**Mehrfach-beschäftigte**	**Mindest-Arbeitszeit**	**Alters-grenzen**	**1. Sektor**
SLI		X					1h	19-63/65	X
SAKE	X[a]	X[b]	X[a]	X[a]		[c]	1h	15-	X[a]
BESTA	X		X	X[d]	X[e]	X	6h	keine	
AGR-BZ	X	?	X	X	X	X	?	keine	X

[a] Kann geschätzt werden.

[b] Die Betriebsarbeitszeit wird in der SAKE nicht erhoben. Deshalb können nur die Einpersonenbetriebe, in welchen weniger als 20 Stunden gearbeitet wird, geschätzt werden.

[c] Aus der SAKE kann die Mehrfachbeschäftigung geschätzt werden.

[d] Ab 2. Quartal 2003.

[e] Kann auf Niveau NOGA ab 1. Quartal 2004 geschätzt werden.

8.6 Notationen und Schätzer der Populationsgrösse

Unser Ziel ist es einerseits die revidierte Methode mathematisch zu beschreiben und andererseits die Varianz der Schätzer der Population zu schätzen. Für die exakte Schätzung der Population müsste für jede Person in der SAKE, respektive für jeden Beschäftigten, der in der BESTA gezählt wird, die gemeinsame Einschlusswahrscheinlichkeit berechnet werden, dass die Person in der SAKE ist und in der BESTA gezählt wird. Dies setzt jedoch eine Identifikation aller Personen in der Vereinigung der Stichproben der BESTA und der SAKE voraus. Dies ist nicht möglich, da die Beschäftigten in der BESTA nicht identifiziert werden können. Wie aus dem Abschnitt 8.5 hervorgegangen ist, ist es nicht möglich, die BESTA vollständig mit der SAKE zu harmonisieren. Somit muss bei sämtlichen Schätzern, die wir im folgenden vorschlagen, von einem Bias ausgegangen werden.
In diesem Abschnitt werden wir die Schätzer der Populationsgrösse der SLI (Lohnindices) mathematisch beschreiben. Im Abschnitt 8.7 werden wir die Varianz der Schätzer schätzen.
Das Vorgehen für die Schätzung der Populationsgrösse der Referenzperiode ist das gleiche wie das Vorgehen für die vierteljährliche Schätzung der Populationsgrösse, die für den Lohnmassenindex verwendet wird. Deshalb werden wir diesen Aspekt im folgenden vernachlässigen. Zudem beschränken wir uns vorerst auf die Schätzung der Populationsgrösse im 2. und 3. Sektor. Das Vorgehen für den 1. Sektor ist zur Zeit noch unklar.
Wir werden im folgenden die Zeitkomponente t zur Vereinfachung nur bei vierteljährlich erhobenen oder geschätzten Grössen verwenden. Zusätzlich verwenden wir folgende Notationsregeln: hochgestellte Indices bezeichnen die Datenquelle mit dem Anfangsbuchstaben des Namens der Datenquelle. Die übrigen Indices bezeichnen Untergruppen oder die Zeitkomponente. Es sei
Ω^S : die Population der SAKE
s : die Stichprobe der SAKE
ω_k^s : Gewichtung der Person k in der SAKE. Diese beinhaltet die Einschlusswahrscheinlichkeit und die Nachschichtung (siehe [1]).
Für die Harmonisierung der Daten der BESTA und der SAKE definieren wir Untersuchungsbereiche:
E_d : die Untersuchungsbereiche, durch die NOGA2, das Geschlecht und den Beschäftigungsgrad definiert, aufgrund der Population der SLI.
$E_{l,d}^S$: Beschäftigte (Lohnempfänger) in $\Omega^S \cap E_d$ ohne Inhaber, Lehrlinge, Heimarbeiter und ohne selbständig Erwerbende ohne Mitarbeiter, die weniger als 20 Stunden pro Woche arbeiten, sowie beschäftigte Personen unter 19 Jahren, Unabhängige und Mitarbeitende im Familienbetrieb ohne Entlöhnung und beschäftigte Männer über 64 sowie beschäftigte Frauen über 62 Jahren in $E_d^S \cap \Omega^S$. Eingeschlossen sind hier Personen, die weniger als 6 Stunden pro Woche arbeiten. Diese

Definition der Beschäftigten in der SAKE entspricht der Definition der Lohnempfänger in den SLI.
$N^s_{l,d}$: Anzahl Lohnempfänger in $E^S_{l,d}$.
$\hat{N}^s_{l,d}$: Schätzer von $N^s_{l,d}$ aus der SAKE.
$E^S_{b,d}$: Beschäftigte in $\Omega^S \cap E_d$ ohne Personen, die weniger als 6 Stunden pro Woche arbeiten, sowie ohne selbständig Erwerbende ohne Mitarbeiter, die weniger als 20 Stunden pro Woche arbeiten. Diese Definition der Beschäftigten in der SAKE entspricht der Definition der Beschäftigten in der BESTA ab dem 3. Quartal 2003[7], ausser dass aus der SAKE die Anzahl beschäftigter Grenzgänger nicht geschätzt werden kann. $N^s_{b,d}$: Anzahl Erwerbstätige in $E^S_{b,d}$.
$\hat{N}^s_{b,d}$: Schätzer von $N^s_{b,d}$ aus der SAKE.

$N^s_{a,d}$: Anzahl Arbeitgeber in $E^S_{b,d}$.
$n^s_{a,k}$: Anzahl Arbeitgeber des Erwerbstätigen $k \in E^S_{e,d}$. Wir definieren $n^s_{a,k} = 0$, wenn $k \notin E^S_{b,d}$.
$\hat{N}^s_{a,d} = \sum_{k \in s \cap E^S_{e,d}} \omega^s_k n^s_{a,k}$: Schätzer der Anzahl Arbeitgeber in E^S_d aus der SAKE.
$\hat{N}^s_{p,d} = \sum_{k \in s \cap E_d} \omega^s_k$: aus der SAKE geschätzte Populationsgrösse von $E_d \cap \Omega^S$.
$\hat{N}_{l,d,t}$: vierteljährlicher Schätzer der Anzahl Lohnempfänger in $E^S_{l,d}$ anhand der BESTA und der SAKE.
$\hat{N}_{l,j,t}$: vierteljährlicher Schätzer der Anzahl Lohnempfänger in der Lohngruppe $U_j \cap E^S_{l,d}$ anhand der BESTA und der SAKE.
$\hat{N}^s_{l,j} = \sum_{k \in s \cap U_j \cap E^S_{l,d}} \omega^s_k n^s_{a,k}$: der Schätzer der Anzahl Lohnempfänger in der Lohngruppe $U_j \cap E^S_{l,d}$ anhand der SAKE.
Die bereits erwähnte Anzahl Lohnempfänger im Untersuchungsbereich $E^S_{l,d}$ kann anhand der SAKE folgendermassen geschätzt werden
$\hat{N}^s_{l,d} = \sum_{j \in J, U_j \cap E^S_{l,d} \neq \emptyset} \hat{N}^s_{l,j} = \sum_{k \in s \cap E^S_{l,d}} \omega^s_k n^s_{a,k}$.

Niveau Lohngruppe
$r_{l,d,j}$: Anteil der beschäftigten Lohnempfänger in der Lohngruppe U_j an der Anzahl beschäftigter Lohnempfänger im Untersuchungsbereich E_d. Zur Vereinfachung der Schreibweise schreiben wir nur $r_{l,d,j}$, da implizit klar ist, dass es sich um Lohnempfänger handelt.

$$r_{l,d,j} \doteq \begin{cases} \frac{N_{l,j}}{N_{l,d}} & \text{wenn } U_j \cap E^S_{l,d} \neq \emptyset \\ 0 & \text{andernfalls} \end{cases}$$

[7] Vor dem 3. Quartal 2003 müssen die Heimarbeiter ebenfalls ausgeschlossen werden.

$\hat{r}^s_{l,d,j}$: Schätzer von $r_{l,d,j}$ aus der SAKE.

$$\hat{r}^s_{l,d,j} = \begin{cases} \frac{\hat{N}^s_{l,j}}{\hat{N}^s_{l,d}} & \text{wenn } U_j \cap E^S_{l,d} \cap s \neq \emptyset \\ 0 & \text{andernfalls} \end{cases}$$

Dieser Schätzer hat einen Bias, B, der nicht berechnet werden kann. Dieser kann jedoch vernachlässigt werden, wenn der Biasquotient, BR, klein ist. Der Schätzer des Biasquotienten von $\hat{r}^s_{l,d,j}$ ist folgendermassen definiert

$$\widehat{BR}(\hat{r}^s_{l,d,j}) \doteq \frac{\hat{B}(\hat{r}^s_{l,d,j})}{\sqrt{\hat{V}ar(\hat{r}^s_{l,d,j})}}$$

Der Biasquotient ist nach oben begrenzt (siehe [15], pp. 176):

$$\left[\widehat{BR}(\hat{r}^s_{l,d,j})\right]^2 \leq \frac{\hat{V}ar(\hat{N}^s_{l,d})}{\left(\hat{N}^s_{l,d}\right)^2} = cv(\hat{N}^s_{l,d})^2$$

Der Bias kann vernachlässigt werden, wenn der Biasquotient kleiner als 0.1 ist, bis 0.5 kann er noch akzeptiert werden. Ist er grösser als 0.5, so ist der Schätzer derart verzerrt, dass er nicht mehr verwendet werden kann.
Aufgrund der oberen Grenze für den Biasquotienten können wir eine obere Grenze für den Bias im Quadrat schätzen.

$$\hat{B}(\hat{r}^s_{l,d,j})^2 = \hat{V}ar(\hat{r}^s_{l,d,j})\widehat{BR}(\hat{r}^s_{l,d,j})^2 \leq \hat{V}ar(\hat{r}^s_{l,d,j})cv(\hat{N}^s_{l,d})^2$$

Über das Vorzeichen des Bias können wir leider keine Aussage machen.

Harmonisierung der BESTA und der SAKE
Für die Harmonisierung der BESTA mit der SAKE müssen Unterschiede auf dem Niveau der Definitionen und des Stichprobenrahmens miteinbezogen werden wie sie im Abschnitt 8.5 aufgezeigt wurden. Es sei
$\hat{N}^B_{d,t}$: vierteljährlicher Schätzer der Anzahl Beschäftigten bezüglich der NOGA2, dem Geschlecht und dem Beschäftigungsgrad, d.h. in E^S_d, aus der BESTA.
N^z_d : die Anzahl erwerbstätige Saisonniers, Kurzaufenthalter und Grenzgänger nach NOGA2, Geschlecht und Beschäftigungsgrad aus dem ZAR[8] in E^S_d.

Die Harmonisierung der Anzahl Erwerbstätiger mit der Anzahl Beschäftigten wird durch die d Parameter $\alpha_{d,t}$ gewährleistet, welche durch die Gleichung (119)

[8]Die Daten des ZAR sind noch nicht zugesichert. Wir gehen jedoch davon aus, dass diese verwendet werden können.

geschätzt werden. Dabei wird der Schätzer der Anzahl Beschäftigter aus der BESTA dem Schätzer der Anzahl Beschäftigter aus der SAKE und dem ZAR, $\hat{N}^{s,z}_{b,d}$, gegenübergestellt:

$$\hat{N}^{B}_{d,t} = \hat{\alpha}_{d,t}\hat{N}^{s,z}_{b,d} \doteq \hat{\alpha}_{d,t}\left(\hat{N}^{s}_{b,d} + N^{z}_{d}\right)$$

$\hat{\alpha}_{d,t}$ wird dann folgendermassen geschätzt

$$\hat{\alpha}_{d,t} = \frac{\hat{N}^{B}_{d,t}}{\hat{N}^{s}_{b,d} + N^{z}_{d}} = \frac{\hat{N}^{B}_{d,t}}{\hat{N}^{s,z}_{b,d}} \tag{119}$$

Durch die Parameter $\hat{\alpha}_{d,t}$ wird der Entwicklung der Beschäftigung und den nicht schätzbaren, methodenbedingten Unterschieden[9] der Schätzer Rechnung getragen. Wir gehen dabei davon aus, dass diese Entwicklung für die beschäftigten Saisonniers, Kurzaufenthalter, Grenzgänger und die Anzahl Arbeitgeber pro Erwerbstätiger dieselbe ist, wie für die Beschäftigten Anzahl der BESTA.

Schätzung der Populationsgrösse
Für die Schätzung der Populationsgrösse haben wir zwei Möglichkeiten. Entweder wir basieren unsere Schätzung auf die Daten der BESTA und harmonisieren diese mit der SAKE und den SLI. Oder wir stützen uns auf die Daten der SAKE.

Wenn wir unsere Schätzung auf den Daten der SAKE abstützen, dann ist der Schätzer für die Anzahl Lohnempfänger in E_d

$$\hat{N}_{l,d,t} = \hat{\alpha}_{d,t}\hat{N}^{s}_{l,d} = \hat{N}^{B}_{d,t}\frac{\hat{N}^{s}_{l,d}}{\hat{N}^{s,z}_{b,d}} \tag{120}$$

Das heisst, wir gehen davon aus, dass sich die Anzahl Lohnempfänger zum Schätzer der Anzahl Lohnempfänger in der SAKE gleich verhält, wie der Schätzer der Anzahl Beschäftigter aus der BESTA zum Schätzer der Anzahl Beschäftigter aus der SAKE.

Folgendes Vorgehen wird gewählt, falls wir uns auf die Daten der BESTA abstützen, um die Populationsgrösse der Lohnempfänger zu schätzen. Der Schätzer der Anzahl Lohnempfänger in E_d wird vierteljährlich aufgrund des Schätzers der Anzahl Beschäftigte aus der BESTA und einem Korrektionsterm anhand der BESTA,

[9] c.f. [6].

der SAKE und dem ZAR geschätzt.

$$\begin{aligned}
\hat{N}_{l,d,t} &= \hat{N}^B_{d,t} - \hat{\alpha}_{d,t}\left(\hat{N}^{s,z}_{b,d} - \hat{N}^s_{l,d}\right) \\
&= \hat{N}^B_{d,t} - \frac{\hat{N}^B_{d,t}}{\hat{N}^s_{b,d} + N^z_d}\left(\hat{N}^{s,z}_{b,d} - \hat{N}^s_{l,d}\right) \\
&= \hat{N}^B_{d,t}\left(\frac{\hat{N}^s_{l,d}}{\hat{N}^s_{b,d} + N^z_d}\right) \\
&= \hat{N}^B_{d,t}\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}} \qquad (121)
\end{aligned}$$

Somit entspricht die Schätzung (121) der Schätzung (120), was zeigt, dass die beiden gewählten Ansätze gleich gesetzt werden können.
Die Schätzung der Anzahl Lohnempfänger in der Lohngruppe U_j kann folgendermassen ausgedrückt werden

$$\begin{aligned}
\hat{N}_{l,j,t} &= \hat{N}_{l,d,t} \cdot \hat{r}^s_{l,d,j} = \frac{\hat{N}_{d,t}}{\hat{N}^s_{l,d}}\hat{N}^s_{l,j} \qquad \forall U_j \subset E_d \\
&\overset{(120)}{=} \frac{\hat{\alpha}_{d,t}\hat{N}^s_{l,d}}{\hat{N}^s_{l,d}}\hat{N}^s_{l,j} = \hat{\alpha}_{d,t}\hat{N}^s_{l,j} \qquad \forall U_j \subset E_d \qquad (122)
\end{aligned}$$

8.7 Varianzschätzer der Populationsgrösse

Der Varianzschätzer von (120) wird unter der Annahme, dass die SAKE und die BESTA zwei unabhängige Erhebungen sind, mit einer *Taylor*-Linearisierung angenähert.

$$\begin{aligned}
\hat{V}ar(\hat{N}_{l,d,t}) &= \hat{V}ar(\hat{\alpha}_{d,t}\hat{N}^s_{l,d}) \\
&= \hat{V}ar\left(\frac{\hat{N}^B_{d,t}}{\hat{N}^{s,z}_{b,d}}\hat{N}^s_{l,d}\right) = \hat{V}ar\left(\hat{N}^B_{d,t}\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right) \\
&\approx \left(\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right)^2 \hat{V}ar(\hat{N}^B_{d,t}) + \left(\hat{N}^B_{d,t}\right)^2 \hat{V}ar\left(\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right) \\
&\quad + 2\hat{N}^B_{d,t}\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}} \underbrace{\hat{C}ov\left(\hat{N}^B_{d,t}, \frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right)}_{=0, \text{ da unabhängige Erhebungen}} \\
&\approx \left(\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right)^2 \hat{V}ar(\hat{N}^B_{d,t}) + \left(\hat{N}^B_{d,t}\right)^2 \hat{V}ar\left(\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right) \qquad (123)
\end{aligned}$$

Der erste Varianzschätzer auf der rechten Seite in (123) wird in der BESTA geschätzt. Der zweite Term wird mit einer weiteren *Taylor*-Linearisierung angenähert.

$$\begin{aligned}
\left(\hat{N}^B_{d,t}\right)^2 \hat{V}ar\left(\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right) &\approx \hat{\alpha}^2_{d,t}\Bigg[\hat{V}ar(\hat{N}^s_{l,d}) + \left(\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\right)^2 \hat{V}ar(\hat{N}^{s,z}_{b,d}) \\
&\quad -2\frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}\hat{C}ov(\hat{N}^s_{l,d}, \hat{N}^{s,z}_{b,d})\Bigg] \\
&= \hat{\alpha}^2_{d,t}\Big[\hat{V}ar(\hat{N}^s_{l,d}) + (\hat{\beta}^{s,z}_{l,b,d})^2\hat{V}ar(\hat{N}^{s,z}_{b,d}) \\
&\quad -2\hat{\beta}^{s,z}_{l,b,d}\hat{\rho}_s(\hat{N}^s_{l,d}, \hat{N}^{s,z}_{b,d})\sqrt{\hat{V}ar(\hat{N}^s_{l,d})\hat{V}ar(\hat{N}^{s,z}_{b,d})}\Big] \\
&= \hat{\alpha}^2_{d,t}\Big[\hat{V}ar(\hat{N}^s_{l,d}) + (\hat{\beta}^{s,z}_{l,b,d})^2\hat{V}ar(\hat{N}^s_{b,d}) \\
&\quad -2\hat{\beta}^{s,z}_{l,b,d}\hat{\rho}_s(\hat{N}^s_{l,d}, \hat{N}^s_{b,d})\sqrt{\hat{V}ar(\hat{N}^s_{l,d})\hat{V}ar(\hat{N}^s_{b,d})}\Big]
\end{aligned} \tag{124}$$

wobei $\hat{\beta}^{s,z}_{l,b,d} = \frac{\hat{N}^s_{l,d}}{\hat{N}^{s,z}_{b,d}}$ der Schätzer des Harmonisierungs-Parameters zwischen der Anzahl Lohnempfänger aus der SAKE und der Anzahl Beschäftigter aus der SAKE ist. $\hat{\rho}_{()}$ steht für die geschätzte Stichprobenkorrelation. Die Daten aus dem ZAR haben keinen Einfluss auf die Schätzer der Varianz, der Kovarianz und der Stichprobenkorrelation: $\hat{V}ar(\hat{N}^{s,z}_{b,d}) = \hat{V}ar(\hat{N}^s_{b,d})$, $\hat{C}ov(\hat{N}^s_{l,d}, \hat{N}^{s,z}_{b,d}) = \hat{C}ov(\hat{N}^s_{l,d}, \hat{N}^s_{b,d})$ und somit auch $\hat{\rho}_s(\hat{N}^s_{l,d}, \hat{N}^{s,z}_{b,d}) = \hat{\rho}_s(\hat{N}^s_{l,d}, \hat{N}^s_{b,d})$.
Die Varianzschätzer in (124) werden mit dem Macro 'jackanz' aus der SAKE geschätzt, siehe [9]. Den Kovarianzschätzer in (124) können wir mit einem um weniges abgeänderten Version des Makros 'jackanz' schätzen, das wir 'covjackanz' nennen, siehe Anhang E.
Das Vorgehen in (124) werden wir im Folgenden noch einmal antreffen. Deshalb

definieren wir einen Operator Ψ, der dieses Vorgehen beschreibt. Es sei

$$\begin{aligned}
\Psi(\hat{N}_X^s) &\doteq \left(\hat{N}_{d,t}^B\right)^2 \hat{V}ar\left(\frac{\hat{N}_X^s}{\hat{N}_{b,d}^{s,z}}\right) \\
&\doteq \left(\hat{N}_{d,t}^B\right)^2 \hat{V}ar(\hat{R}) \\
&= (\hat{\alpha}_{d,t})^2 \left[\hat{V}ar(\hat{N}_X^s) + \hat{R}^2\hat{V}ar(\hat{N}_{b,d}^{s,z}) - 2\hat{R}\hat{C}ov\left(\hat{N}_X^s, \hat{N}_{b,d}^{s,z}\right)\right] \\
&= (\hat{\alpha}_{d,t})^2 \Big[\hat{V}ar(\hat{N}_X^s) + \hat{R}^2\hat{V}ar(\hat{N}_{b,d}^s) \\
&\quad -2\hat{R}\hat{\rho}_s\left(\hat{N}_X^s, \hat{N}_{b,d}^s\right)\sqrt{\hat{V}ar(\hat{N}_X^s), \hat{V}ar(\hat{N}_{b,d}^s)}\Big]
\end{aligned} \tag{125}$$

Die Varianzschätzer in (125) werden mit dem Macro 'jackanz' aus der SAKE geschätzt. Der Kovarianzschätzer kann wie oben erwähnt, mit dem Macro 'covjackanz' geschätzt werden.
Mit dieser Notation können wir den Varianzschätzer (123) folgendermassen ausdrücken

$$\hat{V}ar(\hat{N}_{l,d,t}) \approx \left(\frac{\hat{N}_{l,d}^s}{\hat{N}_{b,d}^{s,z}}\right)^2 \hat{V}ar(\hat{N}_{d,t}^B) + \Psi(\hat{N}_{l,d}^s)$$

Somit haben wir die Varianz der Anzahl Lohnempfänger im Untersuchungsbereich E_d, mit der SAKE als Basis der Schätzung geschätzt. Es verbleibt die Schätzung der Varianz des Schätzers der Anzahl Lohnempfänger in den einzelnen Lohngruppen U_j. Aufgrund von (122) ist die Varianz von $\hat{N}_{j,t}$

$$\hat{V}ar(\hat{N}_{l,j,t}) = \hat{V}ar(\hat{\alpha}_{d,t}\hat{N}_{l,j}^s) = \hat{V}ar\left(\frac{\hat{N}_{d,t}^B}{\hat{N}_{b,d}^{s,z}}\hat{N}_{l,j}^s\right) = \hat{V}ar\left(\hat{N}_{d,t}^B\frac{\hat{N}_{l,j}^s}{\hat{N}_{b,d}^{s,z}}\right) \tag{126}$$

(126) entspricht in der Form (123). Somit können die Formeln (123) bis (125) direkt angepasst werden:

$$\begin{aligned}
\hat{V}ar(\hat{N}_{l,j,t}) &\approx \left(\frac{\hat{N}_{l,j}^s}{\hat{N}_{b,d}^{s,z}}\right)^2 \hat{V}ar(\hat{N}_{d,t}^B) + \left(\hat{N}_{d,t}^B\right)^2 \hat{V}ar\left(\frac{\hat{N}_{l,j}^s}{\hat{N}_{b,d}^{s,z}}\right) \\
&= \left(\frac{\hat{N}_{l,j}^s}{\hat{N}_{b,d}^{s,z}}\right)^2 \hat{V}ar(\hat{N}_{d,t}^B) + \Psi(\hat{N}_{l,j}^s)
\end{aligned} \tag{127}$$

Die Kovarianz in $\Psi(\hat{N}_{l,j}^s)$ kann mit dem Macro 'covjackanz' geschätzt werden und die Varianz-Terme mit dem Macro 'jackanz'.

9 Beispiel anhand der BESTA99 und der SAKE99

9.1 SAKE99 und BESTA99

Von den theoretischen 2016 Lohngruppen können in der SAKE99 deren 1013 gebildet werden (c.f. Tabelle 14). In 776 dieser Gruppen wies die SAKE99 weniger als 10 Antwortende auf, dies entspricht 76.6% aller Lohngruppen.
Bei 17 Antwortenden konnte die Anzahl Arbeitgeber nicht bestimmt werden, oder die Frage traf nicht zu. Gewichtet entspricht dies rund 0.37% der Arbeitnehmer (7'060 Arbeitnehmer). Diese wurden aus den Schätzungen vorerst ausgeschlossen.
Wir haben das zweite Trimester der BESTA99 als Periode t verwendet. Da in den Daten der BESTA die Gewichtung erst ab dem zweiten Trimester 2000 vorhanden ist, haben wir den Schätzer der Varianz des Schätzers der Anzahl Beschäftigter aus diesem Jahr verwendet. Von den theoretisch 112 Gruppen, die durch die Kreuzung der NOGA28 mit dem Geschlecht und dem Beschäftigungsgrad (vollzeit, teilzeit) entstehen, konnte anhand der BESTA die Anzahl Beschäftigter in 104 Gruppen geschätzt werden. Die 8 restlichen Gruppen befinden sich alle im Sektor 1. Die Anzahl Beschäftigter in diesen Gruppen muss somit aus einer Erhebung des 1. Sektors geschätzt werden. Eine vierteljährliche Erhebung besteht für den ersten Sektor zur Zeit nicht, deshalb ist es fraglich ob für den 1. Sektor vierteljährliche Lohnindices geschätzt werden können.
Für die Varianzschätzer hatten wir das SAS Programm der Sektion EWL zur Verfügung. Da wir das SAS Programm von EWL für unsere Bedürfnisse hätten anpassen müssen, haben wir es vorgezogen die SAS Prozedur `proc surveymeans` für die Varianzschätzung zu verwenden. Die so geschätzten Varianzen wichen nur minim von den von EWL veröffentlichten Schätzern ab.

9.2 Schätzung der Population

Für die Schätzung der Population fehlten uns die Daten der Grenzgänger und Saisonniers. Aufgrund der Schätzung der Anzahl Beschäftigter aus der SAKE und der BESTA gab es in der SAKE Beschäftigtengruppen, die im Vergleich zur BESTA unterschätzt wurden. Wir haben die Hypothese aufgestellt, dass die Grenzgänger und Saisonniers sich in diesen Gruppen befinden. Deshalb haben wir in diesen Gruppen proportional zum Unterschied mit der BESTA 180'000 solche Lohnempfänger aufgrund der Daten der Erwerbstätigenstatistik [2] generiert. Dieses Vorgehen ist methodologisch nicht einwandfrei, da unsere Hypothese auf keiner objektiven Argumentation basiert. Da wir davon ausgehen, dass diese Daten in Zukunft aus einem Register kommen, und keine Varianz aufweisen, führt der Einbezug dieser Lohnempfänger dazu, dass die Schätzer der Varianz die Varianz un-

terschätzen werden, im Gegensatz dazu, wenn wir davon ausgehen, dass die Daten dieser Lohnempfänger Schätzer mit einer Varianz sind. Ebenso werden in den Lohngruppen, in welchen Grenzgänger, etc. generiert wurden, der Variationskoeffizient unterschätzt, falls die Anzahl dieser Lohnempfänger trotzdem geschätzt werden muss.
Aus der Tabelle 15 ist ersichtlich, dass die Schätzung der Lohnempfänger in den Untersuchungsbereichen, die durch die NOGA28, das Geschlecht und den Beschäftigungsgrad definiert sind, nur in 102 Untersuchungsbereichen möglich war. Weitere 6 Untersuchungsbereiche befinden sich im 1. Sektor und der Rest der Untersuchungsbereiche konnte anhand der Daten der BESTA und der SAKE nicht gebildet werden. Aus der Schätzung der Harmonisierungsparameter $\hat{\alpha}_{d,t}$ ist ersichtlich, dass im Durchschnitt der Schätzer der Anzahl Beschäftigter der BESTA99II in den Untersuchungsbereichen um 4.11% höher ist als in der SAKE99. Insgesamt ist der Schätzer der Anzahl Beschäftigter basierend auf der BESTA99II jedoch nur 1.39% höher als der Schätzer der Anzahl Beschäftigter mit der SAKE als Basis inklusive der Grenzgänger. Dies erstaunt nicht, da wir den Unterschied mit der Generierung der Anzahl Saisonniers, Grenzgänger, etc. "optimiert" haben. Der Varianzschätzer der Anzahl Beschäftigten haben wir dem zweiten Trimester der BESTA2000 entnommen, da dieser für 1999 aufgrund von fehlenden Gewichten in den Daten nicht geschätzt werden konnte. Aus der Tabelle 15 und den Abbildungen 5 und 6 ist ersichtlich, dass in beinahe der Hälfte der Untersuchungsbereiche der Variationskoeffizient des Schätzers der Anzahl Lohnempfänger kleiner als 5% ist, da der Median bei 6.4% liegt. In mehr als 5% der Untersuchungsbereiche liegt die Anzahl Lohnempfänger zwischen 0, keine Lohnempfänger, und dem Doppelten des Schätzers der Anzahl Lohnempfänger, da ihr Variationskoeffizient grösser als 50% ist.
Für die Schätzung der Anzahl Lohnempfänger in U_j haben wir die Lohngruppen, wie von LOHN vorgeschlagen, anhand der Variablen NOGA28, Geschlecht, Beschäftigungsgrad, Altersklassen und ISCO-Code und wie in den Abschnitten 8.6 und 8.7 beschrieben, gebildet.
In 1013 Lohngruppen von insgesamt 2016, war es möglich die Varianz des Schätzers der Anzahl Lohnempfänger zu schätzen. Aus den Abbildungen 7 und 8 ist ersichtlich, dass in 10% der Lohngruppen der Variationskoeffizient kleiner als 5% ist. In 83 Lohngruppen (8.19%) kann keine Aussage über die Anzahl Lohnempfänger mehr gemacht werden, da der entsprechende Variationskoeffizient grösser als 50% ist. Mehr als die Hälfte der Lohngruppen weist einen Variationskoeffizienten von mehr als 14% auf. Somit genügt die Mehrheit der Schätzer der Population der Lohngruppen unseren Qualitätsansprüche an Schätzer der Population nicht. Die Lohngruppen können somit nicht als Einheit für die Schätzung der Lohnindices verwendet werden. Das heisst, es ist davon abzusehen Lohnindices auf diesem Niveau zu schätzen, da es zu fein ist.

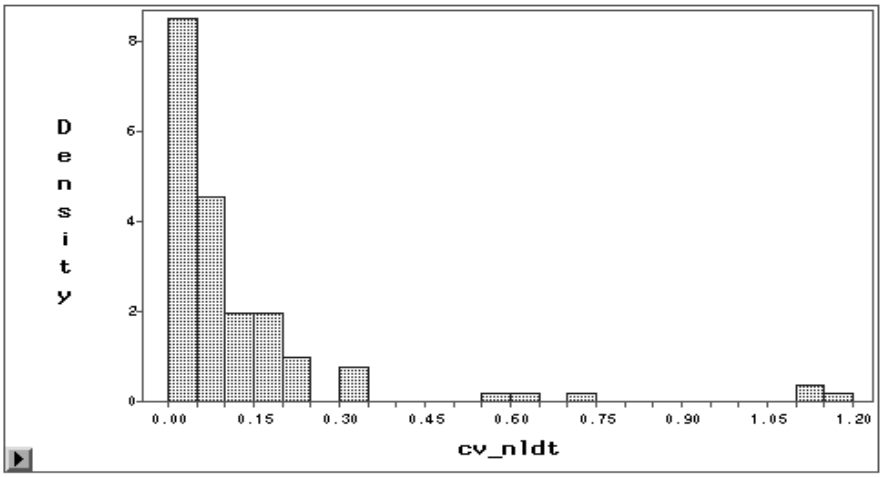


Abbildung 5: Verteilung der Variationskoeffizienten der Schätzer der Anzahl Lohnempfänger anhand der BESTA99II und der SAKE99 (’cv_nldt’) in den Untersuchungsbereichen, die durch die Kreuzung der NOGA28 mit dem Geschlecht und dem Beschäftigungsgrad entstanden sind.

Quantiles

100%	Max	1.1577	99.0%	1.1451
75%	Q3	0.1350	97.5%	1.1212
50%	Med	0.0640	95.0%	0.5532
25%	Q1	0.0295	90.0%	0.2312
0%	Min	0.0062	10.0%	0.0185
	Range	1.1515	5.0%	0.0161
	Q3-Q1	0.1056	2.5%	0.0117
	Mode	.	1.0%	0.0077

Abbildung 6: Percentile der Verteilung der Variationskoeffizienten der Schätzer der Anzahl Lohnempfänger anhand der BESTA99II und der SAKE99 (’cv_nldt’).

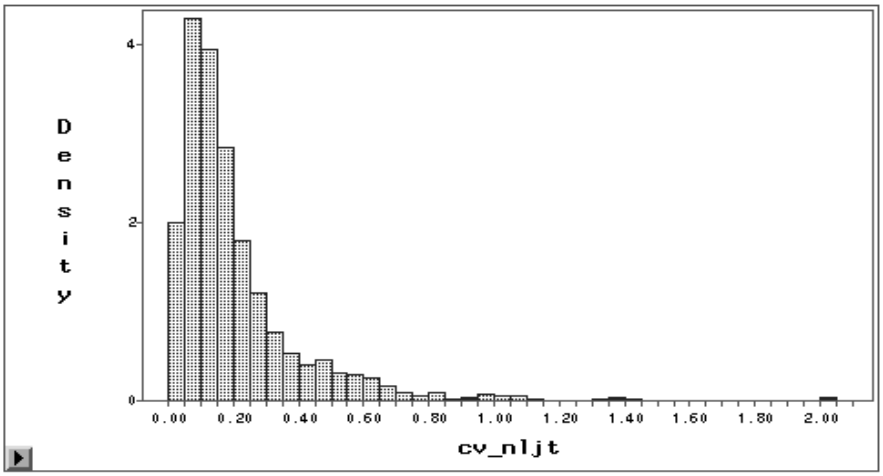


Abbildung 7: Verteilung der Variationskoeffizienten der Schätzer der Anzahl Lohnempfänger anhand der BESTA99II und der SAKE99 (’cv_nljt’) in 1013 Lohngruppen.

Quantiles

100%	Max	2.0226	99.0%	1.0484
75%	Q3	0.2543	97.5%	0.7678
50%	Med	0.1435	95.0%	0.6050
25%	Q1	0.0883	90.0%	0.4743
0%	Min	0.0080	10.0%	0.0500
	Range	2.0146	5.0%	0.0388
	Q3-Q1	0.1660	2.5%	0.0267
	Mode	0.0336	1.0%	0.0210

Abbildung 8: Percentile der Verteilung der Variationskoeffizienten der Schätzer der Anzahl Lohnempfänger anhand der BESTA99II und der SAKE99 (’cv_nljt’) in 1013 Lohngruppen.

9.3 Bias der Populations-Schätzer

Wie im Abschnitt 8.6 kurz erwähnt, müssen wir mit einem Bias für die Schätzer der Populationsgrösse der Lohngruppen rechnen. Da wir den Bias nicht exakt schätzen können, sondern nur eine obere Grenze für den absoluten Wert des Bias, verwenden wir diesen als Schätzer für den Bias. Mit diesem Vorgehen sind wir sicher, dass unsere Schätzer einen kleineren Bias aufweisen, als den hier geschätzten. Ist letzterer vernachlässigbar so können wir ihn mit Sicherheit vernachlässigen. Andernfalls können wir nichts genaueres sagen. Somit setzen wir aufgrund von (119)

$$|\hat{B}(\hat{r}^s_{l,d,j})| \doteq cv(\hat{N}^s_{l,d})\sqrt{\hat{V}ar(\hat{r}^s_{l,d,j})} \tag{128}$$

Vorausgesetzt wir können den Bias von $r^s_{l,d,j}$ (mit seinem Vorzeichen) schätzen, so können wir einen erwartungstreuen Schätzer $\tilde{r}^s_{l,d,j}$ von $r^s_{l,d,j}$ definieren, so dass

$$\tilde{r}^s_{l,d,j} \doteq \hat{r}^s_{l,d,j} - \hat{B}(\hat{r}^s_{l,d,j}) \doteq \frac{\tilde{N}^s_{l,j}}{\hat{N}^s_{l,d}}. \tag{129}$$

Wir versuchen nun einen Schätzer des Bias von $\hat{N}_{l,j,t}$ zu schätzen. Aus dem Schätzer der Anzahl Lohnempfänger in den Lohngruppen U_j (122) wird mit den obigen Schätzern somit

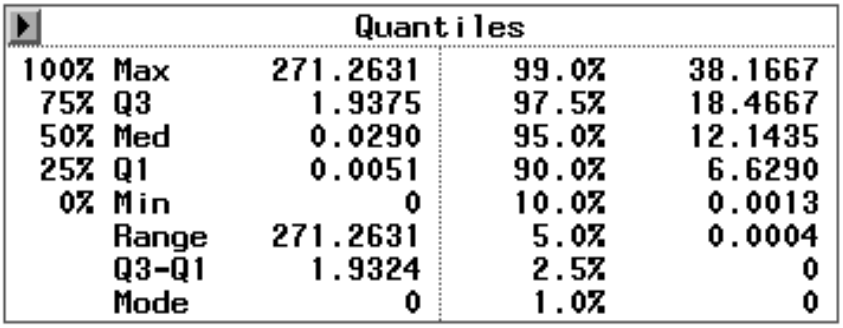
$$\begin{aligned}
\hat{N}_{l,j,t} &= \hat{N}_{l,d,t} \cdot (\tilde{r}^s_{l,d,j} + \hat{B}(\hat{r}^s_{l,d,j})) = \frac{\hat{N}_{l,d,t}}{\hat{N}^s_{l,d}}(\tilde{N}^s_{l,j} + \hat{B}(\hat{r}^s_{l,d,j})\hat{N}^s_{l,d}) \qquad \forall U_j \subset E_d \\
&\overset{(120)}{=} \frac{\hat{\alpha}_{d,t}\hat{N}^s_{l,d}}{\hat{N}^s_{l,d}}(\tilde{N}^s_{l,j} + \hat{B}(\hat{r}^s_{l,d,j})\hat{N}^s_{l,d}) \\
&= \hat{\alpha}_{d,t}\tilde{N}^s_{l,j} + \hat{\alpha}_{d,t}\hat{B}(\hat{r}^s_{l,d,j})\hat{N}^s_{l,d} \qquad \forall U_j \subset E_d \\
&\doteq \hat{\alpha}_{d,t}\tilde{N}^s_{l,j} + \hat{B}(\hat{N}_{l,j,t})
\end{aligned} \tag{130}$$

wobei wir davon ausgehen, dass $\tilde{N}_{l,j,t} \doteq \hat{\alpha}_{d,t}\tilde{N}^s_{l,j}$ ein erwartungstreuer Schätzer von $N_{j,t}$ ist. Anhand der Definition des Bias von $\hat{N}_{l,j,t}$ in (130) können wir eine obere Grenze für den absoluten Biasquotienten schätzen, ohne dass wir das Vorzeichen des Bias kennen müssen:

$$\begin{aligned}
|\widehat{BR}(\hat{N}_{l,j,t})| &= \frac{|\hat{B}(\hat{N}_{l,j,t})|}{\hat{V}ar(\hat{N}_{l,j,t})} \\
\overset{(130)}{\Rightarrow} |\widehat{BR}(\hat{N}_{l,j,t})| &= \frac{\hat{\alpha}_{d,t}|\hat{B}(\hat{r}^s_{l,d,j})|\hat{N}^s_{l,d}}{\hat{V}ar(\hat{N}_{l,j,t})}
\end{aligned} \tag{131}$$

Wir wenden das Kriterium aus [15] an, um zu bestimmen, ob der Bias vernachlässigbar ist oder nicht, das heisst, wenn der Biasquotient kleiner als 0.5 ist, gehen

wir davon aus, dass der Bias vernachlässigbar ist. Aus der Abbildung 9 ist ersichtlich, dass für eine Mehrheit der Lohngruppen der Bias vernachlässigbar ist. In 353 der 1013 Lohngruppen, beinahe 35% der Lohngruppen, müssen wir jedoch mit zum Teil massiven Verzerrungen rechnen.



Quantiles

100%	Max	271.2631	99.0%	38.1667
75%	Q3	1.9375	97.5%	18.4667
50%	Med	0.0290	95.0%	12.1435
25%	Q1	0.0051	90.0%	6.6290
0%	Min	0	10.0%	0.0013
	Range	271.2631	5.0%	0.0004
	Q3-Q1	1.9324	2.5%	0
	Mode	0	1.0%	0

Abbildung 9: Percentile der Verteilung des absoluten Biasquotienten des Schätzers der Anzahl Lohnempfänger in 1013 Lohngruppen ($|\widehat{BR}(\hat{N}_{l,j,t})|$).

9.4 Genauigkeit der Lohnindices-Schätzer

Wir werden in diesem Abschnitt die Genauigkeit des Schätzers des Lohnmassenindex und des Durchschnittslohnindex untersuchen. Hierfür konstruieren wir für den Lohnmassenindex einen 'Pseudo-Lohnmassenindex' anhand des Schätzers der Anzahl Beschäftigter aus der BESTA und dem Schätzer des Durchschnittslohns aus den Daten der SSUV. Für den Durchschnittslohnindex konstruieren wir eine untere Grenze für den Varianzschätzer. Anhand dieser Grenze zeigen wir auf, mit welcher Genauigkeit wir im besten Fall rechnen können. Um eine untere Grenze zu schätzen werden wir versuchen, die Schätzgrössen von $t_0 = 1993$ durch Schätzgrössen von $t = 1999$ zu ersetzen. Denn wir haben keinen Zugriff mehr auf die in unseren Formeln vorhandenen Schätzer von 1993, das heisst auf das von uns angestrebte Niveau der Schätzer von 1993.
Aus den aktuellen Publikationen ist zu schliessen, dass das 'erlaubte' Vertrauensintervall nicht grösser als 0.001, respektive 0.1 sein darf. Somit kann aus dem 95%-Vertrauensintervall der maximale Wert für den erlaubten Variationskoeffizi-

enten geschätzt werden

$$\begin{aligned} cv_{max}\left(\hat{I}^{D}_{d,t,t_0,\hat{\pi}}\right)\cdot 1.96\cdot 113 &= 0.05 \\ \Rightarrow cv_{max}\left(\hat{I}^{D}_{d,t,t_0,\hat{\pi}}\right) &= 2.26\cdot 10^{-4} \end{aligned} \tag{132}$$

9.4.1 Lohnmassenindex

Bevor wir einen 'Pseudo-Lohnmassenindex' konstruieren, versuchen wir zuerst eine untere Grenze des Varianzschätzers des Lohnmassenindex zu schätzen.
Zur Erinnerung geben wir den Varianzschätzer für den Lohnmassenindex der Lohngruppe j nochmals wieder, wie in (51)

$$\begin{aligned} \breve{V}ar\left(\hat{I}^{M}_{j,t,t_0}\right) &\approx \frac{1}{\left(\hat{M}_{j,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)+\left(\hat{I}^{M}_{j,t,t_0}\right)^2\hat{V}ar\left(\hat{M}_{j,t_0,\hat{\pi}}\right)\right] \\ &= \frac{1}{\left(\hat{M}_{j,t_0,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)+\left(\frac{\hat{M}_{j,t,\hat{\pi}}}{\hat{M}_{j,t_0,\hat{\pi}}}\right)^2\hat{V}ar\left(\hat{M}_{j,t_0,\hat{\pi}}\right)\right] \end{aligned}$$

Wir gehen davon aus, dass sich die nominale Lohnmasse seit 1993 mit $t = 1999 > t_0$ vergrössert hat, das heisst $\hat{M}_{j,t,\hat{\pi}} > \hat{M}_{j,t_0,\hat{\pi}}$. Somit ist

$$\begin{aligned} \breve{V}ar\left(\hat{I}^{M}_{j,t,t_0}\right) &> \frac{1}{\left(\hat{M}_{j,t,\hat{\pi}}\right)^2}\left[\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)+\hat{V}ar\left(\hat{M}_{j,t_0,\hat{\pi}}\right)\right] \\ &\overset{\substack{\hat{V}ar\left(\hat{M}_{j,t_0,\hat{\pi}}\right)\approx \\ \hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)}}{\approx} \frac{2\hat{V}ar\left(\hat{M}_{j,t,\hat{\pi}}\right)}{\left(\hat{M}_{j,t,\hat{\pi}}\right)^2} = 2cv\left(\hat{M}_{j,t,\hat{\pi}}\right)^2 \end{aligned} \tag{133}$$

Da die Form der Varianzschätzer der Schätzer des Lohnmassenindex über alle Lohngruppen und im Untersuchungsbereich E_d die gleiche Form haben, wie der Varianzschätzer für den Lohnmassenindex der Lohngruppe j, können wir die obige Ungleichheit (133) direkt anpassen:

$$\breve{V}ar\left(\hat{I}^{M}_{t,t_0}\right) > 2cv\left(\hat{M}_{t,\hat{\pi}}\right)^2 \tag{134}$$

$$\breve{V}ar\left(\hat{I}^{M}_{d,t,t_0}\right) > 2cv\left(\hat{M}_{d,t,\hat{\pi}}\right)^2 \tag{135}$$

Wir stellen die Hypothese auf, dass die Varianz der Schätzer der Lohnmassenindices als konstant betrachtet werden kann, das heisst $\sigma^2_{j,t} = \sigma^2_{j,t-1}$ in der Notation

des Abschnitts 4.2.5. Vorausgesetzt die Lohnmassenindices zweier aufeinander folgender Perioden korrelieren positiv miteinander, $\hat{\rho}_s\left(\hat{I}^M_{j,t,t_0}, \hat{I}^M_{j,t,t_0}\right) \in [0,1]$, ist der Varianzschätzer der Veränderung des Lohnmassenindex der Lohngruppe j folgendermassen begrenzt (siehe 55)

$$\hat{V}ar(\Delta^M_{j,t}) \in \left[0, 2\sigma^2_{j,t}\right] \supset \left[0, 4cv\left(\hat{M}_{j,t,\hat{\pi}}\right)^2\right].$$

Gleiches gilt für die Varianzschätzer der Veränderung der Lohnmassenindices über alle Lohngruppen und im Untersuchungsbereich E_d

$$\hat{V}ar(\Delta^M_{t}) \in \left[0, 2\sigma^2_{t}\right] \supset \left[0, 4cv\left(\hat{M}_{t,\hat{\pi}}\right)^2\right]$$
$$\hat{V}ar(\Delta^M_{d,t}) \in \left[0, 2\sigma^2_{d,t}\right] \supset \left[0, 4cv\left(\hat{M}_{d,t,\hat{\pi}}\right)^2\right].$$

Durch folgende Überlegungen versuchen wir eine untere Grenze für den Varianzschätzer $\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right)$ zu schätzen, um die Genauigkeit von $\breve{V}ar\left(\hat{I}^M_{d,t,t_0}\right)$ und um schlussendlich $\hat{V}ar(\Delta^M_{d,t})$ zu schätzen.

$$\begin{aligned}\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) &\overset{(42)}{=} \sum_{j\in J_d}\nu^2_{j,t}\hat{V}ar(\hat{N}_{l,j,t}) + \sum_{j\in J_d}(1-\hat{f}_{j,t})\hat{N}_{l,j,t}\hat{g}_{j,t} \\ &\quad -2\sum_{j\in J_d}\hat{\rho}_s(n_{j,t},\hat{N}_{l,j,t})\sqrt{(1-\hat{f}_{j,t})\hat{N}_{j,t}\hat{g}_{j,t}\hat{V}ar(\hat{N}_{l,j,t})} \\ &\doteq \sum_{j\in J_d}\nu^2_{j,t}\hat{V}ar(\hat{N}_{l,j,t}) + \hat{A}_{d,t}\end{aligned} \tag{136}$$

wobei

$$\begin{aligned}\hat{A}_{d,t} &\geq \sum_{j\in J_d}(1-\hat{f}_{j,t})\hat{N}_{l,j,t}\hat{g}_{j,t} \\ &\quad -2\sum_{j\in J_d}\sqrt{(1-\hat{f}_{j,t})\hat{N}_{l,j,t}\hat{g}_{j,t}\hat{V}ar(\hat{N}_{l,j,t})} \\ &= \sum_{j\in J_d}\sqrt{(1-\hat{f}_{j,t})\hat{N}_{l,j,t}\hat{g}_{j,t}}\left[\sqrt{(1-\hat{f}_{j,t})\hat{N}_{l,j,t}\hat{g}_{j,t}} - 2\sqrt{\hat{V}ar(\hat{N}_{l,j,t})}\right]\end{aligned} \tag{137}$$

Um eine untere Grenze für $\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right)$ schätzen zu können, muss $\hat{A}_{d,t}$ nach unten begrenzt sein. Da $(1-\hat{f}_{j,t})\hat{N}_{l,j,t}\hat{g}_{j,t} = \frac{\hat{N}_{l,j,t}(\hat{N}_{l,j,t}-n_{j,t})}{n_{j,t}} > 0$ ist, kann gezeigt

werden dass $\hat{A}_{d,t}$ immer positiv ist, wenn $\frac{\hat{N}_{l,j,t}(\hat{N}_{l,j,t}-n_{j,t})}{4n_{j,t}} > \hat{V}ar(\hat{N}_{l,j,t})$ ist. Unter der Annahme, dass $n_{j,t} = \hat{\pi}_{j,t}\hat{N}_{l,j,t}$ ist $\hat{A}_{d,t}$ immer positiv, wenn

$$\hat{\pi}_{j,t} < \frac{\hat{N}_{j,t}}{4\hat{V}ar(\hat{N}_{l,j,t}) + \hat{N}_{l,j,t}}$$

Anhand der 1013 Lohngruppen, für welche $\hat{N}_{l,j,t}$ und $\hat{V}ar(\hat{N}_{l,j,t})$ geschätzt werden konnten, haben wir die obere Limite für die Unfallwahrscheinlichkeit geschätzt, damit wir eine untere Grenze für $A_{d,t}$ schätzen können. Aus den Percentilen von $\hat{\pi}_{j,t}$ geht hervor, dass diese Grenze vorwiegend für Lohngruppen mit einer geringen Unfallwahrscheinlichkeit geschätzt werden kann, c.f. Abbildungen 10 und 11. Mit dieser Schätzung der Unfallwahrscheinlichkeit unterschätzen wir die Anzahl Verunfallter um mehr als einen Faktor 4 (32'656 anstatt 144'169). Anhand dieses Beispiels ist ersichtlich, welche Auswirkungen der Bias in den Schätzern der Population und die Approximationen in den Varianzschätzungen haben.
Somit sehen wir uns gezwungen, das Vorhaben eine untere Grenze des Varianzschätzers des Lohnmassenindex zu schätzen, aufzugeben, da keine untere Grenze gefunden werden konnte, die grösser als 0 ist.

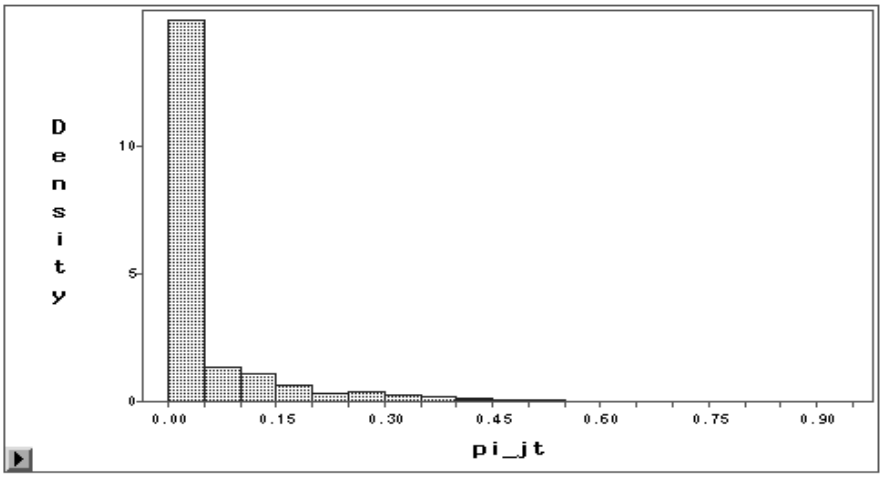
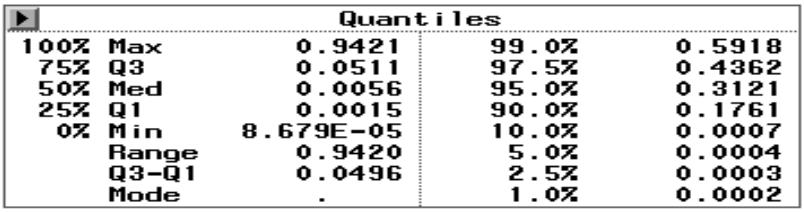


Abbildung 10: Verteilung der Schätzer der Unfallwahrscheinlichkeit aufgrund der Schätzer der Anzahl Lohnempfänger basierend auf der BESTA99II und der SAKE99 ('pi_jt') in 1013 Lohngruppen.



Quantiles

100%	Max	0.9421	99.0%	0.5918
75%	Q3	0.0511	97.5%	0.4362
50%	Med	0.0056	95.0%	0.3121
25%	Q1	0.0015	90.0%	0.1761
0%	Min	8.679E-05	10.0%	0.0007
	Range	0.9420	5.0%	0.0004
	Q3-Q1	0.0496	2.5%	0.0003
	Mode	.	1.0%	0.0002

Abbildung 11: Percentile der Verteilung der Schätzer der Unfallwahrscheinlichkeit aufgrund der Schätzer der Anzahl Lohnempfänger basierend auf der BESTA99II und der SAKE99 ('pi_jt') in 1013 Lohngruppen.

Simulation des Lohnmassenindex

Der Lohnmassenindex soll die Veränderung der ausgeschütteten Lohnmasse vierteljährlich wiedergeben. Für die meisten Lohnempfänger fluktuiert der Lohn innerhalb eines Jahres nicht, da er vertraglich auf ein Jahr fixiert ist. Diese Konstanz wird durch Stellenwechsel und Änderungen in der Hierarchie gestört. Weitere Fluktuationen im Lohn gehen von nicht ständig angestellten Lohnempfängern aus, welche keine fixen Löhne beziehen, jedoch eine untergeordnete Rolle spielen dürften.

Zusätzlich haben Veränderungen der Anzahl Lohnempfänger einen Einfluss auf die Fluktuationen des Lohnmassenindex. Über eine längere Beobachtungsperiode müssten tendenzielle und periodische Veränderungen der Anzahl Lohnempfänger im Lohnmassenindex sichtbar sein. Zum Beispiel hat der Wandel zu einer Dienstleistungsgesellschaft, die sich auf gewisse Branchen mit ihren spezifischen Löhnen spezialisiert einen Einfluss auf die Anzahl Lohnempfänger in den einzelnen Branchen. Ein periodischer Einfluss auf die Anzahl Lohnempfänger ist vor allem in der Baubranche zu beobachten. Es stellt sich die Frage, inwiefern Restrukturationen in Grossbetrieben einen Einfluss auf den Lohnmassenindex haben. Hierfür müssen Analysen auf einem genügend feingliedrigen Niveau durchgeführt werden.

Zu den Struktur bedingten und Lohn bedingten Fluktuationen kommt die Variabilität aufgrund der Stichprobenvarianz (51), (63) und (68). Diese zeichnet sich durch eine Unsicherheit aus, die je nach Genauigkeit des Schätzers des Lohnmassenindex einen einschneidenden Einfluss auf die Interpretation der Fluktuationen haben kann.

Wir haben zur Veranschaulichung der zu erwartenden Fluktuationen einen 'Pseudo-Lohnmassenindex' geschätzt. Da der Aufwand zu gross gewesen wäre für die Jahre 1993 bis 2000 die Anzahl Lohnempfänger zu schätzen, haben wir für die Schätzung der Lohnmasse den Schätzer der Anzahl Beschäftigter aus der BESTA

mit dem Schätzer des Durchschnittslohns der SSUV-Daten kombiniert.

$$\hat{M}_{d,t,\hat{\pi}} \doteq \hat{N}^B_{d,t}\nu_{d,t}. \tag{138}$$

Als Referenzperiode haben wir das vierte Quartal von 1993, das wir '1993_4' bezeichnen, festgelegt. Der Lohnmassenindex wurde für Männer und Frauen für das erste Quartal 1993 (1993_1) bis zum vierten Quartal 2000 (2000_4) separat geschätzt. Für $d \in \{\text{Männer}, \text{Frauen}\}$ ist somit

$$\begin{aligned} \breve{I}^M_{d,t,t_0} &= \frac{\hat{M}_{d,t,\hat{\pi}}}{\hat{M}_{d,t_0,\hat{\pi}}} \\ \breve{I}^M_{d,t,1993_4} &= \frac{\hat{M}_{d,t,\hat{\pi}}}{\hat{M}_{d,1993_4,\hat{\pi}}} \quad \forall t \in \{1993_1, \ldots, 2000_4\} \end{aligned} \tag{139}$$

Auf die gleiche Weise haben wir einen Index des SSUV-Durchschnittslohns kreiert

$$\breve{I}^S_{d,t,1993_4} = \frac{\nu_{d,t}}{\nu_{d,1993_4}} \quad \forall t \in \{1993_1, \ldots, 2000_4\} \tag{140}$$

und einen Index der Anzahl Beschäftigter aus der BESTA

$$\breve{I}^N_{d,t,1993_4} = \frac{\hat{N}^B_{d,t}}{\hat{N}^B_{d,1993_4}} \quad \forall t \in \{1993_1, \ldots, 2000_4\}. \tag{141}$$

Der Varianzschätzer des Lohnmassenschätzers in (139) hat unter den oben genannten Voraussetzungen und aufgrund von (42), und da die Stichprobe der BESTA unabhängig von den Daten der SSUV erhoben werden, folgende Form

$$\begin{aligned} \hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) &= \nu^2_{d,t}\left[\hat{V}ar(\hat{N}_{d,t}) + \left(1 - \hat{f}_{d,t}\right)\hat{N}_{d,t}\hat{g}_{d,t}\right] \\ &= \nu^2_{d,t}\left[\hat{V}ar(\hat{N}_{d,t}) + \hat{g}_{d,t}(\hat{N}_{d,t} - n_{d,t})\right] \end{aligned} \tag{142}$$

Aus pragmatischen Gründen stellen wir die Hypothese auf, dass der Varianzschätzer der Anzahl Vollzeitbeschäftigter nach Geschlecht aus der BESTA vom 1. Quartal 1993 bis zum 4. Quartal 2000 konstant ist. Wir haben somit den Varianzschätzer des vierten Quartals 2000 für sämtliche Quartale zwischen 1993 bis 2000 verwendet. Aufgrund dieser Hypothese und von (67) kann der Varianzschätzer des Lohnmassenindex für Männer und Frauen folgendermassen ausgedrückt

werden

$$\begin{aligned}
\hat{V}ar\left(\breve{I}^{M}_{d,t,1993_4}\right) &\doteq \frac{1+\left(\breve{I}^{M}_{d,t,1993_4}\right)^2}{\left(\hat{M}_{d,1993_4,\hat{\pi}}\right)^2}\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) \\
&\quad -2\frac{\breve{I}^{M}_{d,t,1993_4}\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right)}{\left(\hat{M}_{d,1993_4,\hat{\pi}}\right)^2}\hat{\rho}_s\left(\hat{M}_{d,t,\hat{\pi}},\hat{M}_{d,1993_4,\hat{\pi}}\right) \\
&\doteq \frac{\hat{V}ar\left(\hat{M}_{d,2000_4,\hat{\pi}}\right)}{\left(\hat{M}_{d,1993_4,\hat{\pi}}\right)^2}\left[1+\left(\breve{I}^{M}_{d,t,1993_4}\right)^2 - \right. \\
&\quad \left. \hat{\rho}_s\left(\hat{M}_{d,t,\hat{\pi}},\hat{M}_{d,1993_4,\hat{\pi}}\right)\right] \qquad (143) \\
&\overset{\hat{\rho}_s\approx 0.9}{\approx} \frac{\hat{V}ar\left(\hat{M}_{d,2000_4,\hat{\pi}}\right)}{\left(\hat{M}_{d,1993_4,\hat{\pi}}\right)^2}\left[0.1+\left(\breve{I}^{M}_{d,t,1993_4}\right)^2\right] \qquad (144)
\end{aligned}$$

wobei wir davon ausgehen, dass die Stichprobenkorrelation der Lohnmassenschätzer zum Lohnmassenschätzer der Referenzperiode mit 0.9 recht gross ist. $d \in \{\text{Männer}, \text{Frauen}\}, t \in \{1993_1, \ldots, 2000_4\} \setminus \{1993_4\}$ und

$$\hat{V}ar\left(\hat{M}_{d,t,\hat{\pi}}\right) = \nu^2_{d,t}\left[\hat{V}ar(\hat{N}_{d,2000_4}) + \hat{g}_{d,t}(\hat{N}_{d,t} - n_{d,t})\right]$$

Für das 4. Quartal 1993 ist (144)

$$\hat{V}ar\left(\breve{I}^{M}_{d,1993_4,1993_4}\right) \approx \frac{\hat{V}ar\left(\hat{M}_{d,2000_4,\hat{\pi}}\right)}{\left(\hat{M}_{d,1993_4,\hat{\pi}}\right)^2}.$$

Da wir nicht über genügend Information verfügen, um die Stichprobenkorrelation zweier aufeinanderfolgender Lohnmassenindices zu schätzen, stützen wir die Vertrauensintervalle auf (143) mit $\hat{\rho}_s\left(\hat{M}_{d,t,\hat{\pi}},\hat{M}_{d,1993_4,\hat{\pi}}\right) = 0.9$ ab. Die Resultate der Simulation finden sich in der Tabelle 16 im Anhang F. Daraus ist ersichtlich, dass mit einem Variationskoeffizienten von 0.59% bis 1.26% für den Lohnmassenindex der Männer und 1.17% bis 2.15% für den Lohnmassenindex der Frauen gerechnet werden muss. Die angestrebte Genauigkeit von ungefähr $2 \cdot 10^{-6} \cdot \breve{I}^{M}_{d,t,1993_4}$ (c.f. 132) kann bei weitem nicht erreicht werden. Aus der Abbildung 12 ist klar ersichtlich, dass der Lohnmassenindex (139) der Männer ein periodisches Verhalten im 4. Quartal aufweist. Dies hat mit der Abnahme der Anzahl Lohnempfänger, wahrscheinlich der Anzahl Lohnempfänger auf dem Bau zu tun. Der Anstieg im ersten

Quartal ist jedoch auf einen Lohnanstieg zurückzuführen. Unerwartet fällt der Index des Durchschnittslohns SSUV der Männer im 2. Quartal jeweils wieder. Ist somit der Anstieg des Durchschnittslohns auf ein höheres Unfallrisiko der besser verdienenden zurückzuführen, da diese sich vor allem anfangs Jahr den Gefahren des Wintersports aussetzen? Oder sind die fehlenden Bauarbeiter, die unterdurchschnittlich verdienen, für den Anstieg des Durchschnittslohns verantwortlich? Wir überlassen es vorerst den Fachleuten, eine Interpretation der Lohnkurve des Index für den SSUV-Durchschnittslohn zu finden. Sobald jedoch die Unfallwahrscheinlichkeit für bestimmte Lohngruppen sich saisonal verändert, müsste dieser Veränderung in der Schätzung des Lohnmassenindex Rechnung getragen werden. Das heisst, die Unfallwahrscheinlichkeit müsste vierteljährlich für jede Lohngruppe geschätzt werden können.
Wir werden nicht weiter auf die Interpretation der Schwankungen eingehen, obwohl aus den Abbildungen 12, 13 und 14 klar hervorgeht, dass diesbezüglich ein Handlungsbedarf besteht.

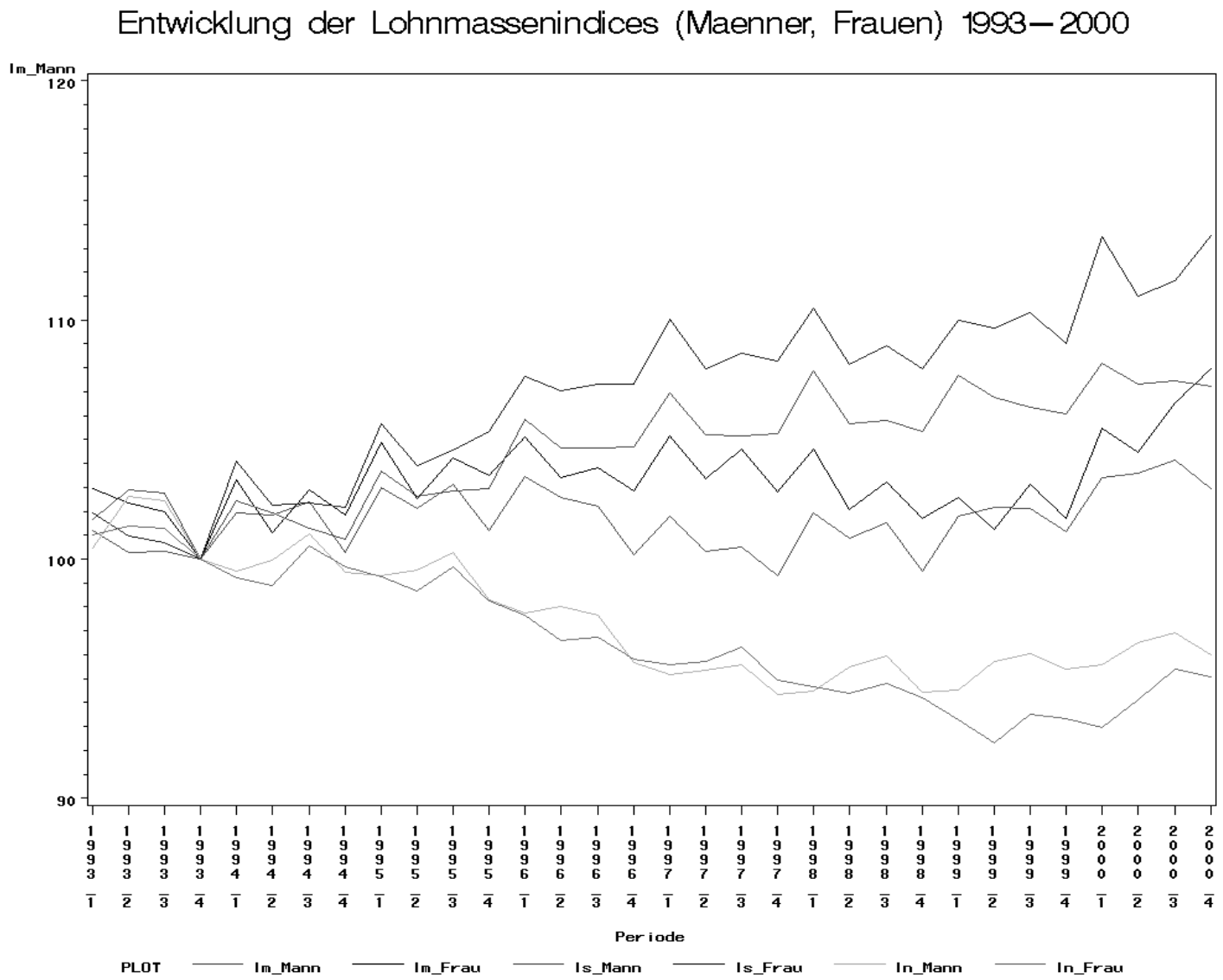


Abbildung 12: Lohnmassenindex der Vollzeitbeschäftigten nach Geschlecht (`Im_Mann`, `Im_Frau`) vom ersten Quartal 1993 bis zum vierten Quartal 2000, siehe (139). Basis: viertes Quartal 1993. Zudem ein Beschäftigtenindex (`In_Mann`, `In_Frau`), siehe (141) und ein Index des SSUV-Durchschnittslohns (`Is_Mann`, `Is_Frau`) nach Geschlecht, siehe (140).

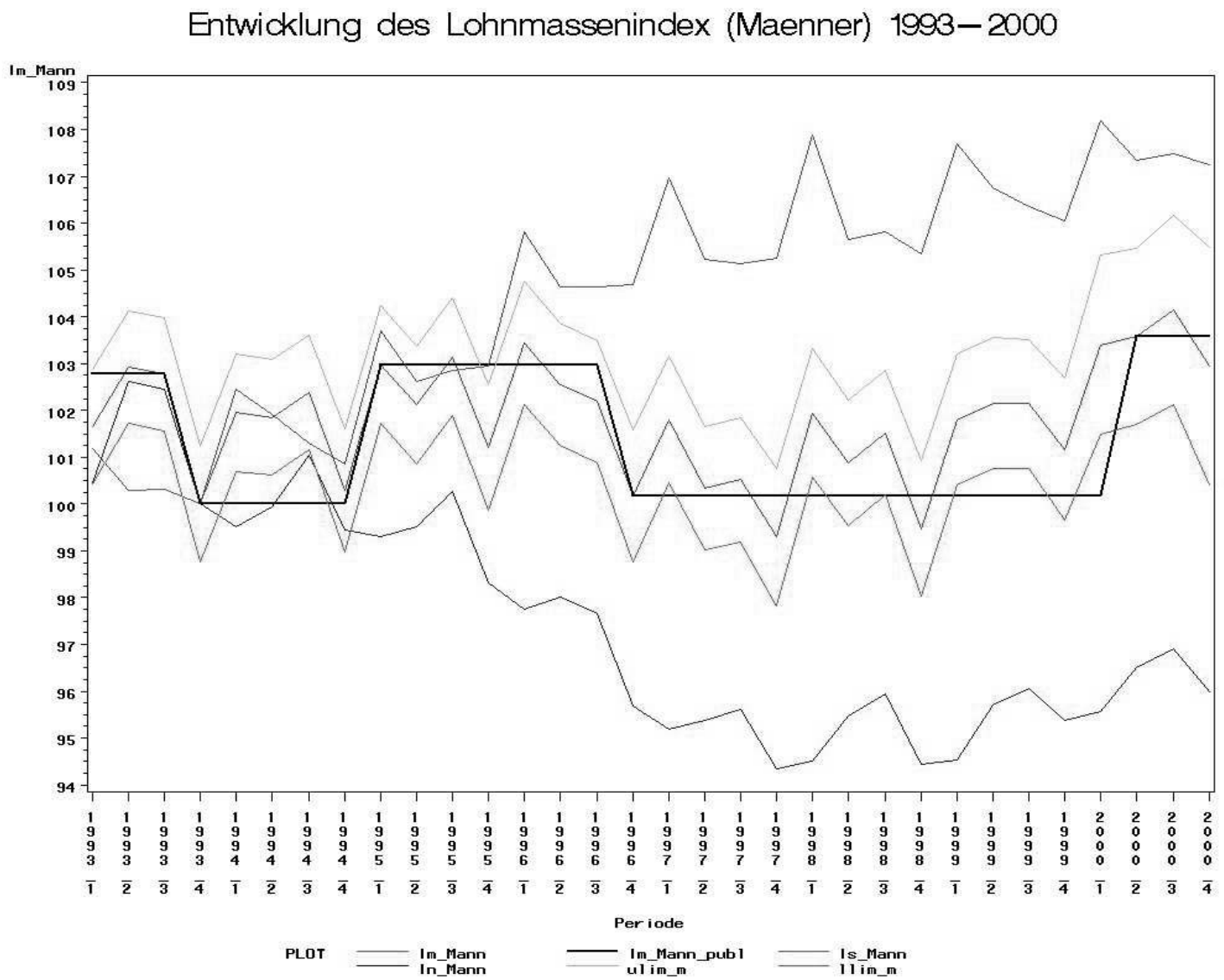


Abbildung 13: Lohnmassenindex der vollzeitbeschäftigten Männer (`lm_Mann`) vom ersten Quartal 1993 bis zum vierten Quartal 2000 mit der unteren und oberen Grenze des Vertrauensintervalls (`llim_m`, `ulim_m`). Zudem ein Beschäftigtenindex der Männer (`ln_Mann`) und ein Index des SSUV-Durchschnittslohns der Männer (`ls_Mann`). `lm_Mann_publ` zeigt die Entwicklung auf, die auf signifikanten Unterschieden der Lohnmassenindices basiert und somit tatsächlich publiziert werden kann. Die Basis aller Indices ist das vierte Quartal 1993.

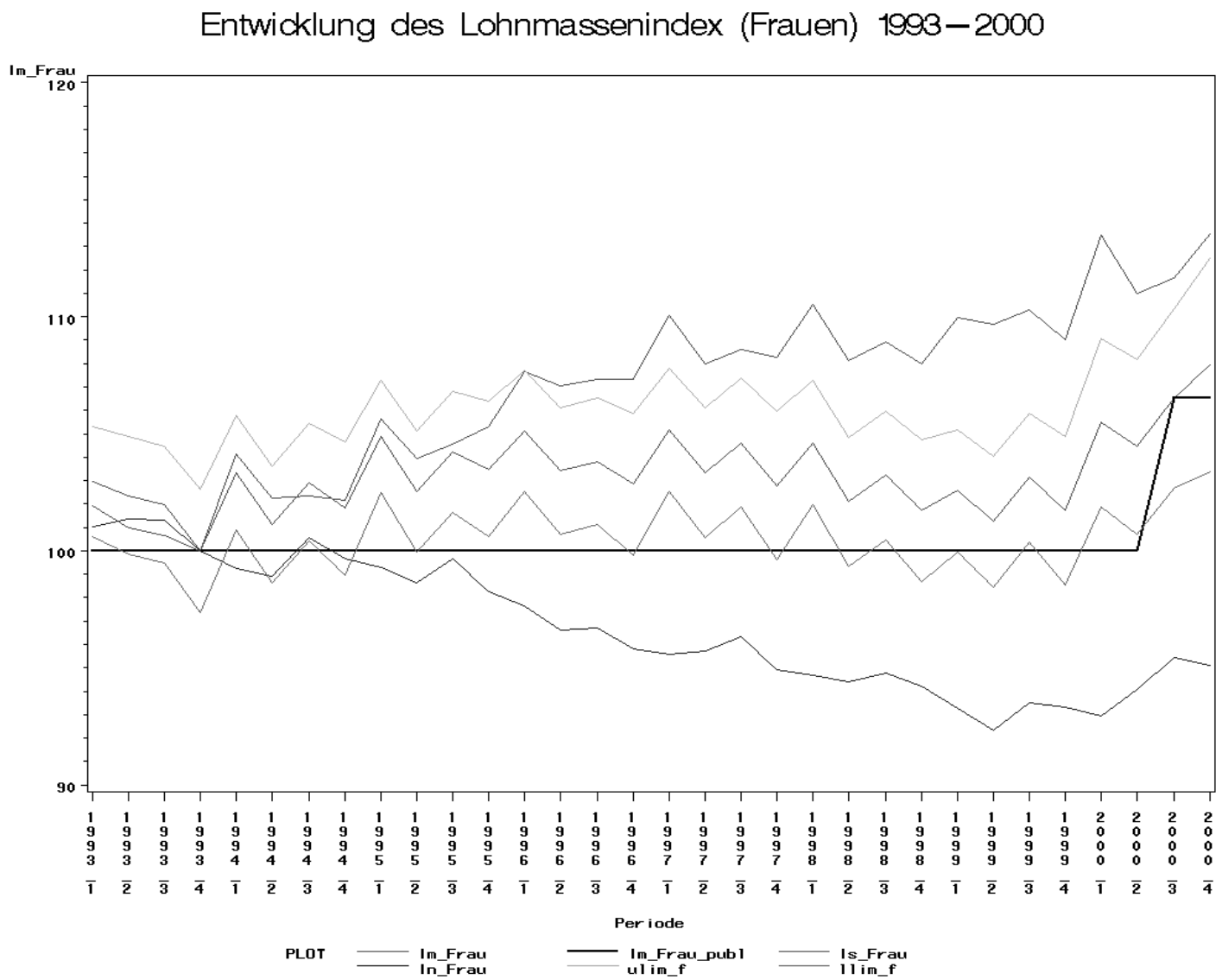


Abbildung 14: Lohnmassenindex der vollzeitbeschäftigten Frauen (`Im_Frau`) vom ersten Quartal 1993 bis zum vierten Quartal 2000 mit der unteren und oberen Grenze des Vertrauensintervalls (`llim_f`, `ulim_f`). Zudem ein Beschäftigtenindex der Frauen (`In_Frau`) und ein Index des SSUV-Durchschnittslohns der Frauen (`Is_Frau`). `Im_Frau_publ` zeigt die Entwicklung auf, die auf signifikanten Unterschieden der Lohnmassenindices basiert und somit tatsächlich publiziert werden kann. Die Basis aller Indices ist das vierte Quartal 1993.

9.4.2 Durchschnittslohnindex

Zur Erinnerung geben wir nochmals den Varianzschätzer des Durchschnittslohnindex in der Lohngruppe j, wie er in der Formel (74) gegeben wurde

$$\breve{V}ar\left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right) \approx \frac{1}{\nu^2_{j,t_0}}\left[\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right) + \left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right)^2 \hat{V}ar\left(\hat{\mu}_{j,t_0,\hat{\pi}}\right)\right]$$

Aufgrund der Homogenität der Lohngruppen in der Zeit, gehen wir davon aus, dass $\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right) \approx \hat{V}ar\left(\hat{\mu}_{j,t_0,\hat{\pi}}\right)$, das heisst, wir gehen von einer konstanten Struktur innerhalb der Lohngruppen aus. Somit haben wir

$$\begin{aligned}
\breve{V}ar\left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right) &\approx \frac{1}{\nu^2_{j,t_0}}\left[\left(1+\left(\hat{I}^D_{j,t,t_0,\hat{\pi}}\right)^2\right)\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right)\right] \\
&\overset{\hat{I}^D_{j,t,t_0,\hat{\pi}}>0}{>} \frac{1}{\nu^2_{j,t_0}}\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right) \\
&\overset{\nu_{j,t}>\nu_{j,t_0}}{>} \frac{1}{\nu^2_{j,t}}\hat{V}ar\left(\hat{\mu}_{j,t,\hat{\pi}}\right) \\
&\overset{(45)}{=} \frac{1}{\nu^2_{j,t}}\frac{\nu^2_{j,t}}{n_{j,t}}\left(1-\hat{f}_{j,t}\right) \\
&= \frac{\hat{N}_{l,j,t}-n_{j,t}}{\hat{N}_{l,j,t}n_{j,t}} \overset{4.2.3}{=} \frac{\hat{N}_{l,j,t}(1-\hat{\pi}_{j,t})}{\hat{N}^2_{l,j,t}\hat{\pi}_{j,t}} = \frac{1-\hat{\pi}_{j,t}}{\hat{N}_{l,j,t}\hat{\pi}_{j,t}} \qquad (145)
\end{aligned}$$

Da wir die Lohngruppen zur Zeit nicht bilden können (ISCO-Code fehlt in den Daten der SSUV), beschränken wir uns darauf, eine untere Grenze für den Varianzschätzer des Durchschnittslohns im Untersuchungsbereich E_d, gebildet durch die NOGA28, das Geschlecht und den Beschäftigungsgrad, zu schätzen. Der Varianzschätzer von $\hat{I}^D_{d,t,t_0,\hat{\pi}}$ ist, c.f. (82),

$$\begin{aligned}
\breve{V}ar\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\right) &\approx \frac{1}{\hat{M}^2_{d,t_0,\hat{\pi}}}\sum_{j\in J_d}\left(\nu^2_{j,t}+\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\nu_{j,t_0}\right)^2\right)\hat{V}ar\left(\hat{N}_{l,j,t_0}\right) + \\
&\frac{1}{\hat{M}^2_{d,t_0,\hat{\pi}}}\sum_{j\in J_d}\underbrace{\left\{\frac{(1-\hat{f}_{j,t})}{n_{j,t}}\left(\hat{N}_{l,j,t_0}\nu_{j,t}\right)^2+\frac{(1-\hat{f}_{j,t_0})}{n_{j,t_0}}\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\hat{M}_{j,t_0,\hat{\pi}}\right)^2\right\}}_{>0\ \forall j} \\
&> \frac{1}{\hat{M}^2_{d,t_0,\hat{\pi}}}\sum_{j\in J_d}\left(\nu^2_{j,t}+\underbrace{\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\nu_{j,t_0}\right)^2}_{>0\ \forall j}\right)\hat{V}ar\left(\hat{N}_{l,j,t_0}\right)
\end{aligned}$$

$$
\begin{array}{ccccc}
 & > & \displaystyle\sum_{j\in J_d} \frac{\nu_{j,t}^2 \hat{V}ar\left(\hat{N}_{l,j,t_0}\right)}{\hat{M}_{d,t_0,\hat{\pi}}^2} & \overset{\hat{M}_{d,t,\hat{\pi}} > \hat{M}_{d,t_0,\hat{\pi}}}{>} & \displaystyle\sum_{j\in J_d} \frac{\nu_{j,t}^2 \hat{V}ar\left(\hat{N}_{l,j,t_0}\right)}{\hat{M}_{d,t,\hat{\pi}}^2} \\
\overset{\hat{V}ar\left(\hat{N}_{l,j,t_0}\right)\approx}{\underset{\approx}{\hat{V}ar\left(\hat{N}_{l,j,t}\right)}} & & \displaystyle\sum_{j\in J_d} \frac{\nu_{j,t}^2 \hat{V}ar\left(\hat{N}_{l,j,t}\right)}{\hat{M}_{d,t,\hat{\pi}}^2} & &
\end{array}
$$

Bis anhin wurden nur die Vollzeitbeschäftigten im SLI berücksichtigt. Somit besitzen wir zurzeit nur die Löhne der Vollzeitbeschäftigten. Dies erlaubt es uns, $\nu_{j,t} > 1000$ zu setzen, wie auch aus den Daten der SSUV von 1999 ersichtlich ist. Dies ergibt

$$
\breve{V}ar\left(\hat{I}_{d,t,t_0,\hat{\pi}}^D\right) > \frac{1'000'000}{\hat{M}_{d,t,\hat{\pi}}^2} \sum_{j\in J_d} \hat{V}ar\left(\hat{N}_{l,j,t}\right)
$$

In einem nächsten Schritt sehen wir, dass auf Niveau NOGA für die Vollzeitbeschäftigten alle Durchschnittslohnindices im Jahre 1999 kleiner als 113 waren. Deshalb setzen wir zur Vereinfachung $\hat{I}_{d,t,t_0,\hat{\pi}}^D = 1.13$, um einen Schätzer für den Variationskoeffizienten zu schätzen. Somit gehen wir davon aus, dass

$$
cv\left(\hat{I}_{d,t,t_0,\hat{\pi}}^D\right) > \widetilde{cv}\left(\hat{I}_{d,t,t_0,\hat{\pi}}^D\right) \doteq \frac{\sqrt{\breve{V}ar\left(\hat{I}_{d,t,t_0,\hat{\pi}}^D\right)}}{1.13} > \frac{1'000}{1.13\hat{M}_{d,t,\hat{\pi}}} \sqrt{\sum_{j\in J_d} \hat{V}ar\left(\hat{N}_{l,j,t}\right)} \tag{146}
$$

In den Lohngruppen, in welchen wir den Varianzschätzer der Anzahl Lohnempfänger nicht schätzen konnten, haben wir den kleinsten Varianzschätzer, der geschätzt werden konnte eingesetzt. Aus den Abbildungen 15 und 16 ist ersichtlich, dass wir unter den oben beschriebenen Voraussetzungen mit einem maximalen Variationskoeffizienten, $\max_d\left(\widetilde{cv}\left(\hat{I}_{d,t,t_0,\hat{\pi}}^D\right)\right)$, rechnen müssen, der grösser als 3% ist, wenn wir die beiden grössten Werte (Ausreisser) vernachlässigen. Dies würde bei einem Lohnindex von 1.13, respektive 113 in der aktuellen Publikationsform des SLI, heissen, dass dieser in einem Intervall liegt, dass grösser ist als $[1.064, 1.196]$, respektive $[106.4, 119.6]$. Wie in 132 aufgezeigt wurde, ist aufgrund der aktuellen Publikationen der maximal erlaubte Variationskoeffizient $2.26 \cdot 10^{-4}$. Aus der Abbildung 16 geht aber hervor, dass der minimale Variationskoeffizient von $\hat{I}_{d,t,t_0,\hat{\pi}}^D$ grösser ist als 0.0031 und somit mehr als 13 mal grösser ist, als der maximal erlaubte. Wir haben in der Tabelle 6 die Vertrauensintervalle mit einem Fehler erster Art von 5% für $\hat{I}_{d,t,t_0,\hat{\pi}}^D \in \mathbb{N} \cap [95, 113]$ simuliert. Dabei haben wir den minimalen und den maximalen Variationskoeffizienten verwendet um die Vertrauensintervalle zu schätzen, das heisst $\min_d\left(\widetilde{cv}\left(\hat{I}_{d,t,t_0,\hat{\pi}}^D\right)\right)$ und

$\max_d \left(\widetilde{cv} \left(\hat{I}^D_{d,t,t_0,\hat{\pi}} \right) \right)$. Es ist klar ersichtlich, dass selbst mit dem minimalen Variationskoeffizienten die Grösse des Vertrauensintervalls immer grösser als ein Indexpunkt ist. Der Durchschnittslohnindex hat somit auf diesem Niveau nicht die derzeit geforderte Aussagekraft und sollte deshalb nie auf diesem Niveau (NOGA28 x Geschlecht der Vollzeitbeschäftigten) mit den derzeitigen Ansprüchen an die Genauigkeit publiziert werden.

Tabelle 6: Simulation der minimalen Vertrauensintervalle mit dem minimalen und dem maximalen Variationskoeffizienten des Schätzers des Durchschnittslohnindex in den Untersuchungsbereichen, welche durch die NOGA28, das Geschlecht und den Beschäftigungsgrad=Vollzeitbeschäftigt gebildet wurden. Die effektiven Vertrauensintervalle sind grösser als die hier geschätzten.

$\hat{I}^D_{d,t,t_0,\hat{\pi}}$	$\mathbf{cv}\left(\hat{\mathbf{I}}^{\mathbf{D}}_{\mathbf{d,t,t_0},\hat{\pi}}\right) = \mathbf{0.0031}$	$\mathbf{cv}\left(\hat{\mathbf{I}}^{\mathbf{D}}_{\mathbf{d,t,t_0},\hat{\pi}}\right) = \mathbf{0.03}$
95	$[94.4, 95.6]$	$[89.4, 100.6]$
96	$[95.4, 96.6]$	$[90.4, 101.6]$
97	$[96.4, 97.6]$	$[91.3, 102.7]$
98	$[97.4, 98.6]$	$[92.2, 103.8]$
99	$[98.4, 99.6]$	$[93.2, 104.8]$
100	$[99.4, 100.6]$	$[94.1, 105.9]$
101	$[100.4, 101.6]$	$[95.1, 106.9]$
102	$[101.4, 102.6]$	$[96.0, 108.0]$
103	$[102.4, 103.6]$	$[96.9, 109.1]$
104	$[103.4, 104.6]$	$[97.9, 110.1]$
105	$[104.4, 105.6]$	$[98.8, 111.2]$
106	$[105.4, 106.6]$	$[99.8, 112.2]$
107	$[106.3, 107.7]$	$[100.7, 113.3]$
108	$[107.3, 108.7]$	$[101.6, 114.4]$
109	$[108.3, 109.7]$	$[102.6, 115.4]$
110	$[109.3, 110.7]$	$[103.5, 116.5]$
111	$[110.3, 111.7]$	$[104.5, 117.5]$
112	$[111.3, 112.7]$	$[105.4, 118.6]$
113	$[112.3, 113.7]$	$[106.4, 119.6]$

Wie bereits erwähnt, interessiert die Veränderung ebenso wie die absolute Höhe des SLI. Für die Schätzung der Varianz der Veränderung $\Delta^D_{d,t}$ zweier aufeinanderfolgender Durchschnittslohnindices stellen wir die Hypothese auf, dass ihre

Stichprobenkorrelation positiv ist, da sich die beiden Schätzer von einer Periode zur anderen nur durch die Schätzung des Durchschnittslohn unterscheiden. $\hat{\rho}_s\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}, \hat{I}^D_{d,t-1,t_0,\hat{\pi}}\right) \in [0,1]$, und dass $\breve{V}ar\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}\right) = \breve{V}ar\left(\hat{I}^D_{d,t-1,t_0,\hat{\pi}}\right)$ ist. Mit der Notation des Abschnitts 4.2.6 ist somit

$$\hat{V}ar(\Delta^D_{d,t}) \in \left[0, 2(\hat{\sigma}^D_{d,t})^2\right] \tag{147}$$

Mit einem Fehler erster Art von 5% ($\alpha = 5\%$) nimmt das Vertrauensintervall der Veränderung folgende Form an

$$\begin{aligned} \Delta^D_{d,t} &\in \left[\Delta^D_{d,t} - 1.96\sqrt{2 * (1-\hat{\rho}_s)}\hat{\sigma}^D_{d,t}, \Delta^D_{d,t} + 1.96\sqrt{2 * (1-\hat{\rho}_s)}\hat{\sigma}^D_{d,t}\right] \\ &\doteq \left[\Delta^D_{d,t} - \hat{\delta}^D_{d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{d,t}\right] \end{aligned} \tag{148}$$

Das 95%-Vertrauensintervall der Veränderung ist unter den oben genannten Voraussetzungen maximal, wenn keine Korrelation zwischen den beiden aufeinanderfolgenden Durchschnittslohnindices vorhanden ist, was jedoch eher unwahrscheinlich ist. $\hat{\rho}_s\left(\hat{I}^D_{d,t,t_0,\hat{\pi}}, \hat{I}^D_{d,t-1,t_0,\hat{\pi}}\right) = 0$:

$$\begin{aligned} \Delta^D_{d,t} &\in \left[\Delta^D_{d,t} - 1.96\sqrt{2}\hat{\sigma}^D_{d,t}, \Delta^D_{d,t} + 1.96\sqrt{2}\hat{\sigma}^D_{d,t}\right] \\ &\doteq \left[\Delta^D_{d,t} - \hat{\delta}^D_{max,d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{max,d,t}\right] \end{aligned} \tag{149}$$

Aus den Abbildungen 17 und 18 geht hervor, dass das Vertrauensintervall in (149) im besten Fall folgende Form annimmt, wobei $\min_d \hat{\delta}^D_{max,d,t} = 0.0096$,

$$\left[\Delta^D_{d,t} - \hat{\delta}^D_{max,d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{max,d,t}\right]_{\min} = \left[\Delta^D_{d,t} - 0.0096, \Delta^D_{d,t} + 0.0096\right]$$

und ohne die Ausreisser maximal folgende Form annimmt, wobei $\max_d \hat{\delta}^D_{max,d,t} = 0.0980$,

$$\left[\Delta^D_{d,t} - \hat{\delta}^D_{max,d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{max,d,t}\right]_{\max} = \left[\Delta^D_{d,t} - 0.0980, \Delta^D_{d,t} + 0.0980\right]$$

Aufgrund der Überlegung, dass ein Teil des Durchschnittslohnindex nur für die Referenzperiode geschätzt wird und dieser Anteil an der Variabilität konstant bleibt, sowie aufgrund der Hypothese, dass die Anzahl Verunfallter von einer Periode zur darauffolgenden in den Untersuchungsbereichen eher konstant bleibt, sprich die gleiche Unfallwahrscheinlichkeit haben, gehen wir davon aus, dass die Stichprobenkorrelation von zwei aufeinanderfolgenden Durchschnittslohnindices

nahe bei 1 ist [10]. Somit ist das Vertrauensintervall wenn $\hat{\rho}_s$ nahe bei 1 ist, kleiner als wenn die Stichprobenkorrelation $\hat{\rho}_s = 0$ ist. Wenn nun also $\hat{\rho}_s \left(\hat{I}^D_{d,t,t_0,\hat{\pi}}, \hat{I}^D_{d,t-1,t_0,\hat{\pi}} \right) = 0.99$ ist, nimmt das 95%-Vertrauensintervall $VI(\Delta^D_{d,t})$ der Veränderung $\Delta^D_{d,t}$ folgende Form an

$$\begin{aligned} VI(\Delta^D_{d,t}) &\doteq \left[\Delta^D_{d,t} - 1.96cv(\Delta^D_{d,t})\Delta^D_{d,t}, \Delta^D_{d,t} + 1.96cv(\Delta^D_{d,t})\Delta^D_{d,t}\right] \\ &= \left[\Delta^D_{d,t} - 1.96\sqrt{0.02}\hat{\sigma}^D_{d,t}, \Delta^D_{d,t} + 1.96\sqrt{0.02}\hat{\sigma}^D_{d,t}\right] \\ &\doteq \left[\Delta^D_{d,t} - \hat{\delta}^D_{min,d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{min,d,t}\right] \end{aligned} \tag{150}$$

Unter den oben beschriebenen Voraussetzungen ist $\hat{\delta}^D_{max,d,t} = 10\hat{\delta}^D_{min,d,t}$. Aus den Abbildungen 19 und 20 geht hervor, dass das Vertrauensintervall in (150) im besten Fall folgende Form annimmt, wenn $\hat{\rho}_s = 0.99$

$$\left[\Delta^D_{d,t} - \hat{\delta}^D_{min,d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{min,d,t}\right]_{\min} = \left[\Delta^D_{d,t} - 0.001, \Delta^D_{d,t} + 0.001\right]$$

wobei $\min_d \hat{\delta}^D_{min,d,t} = 0.001$. Ohne die Ausreisser nimmt das Vertrauensintervall in (150) maximal folgende Form an, wiederum wenn $\hat{\rho}_s = 0.99$ und $\max_d \hat{\delta}^D_{min,d,t} = 0.0098$

$$\left[\Delta^D_{d,t} - \hat{\delta}^D_{min,d,t}, \Delta^D_{d,t} + \hat{\delta}^D_{min,d,t}\right]_{\max} = \left[\Delta^D_{d,t} - 0.0098, \Delta^D_{d,t} + 0.0098\right]$$

Die Veränderung in der aktuellen Publikations-Form geht von einer Genauigkeit von 0.0005 aus. Diese kann für diese Untersuchungsbereiche bei weitem nicht erreicht werden, wenn wir bedenken dass wir oben ein Intervall angegeben haben, von dem wir wissen, dass es kleiner ist als das wirkliche Vertrauensintervall. Mit anderen Worten, eine Veränderung des Durchschnittslohnindex der Vollzeitbeschäftigten auf Niveau NOGA28 x Geschlecht von weniger als 1% (1 Index-Punkt) ist nicht eindeutig interpretierbar.

[10] Die Auswirkungen von saisonalen Schwankungen haben keinen Einfluss auf den jährlich geschätzten Durchschnittslohnindex. Diese müssen aber bei der Schätzung der Genauigkeit des Lohnmassenschätzers berücksichtigt werden.

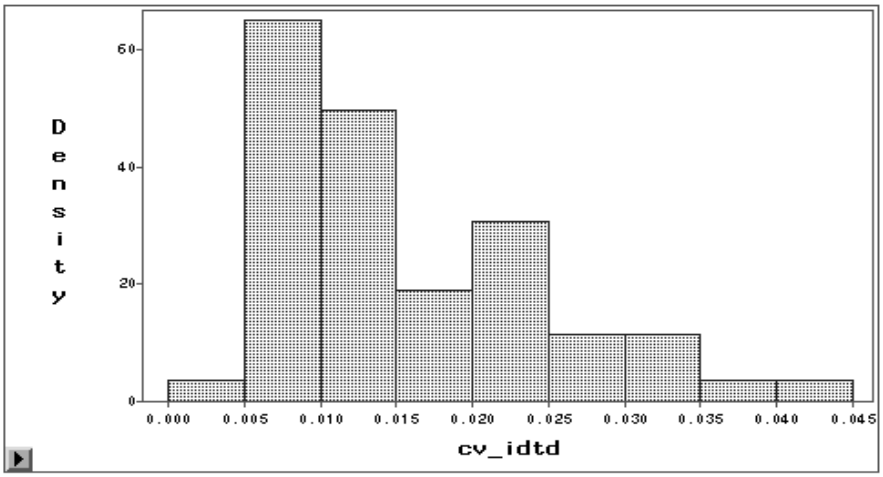


Abbildung 15: Verteilung der Variationskoeffizienten des Schätzers des Durchschnittslohnindex der Vollzeitbeschäftigten ('cv_idtd'). Die Anzahl Lohnempfänger wurde anhand der BESTA99II und der SAKE99 in 52 Lohngruppen geschätzt.

Quantiles

100%	Max	0.0405	99.0%	0.0405
75%	Q3	0.0215	97.5%	0.0398
50%	Med	0.0128	95.0%	0.0313
25%	Q1	0.0089	90.0%	0.0298
0%	Min	0.0031	10.0%	0.0060
	Range	0.0374	5.0%	0.0054
	Q3-Q1	0.0126	2.5%	0.0052
	Mode	.	1.0%	0.0031

Abbildung 16: Percentile der Verteilung der Variationskoeffizienten des Schätzers des Durchschnittslohnindex der Vollzeitbeschäftigten ('cv_idtd'). Die Anzahl Lohnempfänger wurde anhand der BESTA99II und der SAKE99 in 52 Lohngruppen geschätzt.

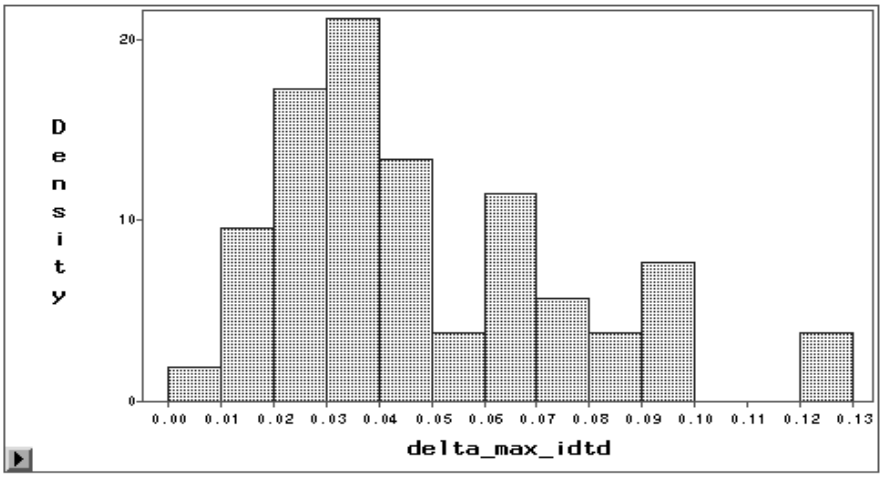


Abbildung 17: Verteilung der halben Länge des Vertrauensintervalls des Veränderungs-Schätzers des Durchschnittslohnindex aufeinanderfolgender Perioden, unter der Annahme, dass keine Korrelation vorhanden ist (’delta_max_idtd’). Schätzung für 52 Lohngruppen.

Quantiles

100%	Max	0.1268	99.0%	0.1268
75%	Q3	0.0674	97.5%	0.1245
50%	Med	0.0402	95.0%	0.0980
25%	Q1	0.0279	90.0%	0.0934
0%	Min	0.0096	10.0%	0.0189
	Range	0.1172	5.0%	0.0169
	Q3-Q1	0.0395	2.5%	0.0163
	Mode	.	1.0%	0.0096

Abbildung 18: Percentile der Verteilung der halben Länge des Vertrauensintervalls des Veränderungs-Schätzers des Durchschnittslohnindex aufeinanderfolgender Perioden, unter der Annahme, dass keine Korrelation vorhanden ist (’delta_max_idtd’). Schätzung für 52 Lohngruppen.

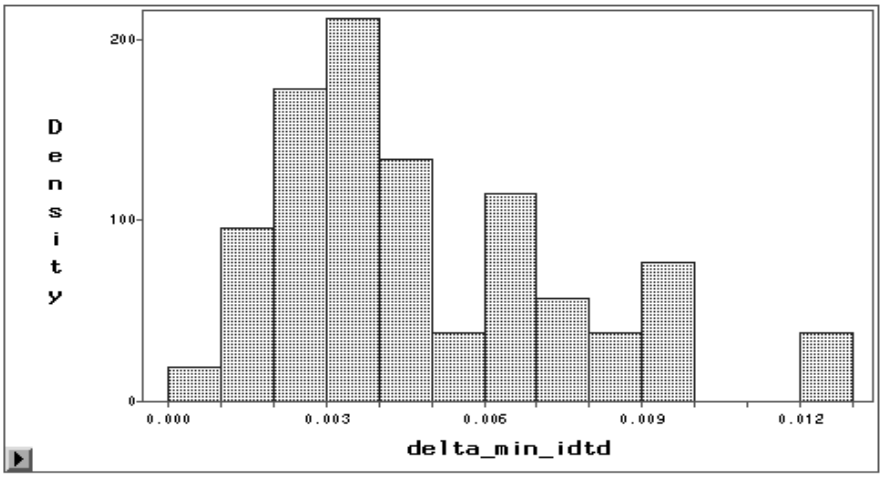


Abbildung 19: Verteilung der halben Länge des Vertrauensintervalls des Veränderungs-Schätzers des Durchschnittslohnindex aufeinanderfolgender Perioden, unter der Annahme, dass die Korrelation 0.99 beträgt (’delta_min_idtd’). Schätzung für 52 Lohngruppen.

Quantiles

100%	Max	0.0127	99.0%	0.0127
75%	Q3	0.0067	97.5%	0.0125
50%	Med	0.0040	95.0%	0.0098
25%	Q1	0.0028	90.0%	0.0093
0%	Min	0.0010	10.0%	0.0019
	Range	0.0117	5.0%	0.0017
	Q3-Q1	0.0040	2.5%	0.0016
	Mode	.	1.0%	0.0010

Abbildung 20: Percentile der Verteilung der halben Länge des Vertrauensintervalls des Veränderungs-Schätzers des Durchschnittslohnindex aufeinanderfolgender Perioden, unter der Annahme, dass die Korrelation 0.99 beträgt (’delta_min_idtd’). Schätzung für 52 Lohngruppen.

Genauigkeit auf derzeitigem Publikations-Niveau

Aus den Tabellen 7, 8 und 9 geht hervor, dass selbst wenn wir gröbere Untersuchungsbereiche bilden, die Veränderung des Durchschnittslohnindex zwischen 1998 und 1999 nicht immer signifikant ist. Wir erachten dabei eine Veränderung mit einem Fehler erster Art α als signifikant, wenn die Veränderung grösser ist als

die halbe Länge des Vertrauensintervalls der Veränderung des Durchschnittslohnindex, d.h. $\Delta^D_{d,99} > \hat{\delta}^D_{d,99}$. Mit anderen Worten, die Veränderung ist signifikant, wenn das 95%-Vertrauensintervall der Veränderung entweder $VI(\Delta^D_{d,99}) \subset \mathbb{R}_+$ oder $VI(\Delta^D_{d,99}) \subset \mathbb{R}_-$ ist. Aufgrund der Definition von $\hat{\delta}^D_{d,99}$ in (148) ist damit auch gesagt, dass die Signifikanz der Veränderung nicht nur von der Grösse der Veränderung abhängt, sondern auch von der Varianz des Durchschnittslohnindex im Untersuchungsbereich d und der Stichprobenkorrelation von zwei aufeinanderfolgenden Durchschnittslohnindices in d. Das heisst, wir können nicht von vornherein davon ausgehen, dass eine Veränderung signifikant ist, wenn sie relativ gross ausfällt (z.B. 1.1 Index-Punkte), wie dies aus der Tabelle 8, NOGA28=26 hervorgeht. Wir haben in den Tabellen 7, 8 und 9 aufgrund der publizierten Durchschnittslohnindices die Veränderung berechnet, womit Abweichungen von 0.1 Punkten zu den veröffentlichten Veränderungen resultieren können. $\hat{\delta}^D_{d,99}$ und $VI(\Delta^D_{d,99})$ wurden mit einem Fehler erster Art von 5% und 10% aufgeführt. Wir schlagen vor, Veränderungen, die bei $\alpha = 5\%$ signifikant sind, ohne Einschränkung zu publizieren. Die restlichen Veränderungen, die bei $\alpha = 10\%$ signifikant sind, sollten hingegen in Klammern publiziert werden, da ihre Qualität zu wünschen übrig lässt. Schlussendlich sollten die Veränderungen, die auch bei $\alpha = 10\%$ nicht signifikant sind, nicht publiziert werden. Mit dieser Regel würden 3 Veränderungen der Tabelle 8 und die Veränderungen auf dem Niveau Sektor x Geschlecht der Tabelle 9 in Klammern publiziert.
In den Tabellen 7 bis 9 haben wir auch den Variationskoeffizienten des Durchschnittslohnindex aufgeführt. Wir überlassen es dem interessierten Leser festzustellen, dass die entsprechenden Vertrauensintervalle des Durchschnittslohnindex grösser sind, als die Intervalle die auf dem Schätzer der Veränderung basieren mit $\alpha = 5\%$. Der Grund dafür ist, dass der Schätzer der Veränderung die zusätzliche Information der vorhergegangenen Periode mit einbezieht und vor allem die Stichprobenkorrelation zweier aufeinanderfolgender Lohnindices im entsprechenden Untersuchungsbereich berücksichtigt. Deshalb ist dieser Schätzer genauer sofern die Stichprobenkorrelation geschätzt werden kann. Wir sehen uns nicht in der Lage, diese analytisch zu beschreiben, da die Kombination von drei Erhebungen eine zu komplexe Ausgangslage darstellt, um eine Stichprobenkorrelation über mehrere Perioden zu schätzen. Wir sehen einzig mit *Jackknife*-Methoden die Möglichkeit die Stichprobenkorrelation schätzen zu können. Dies erfordert jedoch einen zusätzlichen substantiellen Aufwand, den innerhalb dieses Mandats nicht erbringen können. Zur Veranschaulichung haben wir die Stichprobenkorrelation durch die Korrelation $\hat{\rho}$, die von der unterliegenden tatsächlichen Zeitreihe stammt, ersetzt. Wie aus den oben genannten Tabellen ersichtlich ist, schwankt die Korrelation zwischen 0.6753 (Tabelle 8, NOGA28=60-63) und 0.9913 (Tabelle 7 Abschnitt="G, H", Geschlecht="M", resp. Tabelle 8 NOGA28=70-74). In

den meisten Fällen ist die Korrelation grösser als 0.9. Damit wir die Grösse der entsprechenden Vertrauensintervalle nicht unterschätzen, das heisst, dass die Qualität der Schätzer überschätzt wird, muss die Stichprobenkorrelation grösser sein als die Korrelation, die von der beobachteten Zeitreihe stammt: $\hat{\rho}_s > \hat{\rho}$. Aufgrund der grossen Werte von $\hat{\rho}$ haben wir Zweifel daran, dass wir die Qualität mehrerer Schätzer mit der Annahme $\hat{\rho}_s = \hat{\rho}$ nicht überschätzen.
Wenn wir bedenken, dass wir für die Schätzung der Vertrauensintervalle eine untere Grenze des Varianzschätzers des Durchschnittslohnindex verwendet haben, müssen wir damit rechnen, dass eine exakte Schätzung der Varianz dazu führt, dass in der Tabelle 7 von den 14 geschätzten Veränderungen (Veränderungen der Totale ausgenommen) nur noch rund die Hälfte mit $\alpha = 5\%$ signifikant ist (in der Tabelle sind es 9 von 14). Dabei ist eine Veränderung richtigerweise nicht signifikant, da die publizierte Veränderung 0.0 beträgt. In der Tabelle 8 müssen realistischerweise von den 26 geschätzten Veränderungen deren 6 mit $\alpha = 5\%$ als signifikant betrachtet werden (in der Tabelle sind es 9). Das heisst, rund zwei drittel der Veränderungen sind nicht signifikant. Für die zugehörigen Durchschnittslohnindices hätte 1999 nicht eindeutig gesagt werden können, ob sie sich verändert haben oder nicht. Unter diesen Voraussetzungen kann der Durchschnittslohnindex auf diesem Niveau nicht in der aktuellen Form publiziert werden.
Bei rund einem drittel der Indices auf dem Niveau Geschlecht x Wirtschaftsabschnitte, kann nicht gesagt werden, ob sich der Index zwischen 1998 und 1999 verändert hat oder nicht, da der Schätzer der Veränderung bei $\alpha = 5\%$ nicht signifikant ist. Aus der Tabelle 9 ist zu sehen, dass die Veränderung von -0.2 Index-Punkten für den Sektor=2 und das Geschlecht="M" zu klein ist, um uneingeschränkt als signifikant angesehen zu werden. Hingegen kann diese Veränderung mit $\alpha = 10\%$ als signifikant angesehen werden, und dementsprechend in Klammern publiziert werden. Die Veränderung von 0.1 Punkten im gesamten Sektor 2 ist jedoch auch mit $\alpha = 10\%$ nicht signifikant. Ansonsten waren 1999 die Veränderungen auf dem Niveau Sektor x Geschlecht genügend gross, um selbst mit $\alpha = 5\%$ von signifikanten Veränderungen zu sprechen. Bei all diesen Resultaten darf man nicht vergessen, dass wir sicher sein können, die Varianzen unterschätzt zu haben, da die Grenzgänger, etc. geschätzt werden müssen und da wir eine untere Grenze des Varianzschätzers verwendet haben. Somit können wir nicht garantieren, dass die Veränderung von -0.2 Index-Punkten für den Sektor=2 und das Geschlecht="M" bei $\alpha = 10\%$ signifikant ist.
Aus der Tabelle 7 ist ersichtlich, dass auch der Totale Durchschnittslohnindex eine grössere Länge des Vertrauensintervalls aufweist, als in den aktuellen Publikationen verwendet wird. Wenn die Varianz des Durchschnittslohnindex nicht verbessert werden kann, sollte die Form der Publikation angepasst werden. Dabei sollte die Varianzschätzung des Durchschnittslohnindex mit einbezogen werden. Wir schlagen vor, die Lohnindices der Untersuchungsbereiche aufgrund von

$cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ mit $\alpha = 5\%$ zu klassieren. Zum Beispiel könnten drei Klassen der Lohnindices wie folgt gebildet werden: $C_1 = \{d|cv(\hat{I}^D_{d,99,93,\hat{\pi}}) \in [0, 1.0]\}$, $C_2 = \{d|cv(\hat{I}^D_{d,99,93,\hat{\pi}}) \in (1.0, 2.0]\}$ und $C_3 = \{d|cv(\hat{I}^D_{d,99,93,\hat{\pi}}) \in (2.0, \infty)\}$. Die Indices der Klasse C_1 könnten zum Beispiel mit zwei Sternen $(**)$ publiziert werden, die Indices der Klasse C_2 mit einem Stern $(*)$ und die übrigen ohne Sterne.

Für den Vergleich von zwei Durchschnittslohnindices der gleichen Periode in den Untersuchungsbereichen d_1 und d_2 ist aufgrund von (86) und der obigen Überlegungen ein signifikanter Unterschied mit einem Fehler erster Art von 5% vorhanden, wenn der Unterschied der beiden Durchschnittslohnindices grösser als die halbe Länge des Vertrauensintervalls des Schätzers des Unterschieds ist. Das heisst

$$\begin{aligned} |\Lambda^D_{d_1,d_2,t}| &> 1.96\sqrt{(\hat{\sigma}^D_{d_1,t})^2 + (\hat{\sigma}^D_{d_2,t})^2} \\ &\doteq \hat{\lambda}^D_{d_1,d_2,t} \end{aligned} \tag{151}$$

Aus der Tabelle 17 im Anhang G ist ersichtlich, dass auf Niveau Abschnitt x Geschlecht 44 Vergleiche mit einem Fehler erster Art von 5% signifikant sind. Diese Unterschiede können ohne Einschränkung veröffentlicht werden. Weitere 4 Unterschiede müssten mit einem Hinweis auf die eingeschränkte Qualität publiziert werden. Von den 153 möglichen Kombinationen konnten aufgrund von fehlenden Indexwerten nur deren 92 berechnet werden. Von diesen verbleiben somit 44 die keine signifikanten Unterschiede aufweisen. Bei der Publikation sollte darauf geachtet werden, dass diese nicht als Beispiele für die Veranschaulichung einer unterschiedlichen Indexhöhe verwendet werden.
Für die weiteren Untersuchungsbereiche, die derzeit publiziert werden, haben wir aus Platzgründen darauf verzichtet, die Tabellen mit den Kombinationen der Vergleiche in diesem Methodenbericht aufzuführen.

Tabelle 7: Signifikanztest der Veränderung des Durchschnittslohnindex von 1998 zu 1999 nach Abschnitt und Geschlecht (M=Männer, F=Frauen) mit einem Fehler erster Art $\alpha = 5\%$ und $\alpha = 10\%$. Wenn "$\bigcap = 1$", dann ist die Veränderung mit einem Fehler erster Art α signifikant. Dieser Test basiert auf dem Vertrauensintervall der Veränderung $VI(\Delta^D_{d,99})$. Der Varianzschätzer von $\Delta^D_{d,99}$ basiert auf der unteren Grenze des exakten Varianzschätzers des Durchschnittslohnindex. Die Stichprobenkorrelation wurde durch die beobachtete Korrelation $\hat{\rho}$ zwischen zwei aufeinanderfolgenden Durchschnittslohnindices von 1993 bis 2003 ersetzt. $\hat{\delta}^D_{d,99}$ bezeichnet die halbe Distanz des Vertrauensintervalls bei oben erwähnter Korrelation und aufgrund des Fehlers erster Art α. $cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ ist der Schätzer für den Variationskoeffizienten der Durchschnittslohnindices. $VI(\hat{I}^D_{d,99,93,\hat{\pi}})$ bezeichnet das Vertrauensintervall des Durchschnittslohnindex mit $\alpha = 5\%$. In der untersten Linie steht der Totalschätzer. Ein fehlender Wert in der Tabelle bedeutet, dass dieser nicht geschätzt werden konnte.

Abschnitt	**Geschlecht**	$\hat{I}^D_{d,99,93,\hat{\pi}}$	$\Delta^D_{d,99}$	$cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ [%]	$VI(\hat{I}^D_{d,99,93,\hat{\pi}})$ 1999	$\hat{\rho}$	$\hat{\delta}^D_{d,99}$ $\alpha = 5\%$	$VI(\Delta^D_{d,99})$ $\alpha = 5\%$	$\hat{\delta}^D_{d,99}$ $\alpha = 10\%$	$VI(\Delta^D_{d,99})$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
A	M	106.1	0.2			0.9686						
A	F											
C, E	M	105.4	0.2	1.87	[101.5 , 109.3]	0.9362	1.38	[-1.18 , 1.58]	1.16	[-0.96 , 1.36]	0	0
C, E	F											
D	M	104.7	0	0.44	[103.8 , 105.6]	0.9723	0.21	[-0.21 , 0.21]	0.18	[-0.18 , 0.18]	0	0
D	F	107.7	1.2	0.71	[106.2 , 109.2]	0.9841	0.27	[0.93 , 1.47]	0.22	[0.98 , 1.42]	1	1
F	M	104.5	-0.5	0.95	[102.6 , 106.4]	0.9497	0.62	[-1.12 , 0.12]	0.52	[-1.02 , 0.02]	0	0
F	F											
G, H	M	106	1.4	0.75	[104.4 , 107.6]	0.9913	0.21	[1.19 , 1.61]	0.17	[1.23 , 1.57]	1	1
G, H	F	105.7	0.1	0.94	[103.8 , 107.6]	0.9626	0.53	[-0.43 , 0.63]	0.45	[-0.35 , 0.55]	0	0
I	M	104.9	-0.5	0.45	[104.0 , 105.8]	0.8625	0.48	[-0.98 , -0.02]	0.41	[-0.91 , -0.09]	1	1
I	F	101.3	-0.5	0.79	[99.7 , 102.9]	0.8502	0.86	[-1.36 , 0.36]	0.72	[-1.22 , 0.22]	0	0
J, K	M	108.3	1.2	0.52	[107.2 , 109.4]	0.9893	0.16	[1.04 , 1.36]	0.14	[1.06 , 1.34]	1	1
J, K	F	111.2	2	0.51	[110.1 , 112.3]	0.9893	0.16	[1.84 , 2.16]	0.14	[1.86 , 2.14]	1	1
L	M	103.7	-0.7	0.27	[103.2 , 104.2]	0.9156	0.22	[-0.92 , -0.48]	0.19	[-0.89 , -0.51]	1	1
L	F	104.6	-0.6	0.33	[103.9 , 105.3]	0.8973	0.30	[-0.90 , -0.30]	0.26	[-0.86 , -0.34]	1	1
M,N,O	M	104.8	-0.3	0.48	[103.8 , 105.8]	0.9671	0.25	[-0.55 , -0.05]	0.21	[-0.51 , -0.09]	1	1
M,N,O	F	105.2	0.8	0.47	[104.2 , 106.2]	0.9655	0.26	[0.54 , 1.06]	0.21	[0.59 , 1.01]	1	1
Total	M	105.2	0.1	0.22	[104.7 , 105.7]	0.9790	0.09	[0.01 , 0.19]	0.08	[0.02 , 0.18]	1	1
Total	F	106.5	0.7	0.29	[105.9 , 107.1]	0.9854	0.10	[0.60 , 0.80]	0.09	[0.61 , 0.79]	1	1
Total		105.6	0.3	0.17	[105.3 , 105.9]	0.9820	0.07	[0.23 , 0.37]	0.05	[0.25 , 0.35]	1	1

Tabelle 8: Signifikanztest der Veränderung des Durchschnittslohnindex von 1998 zu 1999 nach NOGA28 mit einem Fehler erster Art $\alpha = 5\%$ und $\alpha = 10\%$. Wenn "$\bigcap = 1$", dann ist die Veränderung mit einem Fehler erster Art α signifikant. Dieser Test basiert auf dem Vertrauensintervall der Veränderung $VI(\Delta^D_{d,99})$. Der Varianzschätzer von $\Delta^D_{d,99}$ basiert auf der unteren Grenze des exakten Varianzschätzers des Durchschnittslohnindex. Die Stichprobenkorrelation wurde durch die beobachtete Korrelation $\hat{\rho}$ zwischen zwei aufeinanderfolgenden Durchschnittslohnindices von 1993 bis 2003 ersetzt. $\hat{\delta}^D_{d,99}$ bezeichnet die halbe Distanz des Vertrauensintervalls bei oben erwähnter Korrelation und aufgrund des Fehlers erster Art α. $cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ ist der Schätzer für den Variationskoeffizienten der Durchschnittslohnindices. $VI(\hat{I}^D_{d,99,93,\hat{\pi}})$ bezeichnet das Vertrauensintervall des Durchschnittslohnindex mit $\alpha = 5\%$. In der untersten Linie steht der Totalschätzer. Ein fehlender Wert in der Tabelle bedeutet, dass dieser nicht geschätzt werden konnte.

NOGA28	$\hat{I}^D_{d,99,93,\hat{\pi}}$	$\Delta^D_{d,99}$	$cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ [%]	$VI(\hat{I}^D_{d,99,93,\hat{\pi}})$ 1999	$\hat{\rho}$ $\alpha = 5\%$	$\hat{\delta}^D_{d,99}$ $\alpha = 5\%$	$VI(\Delta^D_{d,99})$	$\hat{\delta}^D_{d,99}$ $\alpha = 10\%$	$VI(\Delta^D_{d,99})$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
1	106.9	-0.1			0.9751						
2	105.0	0.0			0.9016						
10-14/40-41	105.4	-0.1	1.67	[101.9 , 108.9]	0.9550	1.04	[-1.14 , 0.94]	0.87	[-0.97 , 0.77]	0	0
15-16	104.5	0.7	1.20	[102.1 , 106.9]	0.9871	0.39	[0.31 , 1.09]	0.33	[0.37 , 1.03]	1	1
17-18	104.3	-0.9	1.99	[100.2 , 108.4]	0.8625	2.14	[-3.04 , 1.24]	1.79	[-2.69 , 0.89]	0	0
20	105.4	0.1	2.52	[100.2 , 110.6]	0.9623	1.43	[-1.33 , 1.53]	1.20	[-1.10 , 1.30]	0	0
21-22	103.7	0.4	0.86	[102.0 , 105.4]	0.9522	0.54	[-0.14 , 0.94]	0.45	[-0.05 , 0.85]	0	0
19,23-25	108.4	-0.4	0.75	[106.8 , 110.0]	0.9629	0.43	[-0.83 , 0.03]	0.36	[-0.76 , -0.04]	0	1
26	105.1	1.1	1.73	[101.5 , 108.7]	0.9171	1.45	[-0.35 , 2.55]	1.22	[-0.12 , 2.32]	0	0
27-28	104.5	0.6	1.53	[101.4 , 107.6]	0.9613	0.87	[-0.27 , 1.47]	0.73	[-0.13 , 1.33]	0	0
29,34-35	104.1	0.2	0.61	[102.9 , 105.3]	0.9677	0.32	[-0.12 , 0.52]	0.27	[-0.07 , 0.47]	0	0
30-33	105.6	0.3	0.81	[103.9 , 107.3]	0.9646	0.45	[-0.15 , 0.75]	0.38	[-0.08 , 0.68]	0	0
36-37	106.2	0.9	2.00	[102.0 , 110.4]	0.9756	0.92	[-0.02 , 1.82]	0.77	[0.13 , 1.67]	0	1
45	104.6	-0.5	0.83	[102.9 , 106.3]	0.9482	0.55	[-1.05 , 0.05]	0.46	[-0.96 , -0.04]	0	1
50	106.1	0.2	1.82	[102.3 , 109.9]	0.9578	1.10	[-0.90 , 1.30]	0.92	[-0.72 , 1.12]	0	0
51	106.6	1.5	0.72	[105.1 , 108.1]	0.9817	0.29	[1.21 , 1.79]	0.24	[1.26 , 1.74]	1	1
52	105.7	-0.2	0.80	[104.0 , 107.4]	0.9663	0.43	[-0.63 , 0.23]	0.36	[-0.56 , 0.16]	0	0
55	105.2	2.0	2.37	[100.3 , 110.1]	0.9428	1.65	[0.35 , 3.65]	1.39	[0.61 , 3.39]	1	1
60-63	101.8	-0.2	0.54	[100.7 , 102.9]	0.6753	0.86	[-1.06 , 0.66]	0.72	[-0.92 , 0.52]	0	0
64	108.6	-1.0	0.54	[107.4 , 109.8]	0.9227	0.45	[-1.45 , -0.55]	0.38	[-1.38 , -0.62]	1	1
65	111.8	1.6	0.41	[110.9 , 112.7]	0.9827	0.17	[1.43 , 1.77]	0.14	[1.46 , 1.74]	1	1
66	112.7	0.0	0.68	[111.2 , 114.2]	0.9556	0.45	[-0.45 , 0.45]	0.38	[-0.38 , 0.38]	0	0
67	112.3	0.8	1.15	[109.8 , 114.8]	0.9749	0.57	[0.23 , 1.37]	0.48	[0.32 , 1.28]	1	1
70-74	106.2	2.0	0.61	[104.9 , 107.5]	0.9913	0.17	[1.83 , 2.17]	0.14	[1.86 , 2.14]	1	1
75	103.9	-0.7	0.20	[103.5 , 104.3]	0.9132	0.17	[-0.87 , -0.53]	0.14	[-0.84 , -0.56]	1	1
80	104.2	1.0	0.50	[103.2 , 105.2]	0.9570	0.30	[0.70 , 1.30]	0.25	[0.75 , 1.25]	1	1
85	105.3	-0.2	0.52	[104.2 , 106.4]	0.9669	0.28	[-0.48 , 0.08]	0.23	[-0.43 , 0.03]	0	0
90-93	105.1	0.3	0.73	[103.6 , 106.6]	0.9664	0.39	[-0.09 , 0.69]	0.33	[-0.03 , 0.63]	0	0

Tabelle 9: Signifikanztest der Veränderung des Durchschnittslohnindex von 1998 zu 1999 nach Sektor und Geschlecht (M=Männer, F=Frauen) mit einem Fehler erster Art $\alpha = 5\%$ und $\alpha = 10\%$. Wenn "$\bigcap = 1$", dann ist die Veränderung mit einem Fehler erster Art α signifikant. Dieser Test basiert auf dem Vertrauensintervall der Veränderung $VI(\Delta^D_{d,99})$. Der Varianzschätzer von $\Delta^D_{d,99}$ basiert auf der unteren Grenze des exakten Varianzschätzers des Durchschnittslohnindex. Die Stichprobenkorrelation wurde durch die beobachtete Korrelation $\hat{\rho}$ zwischen zwei aufeinanderfolgenden Durchschnittslohnindices von 1993 bis 2003 ersetzt. $\hat{\delta}^D_{d,99}$ bezeichnet die halbe Distanz des Vertrauensintervalls bei oben erwähnter Korrelation und aufgrund des Fehlers erster Art α. $cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ ist der Schätzer für den Variationskoeffizienten der Durchschnittslohnindices. $VI(\hat{I}^D_{d,99,93,\hat{\pi}})$ bezeichnet das Vertrauensintervall des Durchschnittslohnindex mit $\alpha = 5\%$. In der untersten Linie steht der Totalschätzer. Ein fehlender Wert in der Tabelle bedeutet, dass dieser nicht geschätzt werden konnte.

Sektor	**Geschlecht**	$\hat{I}^D_{d,99,93,\hat{\pi}}$	$\Delta^D_{d,99}$	$cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ [%]	$VI(\hat{I}^D_{d,99,93,\hat{\pi}})$	$\hat{\rho}$ $\alpha = 5\%$	$\hat{\delta}^D_{d,99}$ $\alpha = 5\%$	$VI(\hat{I}^D_{d,99,93,\hat{\pi}})$ $\alpha = 10\%$	$\hat{\delta}^D_{d,99}$ $\alpha = 10\%$	$VI(\Delta^D_{d,99})$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
1	M	106.1	0.2			0.9686						
1	F											
2	M	104.6	-0.2	0.41	[103.8 , 105.4]	0.9652	0.22	[104.4 , 104.8]	0.18	[-0.38 , -0.02]	0	1
2	F	107.6	1.1	0.62	[106.3 , 108.9]	0.9876	0.21	[107.4 , 107.8]	0.17	[0.93 , 1.27]	1	1
3	M	105.8	0.4	0.26	[105.3 , 106.3]	0.9872	0.09	[105.7 , 105.9]	0.07	[0.33 , 0.47]	1	1
3	F	106.2	0.6	0.31	[105.6 , 106.8]	0.9805	0.13	[106.1 , 106.3]	0.11	[0.49 , 0.71]	1	1
1	Total	106.5	1.4	0.75		0.9731						
2	Total	105.0	0.1	0.94	[103.1 , 106.9]	0.9717	0.16	[104.8 , 105.2]	0.14	[-0.14 , 0.14]	0	0
3	Total	106.0	-0.5	0.45	[105.1 , 106.9]	0.9874	0.06	[105.9 , 106.1]	0.05	[0.45 , 0.55]	1	1
Total	M	105.2	0.1	0.22	[104.7 , 105.7]	0.9790	0.09	[105.1 , 105.3]	0.08	[0.02 , 0.18]	1	1
Total	F	106.5	0.7	0.29	[105.9 , 107.1]	0.9854	0.10	[106.4 , 106.6]	0.09	[0.61 , 0.79]	1	1
Total		105.6	0.3	0.17	[105.3 , 105.9]	0.9820	0.07	[105.5 , 105.7]	0.05	[0.25 , 0.35]	1	1

10 Zusammenfassung

Das Ziel des vorliegenden Methodenberichts war es, die Lohnindices Schätzung zu formalisieren und die Genauigkeit der Schätzer zu schätzen. Dabei ging es einerseits darum, den bestehenden Lohnindex, den wir Durchschnittslohnindex nennen, formal so zu beschreiben, dass seine Schätzung jederzeit nachvollziehbar ist. Und andererseits haben wir einen neuen Lohnindex eingeführt, den wir Lohnmassenindex nennen. Dieser ist, wie der Name es bereits andeutet, ein Index der die Entwicklung der gesamten Lohnmasse beschreiben soll. Wir haben die Formalisierung zuerst auf einer theoretischen Basis vorgenommen, wobei wir die aktuelle Schätz-Methode miteinbezogen haben, siehe Abschnitt 3. Dabei wurde klar, dass der Durchschnittslohnindex mit einer konstanten Gewichtung, die an die Verunfallten der SSUV gebunden ist, nur über alle Lohngruppen hinweg ein *Laspeyres*-Index ist. Dabei werden die Lohngruppen anhand der Variablen Wirtschaftsklasse (ASWZ), Geschlecht, Arbeitsbereich, Qualifikation, Lohnart, Unfallart gebildet. Sofern der Durchschnittslohnindex für Untersuchungsbereiche, wie zum Beispiel für die Lohnempfänger nach NOGA und Geschlecht, geschätzt werden soll, so muss die Gewichtung für jeden Untersuchungsbereich angepasst werden, siehe Abschnitt 3.2. Auf diese Weise kann ein *Laspeyres*-Index konstruiert werden, wie wir ihn in 3.1 aufgrund von [11] definiert haben. Die gleiche Gewichtung sollte auch für den Lohnmassenindex verwendet werden, wenn dieser für Untersuchungsbereiche geschätzt werden soll. Denn nur mit der angepassten Gewichtung kann die wirkliche Entwicklung der Lohnmasse verfolgt werden, wie aus 3.3 hervorgeht. Somit muss entschieden werden, ob ein *Laspeyres*-Index geschätzt werden soll, oder ob die Gewichtung konstant beibehalten werden soll, respektive, welche Grössen mit den Lohnindices verfolgt werden sollen.
Aufgrund der vorhandenen Daten, konnten wir unsere Schätzungen nur für den 2. und 3. Sektor durchführen. Eine entsprechende Lösung für den 1. Sektor ist nicht in Sicht. Wir werden im weiteren nicht mehr auf diesen Punkt eingehen, und werden immer nur vom 2. und 3. Sektor sprechen.
Bei der Schätzung des aktuellen Durchschnittslohnindex zeigte sich im Abschnitt 4.1, dass diese nur dann die gewünschte Entwicklung wiedergibt, wenn für die Referenzperiode t_0 der Durchschnittslohn der Verunfalltengruppe V_j dem Durchschnittslohn der Lohngruppe U_j entspricht. Ist diese Voraussetzung nicht gegeben, so müssen wir mit einem Bias rechnen. Eine kontrollierte Schätzung setzt notgedrungen voraus, dass die Daten der SSUV als Stichprobe angesehen werden. Bei diesem Ansatz wurde klar, dass die Einschlusswahrscheinlichkeit, sprich Unfallwahrscheinlichkeit, fehlt, um die Schätzung des Durchschnittslohns korrekt durchführen zu können. Insbesondere fehlen die Daten um die Populationsgrösse auf dem Niveau der Lohngruppen zu schätzen. Der Versuch die Einschlusswahrscheinlichkeit aufgrund der SAKE99 zu schätzen (Abschnitt 4.2.2, in welcher eine

Frage bezüglich der Verunfallung gestellt worden war, schlug fehl. Die geschätzte Einschlusswahrscheinlichkeit lag im Durchschnitt 6% unter der Unfallwahrscheinlichkeit, welche anhand der Daten der SSUV geschätzt werden konnte. Es konnte zudem gezeigt werden, dass der korrigierte Schätzer der Einschlusswahrscheinlichkeit von 1999 nicht für das Jahr 1998 angewandt werden kann. Mit anderen Worten, haben wir bisher keinerlei Daten zur Verfügung, um die Einschlusswahrscheinlichkeit mit genügender Genauigkeit schätzen zu können.
Unter der Annahme, dass die Daten der SSUV einer Stichprobe gleich gesetzt werden können, und dass die Einschlusswahrscheinlichkeit geschätzt wird, wurde die Lohnmasse, der Durchschnittslohn, der Durchschnittslohnindex, der Lohnmassenindex sowie die zugehörigen Varianzschätzer im Abschnitt 4.2geschätzt. Die Varianzschätzer mussten auf Approximationen abgestützt werden.
Wir konnten im Abschnitt 4.3 zeigen, dass der Bias in der Schätzung der Populationsgrösse keinen Einfluss auf den Schätzer des Durchschnittslohnindex auf Niveau Lohngruppe hat. Hingegen weisen die Schätzer der Durchschnittslohnindices auf allen anderen Niveaus, sowie der Lohnmassenindices auf allen Niveaus und sämtliche Varianzschätzer aller Schätzer einen Bias auf. Deshalb ist es unabdingbar entweder die Unfallwahrscheinlichkeit für jede Lohngruppe ohne Bias schätzen zu können, oder aber die Populationsgrösse für jede Lohngruppe ohne Bias schätzen zu können.

Der Versuch die fehlende Einschlusswahrscheinlichkeit dadurch zu ersetzen, dass eine Modell-Methode verwendet wird, schlug im Abschnitt 5fehl. Eine dieser Methoden war eine Regressions-Methode. Bei dieser muss anhand dem Selektions-Bias der Tatsache Rechnung getragen werden, dass die Daten nicht aus einer Stichprobe stammen. Hierfür brauchen wir für die gesamte Population der Lohnempfänger eine Indikatorvariable, die besagt, ob ein Lohnempfänger einen Unfall hatte oder nicht. Auf diese Weise sind wir wieder auf das Problem der Unfallwahrscheinlichkeit und der Populationsgrösse gestossen. Ebenso stossen wir bei der Kalibrierungs-Methode wieder auf dieses Problem. Mit dieser Methode wird eine Harmonisierung der SLI mit der LSE angestrebt. Dabei wird die geschätzte Lohnmasse aus den Daten der SSUV mit der Lohnmasse, die aus der LSE geschätzt wird, kalibriert. Für die Schätzung der Lohnmasse fehlt jedoch die Einschlusswahrscheinlichkeit, respektive die Populationsgrösse. Zudem sind die Populationen der beiden Erhebungen unterschiedlich, was eine zusätzliche Harmonisierung voraussetzt. Ein gewichtiger Nachteil dieser Methode ist, dass die Schätzer der Durchschnittslohnindices keine *Laspeyres*-Preisindices sind. Somit ist von dieser Methode abzusehen, da die entsprechenden Lohnindices, die von den EU-Ländern publiziert werden, *Laspeyres*-Preisindices sind.
In einem weiteren Abschnitt, 7, haben wir die Homogenität der Lohngruppen untersucht und formal beschrieben, wie die Lohngruppen zurzeit gebildet werden.

Zudem haben wir aufgezeigt, dass die Definition von Homogenität von der angewandten Methode abhängt. Eine objektive Definition der Homogenität existiert zur Zeit nicht. Dies erschwert es, klare Aussagen zu machen. Die aktuellen homogenen Gruppen wurden durch die Kreuzung der Variablen Wirtschaftsklasse (ASWZ), Geschlecht, Arbeitsbereich, Qualifikation, Lohnart und Unfallart gebildet. Die Bildung dieser homogenen Gruppen hatte zum Ziel, Lohnempfänger mit dem gleichen Lohnniveau zu gruppieren und somit Lohngruppen zu bilden, die bezüglich dem Lohn homogen sind. Durch eine einfache beschreibende Statistik haben wir die aktuellen homogenen Gruppen auf die Variabilität der Löhne untersucht. Dabei sind uns Zweifel aufgekommen, ob die derzeitigen Lohngruppen dem Kriterium der Homogenität entsprechen. Für die Bildung der homogenen Lohngruppen schlagen wir die Anwendung von Klassifikationsbäumen vor. Diese haben den Vorteil, dass die Bildung der Gruppen im Hinblick auf die Homogenität in den Gruppen durchgeführt werden können.

Um die bestehende Serie so gut wie möglich weiterführen zu können, und da keine weiterreichenden Daten vorhanden sind, hat LOHN eine Methode gewählt, die der bisherigen Methode entspricht, wobei die Bildung der Lohngruppen anhand von neuen Variablen durchgeführt werden. Wir haben diese Methode mit den neu verwendeten Daten und Variablen im Abschnitt 8 formal beschrieben und die Qualität der Indices untersucht. Dabei hat sich gezeigt, dass die Kombination von 3 Erhebungen (BESTA, SAKE, SSUV) zum Teil unlösbare Probleme in der Schätzung der Genauigkeit aufwirft. Nur schon die Schätzung der Population in den Lohngruppen basiert auf einem verzerrten, das heisst, nicht erwartungstreuen Schätzer. Sämtliche Varianzen konnten nur anhand von *Taylor*-Linearisierungen approximiert werden. Eines der Hauptprobleme liegt in der unterschiedlichen Definition der Population der drei Erhebungen. Hinzu kommt, dass die drei Erhebungen nicht vollständig harmonisiert werden können. Zum Beispiel fehlt ein Schätzer der Anzahl Grenzgänger, Saisonniers und Kurzaufenthalter auf dem Niveau NOGA x Geschlecht x Beschäftigungsgrad. Dieses Problem wird auch in Zukunft bestehen bleiben, da im neuen Fragebogen der BESTA die Anzahl dieser Lohnempfänger nach Geschlecht gefragt wird, nicht aber der Beschäftigungsgrad. Unseres Wissens gibt es zurzeit keine andere Erhebung, die diese Lohnempfänger explizit befragt. Für unsere Schätzungen haben wir 180’000 Grenzgänger, etc. zufällig in der SAKE generiert, wobei sie proportional zum Unterschied der Schätzer der Beschäftigten aus der SAKE und der BESTA generiert wurden. Wir sind uns der Fragwürdigkeit dieser Methode bewusst. Das Ziel war jedoch, die Schätzmethoden insofern zu testen, dass sie auch die Grenzgänger, etc. mit einbeziehen können. Da wir davon ausgegangen sind, dass diese Lohnempfänger aus dem ZAR oder einem anderen Register stammen werden, und somit keine Varianz mit sich bringen, führt der Einbezug der 180’000 generierten Lohnempfänger zu einer Un-

terschätzung der Varianz und ebenso zu einer Unterschätzung des Variationskoeffizienten. Somit sind die Resultate im Abschnitt 9 zu optimistisch.
Im Abschnitt 9.2 hat sich gezeigt, dass die Schätzung der Anzahl Lohnempfänger in beinahe der Hälfte der 102 Untersuchungsbereiche des 2. und 3. Sektors, gebildet durch die NOGA, das Geschlecht und den Beschäftigungsgrad, genügend genau ist ($cv \leq 5\%$). Hingegen sind in 6 Untersuchungsbereichen die Schätzungen derart schlecht ($cv \geq 50\%$), dass keine Aussage mehr möglich ist. Für die Schätzung der Lohnempfänger in den Lohngruppen, die durch die Kreuzung der Variablen NOGA, Geschlecht, Beschäftigungsgrad, Altersklassen und den ISCO-Code gebildet wurden, haben wir festgestellt, dass rund die Hälfte der Lohngruppen nicht existieren oder nicht gebildet werden können. Wir konnten in 1013 Lohngruppen Schätzungen vornehmen, auch wenn diese zum Teil nur auf einer Beobachtung in der SAKE beruhten. Auf diesem Niveau weisen 10% der 1013 Lohngruppen, eine genügende Genauigkeit des Schätzers der Anzahl Lohnempfänger auf ($cv \leq 5\%$). In mehr als 8% der Lohngruppen kann keine Aussage über die Anzahl Lohnempfänger gemacht werden, da der Variationskoeffizient mehr als 50% beträgt. Mehr als die Hälfte der Lohngruppen weist einen Variationskoeffizienten von über 14% auf, wodurch die Qualität der Schätzer ungenügend ist.
Wir konnten eine obere Grenze für den Biasquotienten des Schätzers der Anzahl Lohnempfänger in den Lohngruppen schätzen. Dabei stellten wir fest, dass in einer Mehrheit der 1013 Lohngruppen der Bias vernachlässigbar ist. Hingegen müssen wir in mehr als einem drittel der Lohngruppen mit einer zum Teil massiven Verzerrung rechnen.
Die darauf folgende Analyse der Genauigkeit der Schätzer der Lohnindices im Abschnitt 9.4.1 hat gezeigt, dass der Varianzschätzer des Schätzers des Lohnmassenindex durch den Variationskoeffizienten des Lohnmassen Schätzers begrenzt ist. Für diesen Schätzer konnten wir keine untere Grenze schätzen, die verschieden von 0 ist. Unsere Methode führte zu einer unteren Grenze, anhand derer wir die Unfallwahrscheinlichkeit klar unterschätzen. Zur Veranschaulichung der zu erwartenden Genauigkeit und Fluktuationen des Lohnmassenindex haben wir einen 'Pseudo-Lohnmassenindex' aus der Anzahl der Beschäftigten aus der BESTA und dem Schätzer des Durchschnittslohns aus der SSUV kreiert. Wenn der Varianzschätzer des realen Lohnmassenindex die gleiche Grössenordnung aufweist, wie der von uns konstruierte 'Pseudo-Lohnmassenindex', dann muss damit gerechnet werden, dass über mehrere Quartale hinweg kein signifikant unterschiedliches Niveau des Lohnmassenindex veröffentlicht werden kann. Um die Varianz des Lohnmassenindex schätzen zu können, muss die Stichprobenkorrelation des Lohnmassenschätzers der Referenzperiode und des Lohnmassenschätzers der jeweiligen Periode t geschätzt werden. Diese Schätzung ist im vorliegenden Methodenbericht nicht aufgeführt, da dies den Rahmen der verfügbaren Ressourcen gesprengt hätte.

Aufgrund der Abbildungen 12 bis 14 ist ersichtlich, dass für die Interpretationen der Fluktuationen des Lohnmassenindex ein Handlungsbedarf von Seiten der Fachleute besteht, da sie einerseits aus der Fluktuation der Anzahl Lohnempfänger und andererseits aus der Fluktuation des Durchschnittslohns besteht. Jede einzelne dieser Fluktuationskomponenten ist nicht immer eindeutig interpretierbar und somit ist die Kombination der beiden nicht ohne weiteres zu interpretieren. Es stellt sich die Frage, ob Fluktuationen der Unfallwahrscheinlichkeit nicht in die Schätzungen miteinbezogen werden müssten, da sich diese bei einer Korrelation mit dem Lohnniveau im Lohnmassenindex und im Durchschnittslohnindex niederschlagen. So haben wir eine Periodizität des Durchschnittslohns der SSUV im 1. und 2. Quartal festgestellt. Erstere zeigt einen Anstieg des Durchschnittslohns auf, der durch den ordentlichen Lohnanstieg erklärbar ist. Im 2. Quartal fällt der Durchschnittslohn jeweils wieder, was wir uns bis anhin nicht erklären können.
Im Gegensatz zum Lohnmassenindex konnten wir unter gewissen Annahmen eine untere Grenze für den Schätzer der Varianz des Durchschnittslohnindex schätzen (siehe Abschnitt 9.4.2). Dadurch konnten wir aufzeigen, dass der Durchschnittslohnindex der Vollzeitbeschäftigten auf dem Niveau NOGA x Geschlecht (x Beschäftigungsgrad=100%) den in den aktuellen Publikationen verwendeten Qualitätsansprüchen nicht genügt. Zum selben Resultat gelangten wir bei der Analyse der Qualität des Schätzers der Veränderung zweier aufeinanderfolgender Durchschnittslohnindices. Wobei wir auch in dieser Schätzung Annahmen gemacht haben, wie zum Beispiel, dass die Varianz des Durchschnittslohnindex auf oben genanntem Niveau für die Periode t die gleiche ist wir für die Periode $t-1$. Auf diesem Niveau kann eine Veränderung des Durchschnittslohnindex, die kleiner als 1% (1 Index-Punkt) ist, nicht mehr interpretiert werden. Somit macht ein Durchschnittslohnindex auf diesem Niveau keinen Sinn.
Die Analyse der Genauigkeit der Schätzer auf den derzeit veröffentlichten Niveaus der Vollzeitbeschäftigten (Total, nach Geschlecht, nach Sektor, nach Geschlecht x Sektor, nach Geschlecht x Wirtschaftsabschnitten, und nach NOGA28) hat für 1999 gezeigt, dass für die gröbsten Untersuchungsbereiche ab einer Veränderung von 0.2 Punkten im besten Fall damit gerechnet werden kann, dass diese Veränderung signifikant ist. Somit konnte auch für den Totalen Durchschnittslohnindex die vorgegebene Genauigkeit nicht erreicht werden, da in den aktuellen Publikationen Veränderungen ab 0.1 Punkten ausgewiesen werden. Gleiches gilt für die Durchschnittslohnindices auf dem Niveau Geschlecht und für den Sektor 3 auf Niveau Geschlecht. Hingegen kann bei rund einem drittel der Indices auf dem Niveau Geschlecht x Wirtschaftsabschnitte nicht gesagt werden, ob sich der Index zwischen 1998 und 1999 verändert hat oder nicht, da der Schätzer der Veränderung bei einem Fehler erster Art von 5% und 10% nicht signifikant ist. Bei der Erhöhung des Fehlers erster Art von 5% auf 10% kann bei drei zusätzlichen Untersuchungsbereichen auf dem Niveau NOGA28 von einer signifikanten Verän-

derung gesprochen werden. Ebenso wird die Veränderung des Durchschnittslohnindex der Männer im 2. Sektor signifikant, wenn $\alpha = 10\%$. Für die bereits bei $\alpha = 5\%$ signifikanten Veränderungen ändert sich durch den Anstieg auf $\alpha = 10\%$ nichts; auch die Länge des Vertrauensintervalls der Veränderung des Totals hat immer noch eine Länge von 0.2 Punkten, gegenüber den 0.1 erlaubten Punkten.
Wird die NOGA28 als Untersuchungsbereich festgelegt, kann selbst eine Veränderung von 1.1 Punkten nicht garantieren, dass diese auch signifikant ist (siehe Tabelle 8, NOGA28=26). Das heisst, alleine eine relativ grosse Veränderung kann nicht garantieren, dass diese signifikant ist. Denn der Schätzer der Varianz des Durchschnittslohnindex sowie die Stichprobenkorrelation haben einen Einfluss auf die Genauigkeit der Schätzung der Veränderung des Durchschnittslohnindex. Unter diesen Voraussetzungen kann der Durchschnittslohnindex mit den derzeitigen Qualitätsansprüchen nicht in der aktuellen Form auf dem Niveau NOGA28 publiziert werden.
Für die geschätzten Veränderungen schlagen wir vor, diese ohne weitere Kennzeichnung zu veröffentlichen, wenn sie bei $\alpha = 5\%$ signifikant sind. Falls dies nicht der Fall sein sollte, aber die Veränderung bei $\alpha = 10\%$ signifikant ist, so sollte diese in Klammern veröffentlicht werden. Ist auch bei $\alpha = 10\%$ keine Signifikanz auszumachen, so sollten für die entsprechenden Veränderungen nur Klammern ohne die Veränderung publiziert werden.
Das gleiche Prinzip sollte für Vergleiche zwischen zwei Durchschnittslohnindices der gleichen Periode aber in verschiedenen Untersuchungsbereichen verwendet werden. Ist der Unterschied bei $\alpha = 5\%$ signifikant, so kann dieser ohne Einschränkungen veröffentlicht werden. Bei einer Signifikanz bei $\alpha = 10\%$, sollte der Vergleich in Klammern und dementsprechend interpretiert werden. Bei keinem signifikanten Unterschied sollte auf eine Veröffentlichung und Interpretation verzichtet werden.
Zusätzlich schlagen wir ein Qualtiätsmerkmal für die Durchschnittslohnindices vor. Wir schlagen vor, die Lohnindices der Untersuchungsbereiche aufgrund von $cv(\hat{I}^D_{d,99,93,\hat{\pi}})$ mit $\alpha = 5\%$ zu klassieren. Zum Beispiel könnten drei Klassen der Lohnindices wie folgt gebildet werden: $C_1 = \{d|cv(\hat{I}^D_{d,99,93,\hat{\pi}}) \in [0, 1.0]\}$, $C_2 = \{d|cv(\hat{I}^D_{d,99,93,\hat{\pi}}) \in (1.0, 2.0]\}$ und $C_3 = \{d|cv(\hat{I}^D_{d,99,93,\hat{\pi}}) \in (2.0, \infty)\}$. Die Indices der Klasse C_1 könnten zum Beispiel mit zwei Sternen $(**)$ publiziert werden, die Indices der Klasse C_2 mit einem Stern $(*)$ und die übrigen ohne Sterne.

11 Schlussfolgerung

Die Erarbeitung von Varianz-Schätzern für die Lohnindices hat durch die Vorgaben, mehrere, nicht harmonisierte Datenquellen für die Schätzung zu verwenden, eine kaum mehr zu bewältigende Komplexität erreicht. Es stellt sich die Frage, ob

ein anderes Notationssystem das Problem zumindest teilweise lösen könnte. Auf dieser Ebene besteht ein Handlungsbedarf seitens der Methodologen.
Aufgrund des vorliegenden Berichts, müssen entweder die Ansprüche an die Lohnindices bezüglich Periodizität und Population gesenkt werden. Dann könnten andere Erhebungen, z.B. die LSE, für die Schätzung der Lohnindices verwendet werden, womit die Genauigkeit der Schätzer verbessert werden und der Bias, der durch die Kombination der BESTA und der SAKE entsteht, verhindert werden könnte. Die Methode für die Verwendung anderer Erhebungen wurde in diesem Methodenbericht nicht beschrieben.
Mit den derzeit angestrebten Zielen und verwendeten Daten können Lohnindices nur auf einem sehr groben Niveau für Vollzeitbeschäftigte mit einer befriedigenden Genauigkeit geschätzt werden. Hinzu kommt, dass die von LOHN angestrebte Methode die Harmonisierung der SAKE, der BESTA und der SSUV voraussetzt. Es hat sich aber gezeigt, dass dies nicht vollständig gelingt. Im weiteren ist die Kombination dieser drei Erhebungen für die zu grossen Varianzen verantwortlich, die wiederum für die den heutigen Ansprüchen nicht genügende Qualität der Lohnindices verantwortlich ist. Eine Kombination dieser drei Erhebungen ist somit nicht anzuraten. Selbst wenn eine Harmonisierung zustande kommen sollte, müssen wir immer noch mit nicht erwartungstreuen Schätzern rechnen, da die Kombination der BESTA mit der SAKE in unserem Fall notgedrungen zu verzerrten Schätzern führt. In rund einem drittel der 1013 Lohngruppen (NOGA28 x Geschlecht x Beschäftigungsgrad x Altersklassen x ISCO-Code), in welchen die Population und ein Varianzschätzer des Schätzers der Population geschätzt werden konnte, ist die Verzerrung zum Teil massiv. Eine Schätzung der Lohnindices aufgrund der BESTA und der SSUV, ohne die SAKE, sprich ohne die Variablen Alter und ISCO-Code, ist kurzfristig die Vernünftigste Lösung, sofern eine Harmonisierung zwischen den beiden Populationen hergestellt werden kann und wenn an der Einführung eines vierteljährlichen Lohnmassenindex festgehalten wird. Wird am vierteljährlichen Lohnmassenindex festgehalten, so muss mit zum Teil wohl nicht erklärbaren Schwankungen gerechnet werden, da nicht beobachtbare Grössen einen Einfluss auf die Fluktuationen des Lohnmassenindex haben können, wie zum Beispiel Schwankungen der Unfallwahrscheinlichkeit. Hinzu kommt, dass bei weitem nicht jedes Quartal ein signifikant unterschiedlicher Lohnmassenindex zu jenem des vorhergehenden Quartals für zum Beispiel vollzeitbeschäftigte Männer veröffentlicht werden kann. Mit anderen Worten kann es unter Umständen Jahre dauern, bis von einem signifikant unterschiedlichen Lohnmassenindex auf diesem Niveau gesprochen und dieser somit interpretiert werden kann.
Falls die Ansprüche nicht gesenkt werden können, so sollte eine Datenbasis geschaffen werden, welche die entsprechenden Schätzungen erlaubt. Mit anderen Worten, müsste idealerweise eine Lohnerhebung alle drei Monate durchgeführt

werden.
Ein weiterer Punkt, der nicht gelöst ist, betrifft den 1. Sektor, der in diesem Bericht weitgehend ausgeschlossen wurde. Zur Zeit verfügen wir über keinerlei Daten, um die Schätzungen die wir für den 2. und 3. Sektor durchgeführt haben, auch für den 1. Sektor durchzuführen. Die wohl realistischste Lösung ist, den 1. Sektor nicht mit dem 2. und 3. Sektor zu veröffentlichen.
Bezüglich der Schätzer der Lohnindices, wie wir sie beschrieben haben, ist es notwendig, die Gewichtung für jede Analyse anzupassen, wenn die Entwicklung des Durchschnittslohns anhand eines *Laspeyres*-Preisindex verfolgt werden soll. Gleiches gilt wenn die Entwicklung der Lohnmasse in Untersuchungsbereichen, wie zum Beispiel der Lohnmasse nach Geschlecht, beobachtet werden soll. Die so genannten "à la carte" Analysen, wie sie von LOHN gewünscht worden sind, sind somit nicht ohne eine Neuberechnung der Gewichtung zu machen.
Ein weiterer Punkt, der grundlegend überdacht werden muss, sind die sogenannten homogenen Gruppen. In der derzeitigen Form weisen sie keine offensichtliche Homogenität bezüglich der Löhne auf. Wir haben zurzeit keine Mittel zur Verfügung, um zu überprüfen, ob die homogenen Gruppen bezüglich der Unfallwahrscheinlichkeit homogen sind. Dies wäre das Kriterium für die Homogenität, falls die Daten der SSUV als Stichprobe betrachtet werden. Unter der Voraussetzung, dass die Daten der SSUV für die Schätzung des Durchschnittslohns dienen soll, sehen wir zurzeit keine andere Methode, als die von uns beschriebene, die verwendet werden könnte und welche die Daten der SSUV nicht als Stichprobe betrachtet.

Aufgrund der Schätzung einer unteren Grenze des Varianzschätzers konnten wir die Qualität der derzeit publizierten Durchschnittslohnindices auf dem Niveau Geschlecht, Geschlecht x Sektor, Geschlecht x Wirtschaftsabschnitt, Sektor, NOGA28 und dem Total untersuchen. Aufgrund dieser Analysen sehen wir es als notwendig an, dass in die derzeitigen Publikationen Qualitätsindikatoren eingeführt werden. Wir schlagen einerseits vor, dass die Lohnindices klassifiziert werden, je nach Grösse des Vertrauensintervalls, bei einem Fehler erster Art von 5%. Im weiteren sollten die Veränderungen die bei einem Fehler erster Art von 5% nicht signifikant sind aber bei 10% signifikant sind, in Klammern publiziert werden. Falls die Veränderung auch bei einem Fehler erster Art von 10% nicht signifikant ist, sollte sie durch leere Klammern ersetzt werden. Veränderungen, die bei einem Fehler erster Art von 5% signifikant sind können ohne weitere Bemerkung publiziert werden. Das gleiche Vorgehen mit den Klammern sollte bei den Vergleichen von zwei Durchschnittslohnindices der gleichen Periode aber in unterschiedlichen Untersuchungsbereichen angewandt werden.
Wir haben anhand der Anzahl Beschäftigter aus der BESTA und dem Schätzer des Durchschnittslohns aus den Daten der SSUV einen 'Pseudo-Lohnmassenindex'

konstruiert, um das Verhalten des Lohnmassenindex zu simulieren. Es zeigte sich deutlich, dass die Fluktuationen dieses 'Pseudo-Lohnmassenindex' einerseits aus der Fluktuation des Durchschnittslohns und andererseits aus der Fluktuation der Beschäftigten besteht. Um Schwankungen des Durchschnittslohns erklären zu können oder damit diese in die Schätzung der Lohnmasse miteinbezogen werden können, muss die Entwicklung der Lohndaten aus der SSUV einer eingehenden Analyse unterzogen werden. Es hat sich ebenso gezeigt, dass Handlungsbedarf bezüglich der Interpretation der Fluktuationen des Lohnmassenindex besteht. Wenn der Varianzschätzer des Lohnmassenindex die gleiche Grössenordnung aufweist, wie der von uns konstruierte 'Pseudo-Lohnmassenindex', dann muss damit gerechnet werden, dass über mehrere Quartale hinweg kein signifikant unterschiedliches Niveau des Lohnmassenindex veröffentlicht werden kann.

Anhang

A Variablen und Wirtschaftszweige (u.a. NOGA28)

Tabelle 10: Variablen für die Bildung der homogenen Gruppen.

Variable	# Modalitäten	Modalitäten
ASWZ	28	siehe Tabelle 11
Geschlecht	2	Mann/Frau
Arbeitsbereich	3	Betrieb/Büro/Verkauf
Art der Lohnzahlung	2	Stundenlohn/Monatslohn
Qualifikation	2	Qualifiziert/nicht oder semi qualif.
Unfallart	2	Berufsunfall/Nicht Berufsunfall

Tabelle 11: ASWZ welche für die Bildung der homogenen Gruppen verwendet wurden und auf welchen der derzeitige Durchschnittslohnindex basiert.

ASWZ	Wirtschaftszweig
02	Gartenbau
03	Forstwirtschaft
11, 12	Energie- und Wasserversorgung, Bergbau
21-23	Nahrungsmitteln, Getränke, Tabak
24, 25	Textilien, Bekleidung, Wäsche
26, 27	Holzberarb., Möbel, Papierwaren
28	Grafische Erzeugnisse
29-32	Leder, chem. Erzeugnisse, Kautschuk
33	Abbau u. Verarbeitung v. Steinen und Erden
34	Metallbe- und -verarbeitung
35	Maschinen- und Fahrzeugbau
36	Elektrotechnik, Elektronik
37	Uhren, Bijouteriewaren
41	Bauhauptgewerbe
41-42	Ausbaugewerbe
51-53	Handelsvermittlung und Grosshandel
55, 56	Einzelhandel
57	Gastgewerbe
58	Reparaturgewerbe
61-65	Verkehr
66	Nachrichtenübermittlung
71	Banken, Finanzgesellschaften
72	Versicherungen
73-76	Immobil., Vermietung, Beratung
81, 82	Unterrichtswesen, F+E
83	Gesundheits- und Veterinärwesen
84-89	Umweltschutz, Kultur, Landesverteidigung, Soz. vers.
91-92	Öffentl. Verwaltung

Tabelle 12: NOGA welche für die Bildung der homogenen Gruppen verwendet werden und für welche der revidierte Durchschnittslohnindex geschätzt werden soll.

NOGA	Wirtschaftszweig
1.12	Gartenbau
02	Forstwirtschaft
11-14, 40-41	Bergbau, Energie- und Wasserversorgung
15, 16	Herstellung v. Nahrungsmitteln u. Getränken; Tabakverarbeitung
17, 18	Herstellung v. Textilien u. Bekleidung
20	Holzbe- und verarbeitung
21, 22	Papier-, Karton-, Verglas- u. Druckgewerbe
19, 23-25	Leder, chem. Erzeugnisse, Kautschuk
26	Herst. v. Produkten aus nichtmetallischen Mineralien
27-28	Metallbe- und verarbeitung
29, 34-35	Maschinen- und Fahrzeugbau
30-33	Herstellung v. elektrischen und elektronischen Geräten; Feinmechanik, Optik
36, 37	Möbel, Bijouteriewaren
45	Baugewerbe
50	Fahrzeughandel und -reparaturen
51	Handelsvermittlung und Grosshandel
52	Detailhandel, Reparatur v. Gebrauchsgütern
55	Gastgewerbe
60-63	Verkehr und Transport
64	Nachrichtenübermittlung
65	Kreditgewerbe
66	Versicherungsgewerbe
67	Mit Kredit- und Versicherungsgewerbe verbundene Tätigkeiten
70-74	Immobil., Vermietung, F+E, Beratung
75	Öffentl. Verwaltung, Landesverteidigung, Soz. vers.
80	Unterrichtswesen
85	Gesundheits- und Sozialwesen
90-93	Erbringung von sonstigen öffentl. u. pers. Dienstleistungen

B Klassifikationsbäume

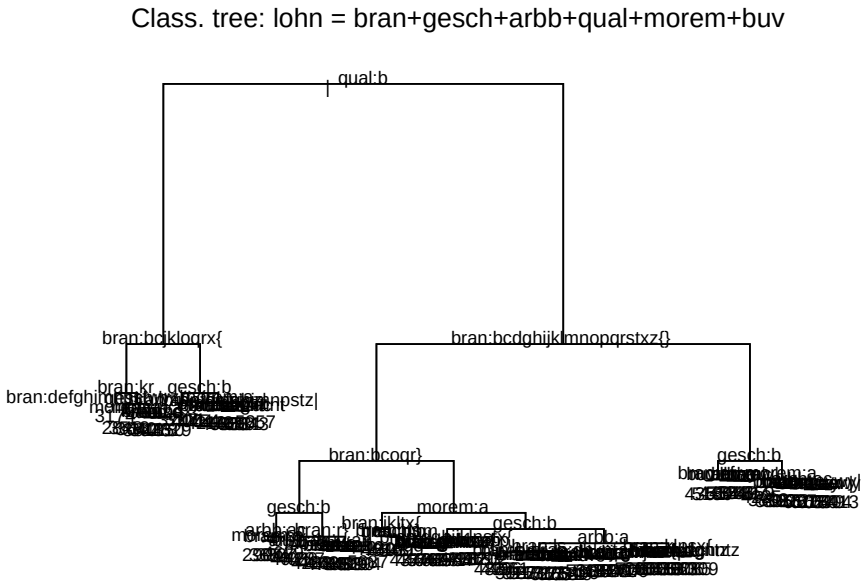


Abbildung 21: Klassifikationsbaum für den Monatslohn (lohn) aufgrund der AS-WZ (bran), dem Geschlecht (gesch), dem Arbeitsbereich (arbb), der Qualifikation (qual), der Art der Lohnzahlung (morem) und der Unfallart (buv). Je länger die Äste, desto grösser ist der Devianzunterschied. Zuoberst ist die Devianz des Baumes am grössten und zu unterst kann sie durch weitere Knotenbildungen nur noch geringfügig verbessert werden.

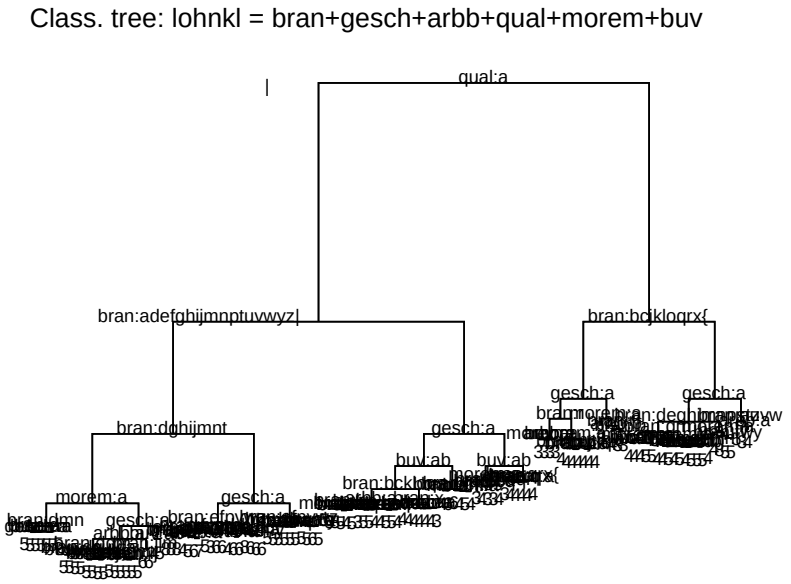


Abbildung 22: Klassifikationsbaum für 1000Fr.-Monatslohnklassen (lohnkl) aufgrund der ASWZ (bran), dem Geschlecht (gesch), dem Arbeitsbereich (arbb), der Qualifikation (qual), der Art der Lohnzahlung (morem) und der Unfallart (buv). Je länger die Äste, desto grösser ist der Devianzunterschied. Zuoberst ist die Devianz des Baumes am grössten und zu unterst kann sie durch weitere Knotenbildungen nur noch geringfügig verbessert werden.

C Einsetzung der NOGA in der SAKE99

Die NOGA Codes in der SAKE werden aufgrund der Firmenadresse und dem Firmennamen automatisch anhand des Betriebs- und Unternehmensregister (BUR) zugeordnet. Danach folgt eine manuelle Zuordnung. Diese Zuordnungen können nicht immer eindeutig gemacht werden, da Name und Adresse Unterschiede zum BUR aufweisen können oder die Firma noch nicht im BUR vorhanden ist. Ist die Zuordnung nicht möglich, so bestehen fehlende Werte in den SAKE Daten. In der SAKE99 hatten 1'112 antwortende Erwerbstätige keinen NOGA Code. Dies entspricht hochgerechnet 13'465 Verunfallten und 106'696 Erwerbstätigen. Deshalb haben wir die NOGA imputiert. Wir haben dies aufgrund der Verteilungen der bestehenden NOGA Codes zufällig mit SAS eingesetzt. Dabei haben wir ein iteratives Vorgehen gewählt, in welchem wir folgende Hilfsvariablen verwendet haben: Geschlecht, Beschäftigungsgrad, ISCO-Code, Altersklasse (15-39, 40-63/65), Anzahl Arbeitgeber, Arbeitsmarktstatus, Betriebsgrösse, Wirtschaftsabschnitt. Einige der Hilfsvariablen wiesen ebenfalls fehlende Werte auf. Bei keiner Übereinstimmung mit diesen Merkmalen, wurde das Modell jeweils reduziert (siehe Tabelle 13).

Tabelle 13: Sequentielle Imputation des NOGA-Codes, zweistellig, in der SAKE99. In der zweiten Kolonne ist die Anzahl Imputationen aufgeführt. Die Hilfsvariablen sind das Geschlecht (sex), der Beschäftigungsgrad (bgrad), der ISCO-Code (isco), Altersklassen (15-39, 40-63/65), die Anzahl Arbeitgeber (ex01), der Arbeitsmarktstatus (bdu1), die Betriebsgrösse (em02) und der Wirtschaftsabschnitt (bmu1).

Imputationsschritt	**Anzahl imp.**	**Hilfsvariablen**							
		sex	**bgrad**	**isco**	**akl**	**ex01**	**bdu1**	**em02**	**bmu1**
1	336	X	X	X	X	X	X	X	X
2	159	X	X	X	X	X	X		X
3	43	X	X	X	X		X		X
4	553	X		X					X
5	21	X		X					

D Schätzung der Populationsgrösse

Tabelle 14: Anzahl verunfallte Personen in der SAKE99 in den 1013 Schichten, welche durch die NOGA, das Geschlecht, den Beschäftigungsgrad , die Berufshauptgruppen (ISCO) und die Altersklassen gebildet werden konnten. '% der Schichten' bezieht sich ausschliesslich auf die Schichten in welchen mindestens eine Person in der SAKE99 vorhanden ist. Die fehlenden NOGA Codes wurden auf die vorhandenen proportional aufgrund der restlichen Schichtungsvariablen verteilt (c.f. Anhang C).

Anzahl Personen in der Schicht	**Anzahl Schichten**	**% der Schichten**
1	298	29.42
2	172	16.98
3	87	8.59
4	46	4.54
5	52	5.13
6	37	3.65
7	32	3.16
8	27	2.67
9	25	2.47
10	14	1.38
11	23	2.27
12	18	1.78
13	14	1.38
14	11	1.09
15	10	0.99
16	14	1.38
17	8	0.79
18	10	0.99
19	6	0.59
20	6	0.59
21	6	0.59
22	10	0.99
23	9	0.89
24	2	0.20
25	10	0.99
26	5	0.49
27	1	0.10
28	4	0.39
29	2	0.20
30	3	0.30
31	5	0.49

Anzahl Personen in der Schicht		Anzahl Schichten	% der Schichten
	32	2	0.20
	33	3	0.30
	34	1	0.10
	35	1	0.10
	36	2	0.20
	37	2	0.20
	38	5	0.49
	39	1	0.10
	40	2	0.20
	44	2	0.20
	45	1	0.10
	46	2	0.20
	47	1	0.10
	49	2	0.20
	52	1	0.10
	53	2	0.20
	54	1	0.10
	56	1	0.10
	57	1	0.10
	58	1	0.10
	59	1	0.10
	64	1	0.10
	68	1	0.10
	69	1	0.10
	85	1	0.10
	89	1	0.10
	96	1	0.10
	103	1	0.10
	110	1	0.10
	115	1	0.10
	116	1	0.10
	176	1	0.10
Total	8116	1013	100.00

Tabelle 15: In der ersten Kolonne ist der Identifikator des Untersuchungsbereichs ('**d**'). In den weiteren drei Kolonnen sind die Variablen aufgelistet, welche den Untersuchungsbereich definieren: NOGA28, Geschlecht (1=Mann, 2=Frau) und der Beschäftigungsgrad (1=vollzeit, 2=teilzeit). In den weiteren Kolonnen folgen der Schätzer der Anzahl Beschäftigter aus der BESTA99II ($\hat{\mathbf{N}}^{\mathbf{B}}_{\mathbf{d,t}}$), die zufällig generierte Anzahl Grenzgänger ($\hat{\mathbf{N}}^{\mathbf{z}}_{\mathbf{d}}$), die Anzahl Personen in der Stichprobe der SAKE ($\mathbf{n}^{\mathbf{s}}$), die Anzahl Beschäftigter aus der SAKE99 geschätzt ($\hat{\mathbf{N}}^{\mathbf{s}}_{\mathbf{b,d}}$), die Anzahl Lohnempfänger aus der SAKE99 ($\hat{\mathbf{N}}^{\mathbf{s}}_{\mathbf{l,d}}$), der Harmonisierungsparameters ($\hat{\alpha}$), die Anzahl Lohnempfänger aus der BESTA und der SAKE ($\hat{\mathbf{N}}_{\mathbf{l,d,t}}$) und schlussendlich noch der Variationskoeffizient des letzten Schätzers ($\mathbf{cv}(\hat{\mathbf{N}}_{\mathbf{l,d,t}})$). In der untersten Zeile sind die Totale wiedergegeben, respektive der Median von $\hat{\alpha}$ und $\mathbf{cv}(\hat{\mathbf{N}}_{\mathbf{l,d,t}})$. Die Werte die für **d**=8 bis 108 gelten, sind schräg geschrieben, ansonsten gelten die Werte für **d**=1 bis 108. Fehlende Werte entsprechen Werten, die nicht geschätzt werden konnten.

d	**NOGA28**	**Geschlecht**	**B.-grad**	$\hat{\mathbf{N}}^{\mathbf{B}}_{\mathbf{d,t}}$ BESTA	$\hat{\mathbf{N}}^{\mathbf{z}}_{\mathbf{d}}$ Zufall	$\mathbf{n}^{\mathbf{s}}$ SAKE	$\hat{\mathbf{N}}^{\mathbf{s}}_{\mathbf{b,d}}$	$\hat{\mathbf{N}}^{\mathbf{s}}_{\mathbf{l,d}}$	$\hat{\alpha}$ $\hat{N}^{B}_{d,t}/(\hat{N}^{s}_{b,d}+\hat{N}^{z}_{d})$	$\hat{\mathbf{N}}_{\mathbf{l,d,t}}$	$\mathbf{cv}(\hat{\mathbf{N}}_{\mathbf{l,d,t}})$
1	1	1	1		0	11	3982	2835			
2	1	2	1		0	2	681	681			
3	1	2	2			2	0	0			
4	2	1	1		0	8	2906	1555			
5	2	2	1		0	1	218	218			
6	2	2	2		0	1	126	126			
7	10	1	1	25582	160	67	25572	25778	0.9942	25628	0.77
8	10	1	2	1130		4	559	1099	2.0226	2223	114.51
9	10	2	1	2269	18	8	1743	1743	1.2879	2245	1.62
10	10	2	2	1988	795	4	1032	1334	1.0880	1452	21.20
11	15	1	1	40178	0	108	40595	38123	0.9897	37731	1.55
12	15	1	2	2549	1104	8	1037	353	1.1904	420	55.32
13	15	2	1	12632	5048	22	6730	5541	1.0725	5943	4.65
14	15	2	2	8526	74	25	6812	7227	1.2382	8949	8.46
15	17	1	1	9105	63	29	8234	6001	1.0974	6586	7.76
16	17	1	2	631	0	2	1512	1512	0.4169	631	4.83
17	17	2	1	10851	4329	21	6356	5193	1.0155	5273	16.31
18	17	2	2	3742	26	17	3568	2514	1.0411	2618	8.78
19	20	1	1	31873	233	78	27441	16680	1.1517	19211	8.04
20	20	2	1	1679	670	3	945	728	1.0395	757	70.65

d	NOGA28	Geschlecht	B.-grad	$\hat{N}^{B}_{d,t}$ BESTA	$\hat{N}^{z}_{d}$ Zufall	n^{s} SAKE	$\hat{N}^{s}_{b,d}$	$\hat{N}^{s}_{l,d}$	$\hat{\alpha}$ $\hat{N}^{B}_{d,t}/(\hat{N}^{s}_{b,d}+\hat{N}^{z}_{d})$	$\hat{N}_{l,d,t}$	$cv(\hat{N}_{l,d,t})$
21	20	2	2	2246	897	10	1248	1304	1.0476	1367	63.33
22	21	1	1	40362	271	107	37981	33560	1.0552	35412	2.07
23	21	1	2	4314	0	18	5245	7055	0.8225	5802	31.16
24	21	2	1	14495	103	46	12767	12434	1.1263	14005	3.26
25	21	2	2	10728	0	42	12121	12280	0.8850	10868	6.90
26	23	1	1	61847	395	169	61301	58649	1.0024	58792	1.79
27	23	1	2	1578	631	7	1162	1184	0.8806	1042	4.72
28	23	2	1	18867	152	52	14683	14654	1.2718	18637	2.17
29	23	2	2	7851	50	27	7776	8989	1.0033	9019	10.26
30	26	1	1	15852	6399	22	6300	5862	1.2484	7318	4.85
31	26	2	1	1998	15	5	1711	1711	1.1575	1981	3.68
32	26	2	2	1239	494	4	696	696	1.0409	724	115.77
33	27	1	1	79888	611	163	65804	57765	1.2029	69483	1.78
34	27	1	2	3266	1315	6	1374	1374	1.2142	1669	16.00
35	27	2	1	10345	70	31	9567	9943	1.0735	10674	10.68
36	27	2	2	6953	0	31	7498	8978	0.9273	8325	7.01
37	29	1	1	105738	803	248	87206	83631	1.2014	100478	1.17
38	29	1	2	3102	22	8	3006	2803	1.0246	2872	10.43
39	29	2	1	12297	79	42	12204	12679	1.0012	12694	5.20
40	29	2	2	6363	44	23	6037	5231	1.0463	5473	3.80
41	30	1	1	80045	504	231	79699	70954	0.9980	70815	2.72
42	30	1	2	3095	1239	6	1554	1235	1.1079	1368	18.50
43	30	2	1	33816	253	83	28272	28709	1.1855	34033	3.97
44	30	2	2	10917	0	39	11467	13271	0.9520	12634	9.41
45	36	1	1	21066	152	47	18332	16326	1.1397	18606	10.16
46	36	1	2	1504	0	6	2092	1387	0.7190	997	112.12
47	36	2	1	4940	1992	7	2018	1638	1.2318	2018	19.76
48	36	2	2	2675	21	12	2345	3172	1.1306	3587	32.64
49	45	1	1	256255	2160	590	190648	155240	1.3291	206324	1.85
50	45	1	2	9382	3753	20	5250	3821	1.0422	3982	18.75
51	45	2	1	11348	88	42	9134	9087	1.2305	11181	15.02
52	45	2	2	15572	108	71	14056	18864	1.0995	20740	13.50

d	NOGA28	Geschlecht	B.-grad	$\hat{N}^{B}_{d,t}$ BESTA	$\hat{N}^{z}_{d}$ Zufall	n^{s} SAKE	$\hat{N}^{s}_{b,d}$	$\hat{N}^{s}_{l,d}$	$\hat{\alpha}$ $\hat{N}^{B}_{d,t} / (\hat{N}^{s}_{b,d} + \hat{N}^{z}_{d})$	$\hat{N}_{l,d,t}$	$cv(\hat{N}_{l,d,t})$
53	50	1	1	61655	536	152	44412	29043	1.3717	39839	8.64
54	50	1	2	3145	1308	8	702	918	1.5645	1436	23.12
55	50	2	1	6848	56	21	5413	5627	1.2522	7046	7.91
56	50	2	2	7749	67	25	5599	4147	1.3676	5671	32.05
57	51	1	1	120269	800	342	114455	105392	1.0435	109977	2.23
58	51	1	2	8579	3424	18	6120	4281	0.8989	3848	21.36
59	51	2	1	37488	260	147	35091	32700	1.0604	34677	4.85
60	51	2	2	26638	170	90	26332	26444	1.0051	26579	10.80
61	52	1	1	98645	0	327	100037	73720	0.9861	72695	4.28
62	52	1	2	15779	6285	38	11291	9969	0.8978	8949	30.90
63	52	2	1	100335	802	308	78930	63203	1.2584	79535	2.60
64	52	2	2	119171	847	333	106039	103795	1.1149	115724	6.33
65	55	1	1	86678	35196	97	31327	24788	1.3030	32299	10.10
66	55	1	2	12664	5225	16	3449	4150	1.4601	6060	21.12
67	55	2	1	75390	30445	93	30027	23560	1.2467	29372	5.55
68	55	2	2	52074	439	102	39219	36187	1.3131	47517	9.46
69	60	1	1	102770	658	304	100465	89958	1.0163	91423	3.21
70	60	1	2	10971	4437	14	4157	2990	1.2766	3817	17.31
71	60	2	1	23692	161	74	21807	20543	1.0785	22155	2.55
72	60	2	2	14079	91	50	13579	14245	1.0299	14671	16.97
73	64	1	1	45792	0	152	48435	47793	0.9454	45184	2.42
74	64	1	2	7977	3213	9	3356	4211	1.2143	5114	4.88
75	64	2	1	16273	0	72	20086	19666	0.8102	15933	2.84
76	64	2	2	16428	0	66	22969	23797	0.7152	17020	6.06
77	65	1	1	64636	420	210	63914	61127	1.0047	61414	3.90
78	65	1	2	4904	37	11	4058	4278	1.1977	5124	18.76
79	65	2	1	34108	0	124	34492	34492	0.9889	34108	1.26
80	65	2	2	15180	0	48	15455	17165	0.9822	16859	5.80
81	66	1	1	33820	228	112	31366	32809	1.0704	35120	2.25
82	66	1	2	1925	933	2	278	125	1.5896	198	10.38
83	66	2	1	17789	7079	48	11065	11132	0.9804	10914	2.76
84	66	2	2	8263	0	28	8663	10048	0.9538	9585	6.03

d	**NOGA28**	**Geschlecht**	**B.-grad**	$\hat{\mathbf{N}}^{\mathbf{B}}_{\mathbf{d,t}}$ BESTA	$\hat{\mathbf{N}}^{\mathbf{z}}_{\mathbf{d}}$ Zufall	$\mathbf{n}^{\mathbf{s}}$ SAKE	$\hat{\mathbf{N}}^{\mathbf{s}}_{\mathbf{b,d}}$	$\hat{\mathbf{N}}^{\mathbf{s}}_{\mathbf{l,d}}$	$\hat{\alpha}$ $\hat{N}^{B}_{d,t}/(\hat{N}^{s}_{b,d}+\hat{N}^{z}_{d})$	$\hat{\mathbf{N}}_{\mathbf{l,d,t}}$	$\mathbf{cv}(\hat{\mathbf{N}}_{\mathbf{l,d,t}})$
85	67	1	1	5212	0	30	10173	9594	0.5123	4915	5.00
86	67	1	2	912		1	0	251			
87	67	2	1	1937	0	16	4697	4048	0.4124	1669	8.91
88	67	2	2	1334	531	3	805	805	0.9981	804	7.74
89	70	1	1	182546	0	653	200079	142464	0.9124	129980	2.88
90	70	1	2	39080	15562	86	24917	16269	0.9654	15707	19.03
91	70	2	1	63546	423	230	59978	50643	1.0521	53280	6.61
92	70	2	2	70648	476	252	66468	72737	1.0553	76763	10.60
93	75	1	1	80726	0	336	104989	105009	0.7689	80741	0.62
94	75	1	2	8429	67	29	8184	8348	1.0216	8529	11.88
95	75	2	1	26093	0	137	35664	36244	0.7316	26517	1.82
96	75	2	2	24423	0	152	52192	57490	0.4679	26902	4.02
97	80	1	1	52997	0	196	69637	68708	0.7611	52290	1.61
98	80	1	2	41778	16605	87	27363	33238	0.9502	31583	8.35
99	80	2	1	40421	0	167	43181	42748	0.9361	40016	2.30
100	80	2	2	80367	0	331	96780	113460	0.8304	94218	3.88
101	85	1	1	72842	0	251	91118	77429	0.7994	61898	5.68
102	85	1	2	21866	154	66	20580	19335	1.0546	20391	6.40
103	85	2	1	111192	741	389	105163	97793	1.0499	102676	2.30
104	85	2	2	166613	0	635	182652	185513	0.9122	169223	3.50
105	90	1	1	49918	0	191	59985	48736	0.8322	40556	2.81
106	90	1	2	18105	7182	57	13473	13481	0.8766	11817	21.16
107	90	2	1	36669	0	165	43457	35212	0.8438	29712	2.95
108	90	2	2	38796	0	254	56006	59598	0.6927	41283	8.66
				3500845	*180000*	10791	*3272820*	3025186	*1.0411*	3013921	*6.40*

E Jackknife-Kovarianzschätzer Methode

Die Anpassung des Macros 'jackanz' beruht auf den folgenden theoretischen Überlegungen. Es seien $\theta_{h,1}$ und $\theta_{h,2}$, zwei zu schätzende Parameter der Schicht h in der Population. Diese werden durch $\hat{\theta}_{h,1}$ und $\hat{\theta}_{h,2}$ geschätzt. Die Stichprobe wird zufällig in jeder Schicht in G_h Gruppen aufgeteilt. Für jede dieser Gruppen definieren wir
$\hat{\theta}_{h,1,(g)}$: Schätzer von $\theta_{h,1}$, von der gleichen funktionalen Form wie $\hat{\theta}_{h,1}$, der auf sämtlichen Daten ausser jenen der Gruppe g der Schicht h basiert.
$\hat{\theta}_{2,(g)}$: Schätzer von $\theta_{h,2}$, von der gleichen funktionalen Form wie $\hat{\theta}_{h,2}$, der auf sämtlichen Daten ausser jenen der Gruppe g der Schicht h basiert.
Weiter definieren wir für $u \in \{1, 2\}$

$$\hat{\theta}_{h,u,g} = G_h \hat{\theta}_{h,u} - (G_h - 1)\hat{\theta}_{h,u,(g)} \tag{152}$$

Der *Jackknife*-Schätzer von $\theta_{h,u}, u \in \{1, 2\}$ ist dann

$$\hat{\theta}_{h,u,JK} = \frac{1}{G_h} \sum_{g=1}^{G_h} \hat{\theta}_{h,u,g} \tag{153}$$

Der *Jackknife*-Kovarianzschätzer von $\hat{\theta}_{h,1}$ und $\hat{\theta}_{h,2}$ nimmt dann folgende Form an

$$\hat{Cov}_{h,JK}(\hat{\theta}_{h,1}, \hat{\theta}_{h,2}) = \frac{1}{G_h(G_h - 1)} \sum_{g=1}^{G_h} (\hat{\theta}_{h,1,g} - \hat{\theta}_{h,1,JK})(\hat{\theta}_{h,2,g} - \hat{\theta}_{h,2,JK}) \tag{154}$$

Für den Schätzer der Kovarianz in (124), $\hat{Cov}(\hat{N}^s_{l,d}, \hat{N}^s_{b,d})$, ist

$$\hat{\theta}_{d,1} = \hat{N}^s_{l,d} \tag{155}$$

$$\hat{\theta}_{d,1,(g)} = c^s_{d,g} \hat{N}^s_{l,d,(g)} \tag{156}$$

$$\text{wobei} \qquad c^s_{d,g} = \frac{\hat{N}^s_{p,d}}{\hat{N}^s_{p,d} - \hat{N}^s_{l,d,(g)}} \tag{157}$$

$$\tag{158}$$

der Korrektur Parameter der Gewichtung ist, der garantiert, dass die Schätzungen sich auf die Population E_d beziehen.
Weiter ist

$$\hat{\theta}_{d,1} = \hat{N}^s_{b,d} \tag{159}$$

$$\hat{\theta}_{d,1,(g)} = c^s_{d,g} \hat{N}^s_{b,d,(g)} \tag{160}$$

$$\text{wobei} \qquad c^s_{d,g} = \frac{\hat{N}^s_{p,d}}{\hat{N}^s_{p,d} - \hat{N}^s_{b,d,(g)}} \tag{161}$$

$$\tag{162}$$

F Simulation des Lohnmassenindex

Tabelle 16: Lohnmassenindex der Vollzeitbeschäftigten nach Geschlecht (`Im_Mann`, `Im_Frau`) vom ersten Quartal 1993 bis zum vierten Quartal 2000, siehe (139), Basis: viertes Quartal 1993. Die unteren und oberen Grenzen des Vertrauensintervalls der Lohnmassenindices (`llim_m` und `ulim_m`, `llim_f` und `ulim_f`) und den zugehörigen Variationskoeffizienten `cv_m` und `cv_f`. Ebenso ein Beschäftigtenindex (`In_Mann`, `In_Frau`), siehe (141) und ein Index des SSUV-Durchschnittslohns (`Is_Mann`, `Is_Frau`) nach Geschlecht, siehe (140). `Im_Mann_publ` und, `Im_Frau_publ` stehen für die publizierbaren Lohnmassenindices.

Periode	Im_ Mann	Im_ Frau	Is_ Mann	Is_ Frau	In_ Mann	In_ Frau	cv_m [%]	cv_f [%]	llim_m llim_m	ulim_m ulim_m	llim_f llim_f	ulim_f ulim_f	Im_Mann _publ	Im_Frau _publ
1993_1	101.6	103.0	101.2	101.9	100.5	101.0	0.61	1.17	100.4	100.6	102.9	105.3	102.8	100.0
1993_2	102.9	102.4	100.3	101.0	102.6	101.4	0.59	1.25	101.7	99.9	104.1	104.9	102.8	100.0
1993_3	102.8	102.0	100.3	100.7	102.5	101.3	0.60	1.25	101.6	99.5	104.0	104.5	102.8	100.0
1993_4	100.0	100.0	100.0	100.0	100.0	100.0	0.61	1.24	98.8	97.4	101.2	102.6	100.0	100.0
1994_1	101.9	103.3	102.4	104.1	99.5	99.2	0.63	1.20	100.7	100.9	103.2	105.8	100.0	100.0
1994_2	101.9	101.1	101.9	102.2	99.9	98.9	0.61	1.26	100.6	98.6	103.1	103.6	100.0	100.0
1994_3	102.4	102.9	101.3	102.3	101.1	100.6	0.61	1.25	101.2	100.4	103.6	105.4	100.0	100.0
1994_4	100.3	101.8	100.8	102.2	99.5	99.7	0.67	1.42	99.0	99.0	101.6	104.7	100.0	100.0
1995_1	103.0	104.9	103.7	105.7	99.3	99.3	0.63	1.17	101.7	102.5	104.2	107.3	103.0	100.0
1995_2	102.1	102.5	102.6	103.9	99.5	98.6	0.63	1.29	100.9	99.9	103.4	105.1	103.0	100.0
1995_3	103.1	104.2	102.9	104.6	100.3	99.7	0.62	1.26	101.9	101.7	104.4	106.8	103.0	100.0
1995_4	101.2	103.5	103.0	105.3	98.3	98.3	0.68	1.42	99.9	100.6	102.6	106.4	103.0	100.0
1996_1	103.4	105.1	105.8	107.7	97.7	97.6	0.65	1.25	102.1	102.5	104.8	107.7	103.0	100.0
1996_2	102.6	103.4	104.6	107.1	98.0	96.6	0.65	1.33	101.3	100.7	103.9	106.1	103.0	100.0
1996_3	102.2	103.8	104.6	107.3	97.7	96.7	0.65	1.33	100.9	101.1	103.5	106.5	103.0	100.0
1996_4	100.2	102.9	104.7	107.3	95.7	95.8	0.72	1.50	98.8	99.8	101.6	105.9	100.2	100.0
1997_1	101.8	105.2	107.0	110.0	95.2	95.6	0.68	1.28	100.4	102.5	103.2	107.8	100.2	100.0
1997_2	100.3	103.3	105.2	108.0	95.4	95.7	0.67	1.37	99.0	100.6	101.7	106.1	100.2	100.0
1997_3	100.5	104.6	105.1	108.6	95.6	96.3	0.68	1.35	99.2	101.9	101.9	107.4	100.2	100.0
1997_4	99.3	102.8	105.2	108.3	94.3	94.9	0.76	1.58	97.8	99.6	100.8	106.0	100.2	100.0
1998_1	101.9	104.6	107.9	110.5	94.5	94.7	0.69	1.30	100.6	101.9	103.3	107.3	100.2	100.0
1998_2	100.9	102.1	105.7	108.2	95.5	94.4	0.68	1.38	99.5	99.3	102.2	104.9	100.2	100.0
1998_3	101.5	103.2	105.8	108.9	95.9	94.8	0.67	1.36	100.2	100.5	102.9	106.0	100.2	100.0
1998_4	99.5	101.7	105.3	108.0	94.4	94.2	0.74	1.52	98.0	98.7	100.9	104.8	100.2	100.0
1999_1	101.8	102.6	107.7	110.0	94.5	93.3	0.70	1.30	100.4	100.0	103.2	105.2	100.2	100.0
1999_2	102.2	101.3	106.7	109.7	95.7	92.3	0.69	1.41	100.8	98.5	103.5	104.1	100.2	100.0
1999_3	102.1	103.1	106.3	110.3	96.0	93.5	0.69	1.37	100.8	100.4	103.5	105.9	100.2	100.0
1999_4	101.2	101.7	106.1	109.0	95.4	93.3	0.77	1.59	99.6	98.6	102.7	104.9	100.2	100.0
2000_1	103.4	105.5	108.2	113.5	95.6	92.9	0.95	1.74	101.5	101.9	105.3	109.1	100.2	100.0
2000_2	103.6	104.4	107.3	111.0	96.5	94.1	0.93	1.82	101.7	100.7	105.5	108.2	103.6	100.0
2000_3	104.1	106.5	107.5	111.6	96.9	95.4	0.99	1.83	102.1	102.7	106.2	110.4	103.6	106.5
2000_4	102.9	108.0	107.2	113.5	96.0	95.1	1.26	2.15	100.4	103.4	105.5	112.5	103.6	106.5

G Vergleich zweier Durchschnittslohnindices

Tabelle 17: Vergleich der Durchschnittslohnindices auf dem Niveau Abschnitt x Geschlecht. d_1 und d_2 steht für die beiden Untersuchungsbereiche, die verglichen werden. α steht für den Fehler erster Art. $\hat{I}^D_{d_1,99,93,\hat{\pi}}$ und $\hat{I}^D_{d_2,99,93,\hat{\pi}}$ sind die veröffentlichten Durchschnittslohnindices von 1999, $\Lambda^D_{d_1,d_2,99} = \hat{I}^D_{d_1,99,93,\hat{\pi}} - \hat{I}^D_{d_2,99,93,\hat{\pi}}$. $\hat{\lambda}^D_{d_1,d_2,99}$ steht für die halbe Länge des Vertrauensintervalls von $\Lambda^D_{d_1,d_2,99}$. Diese wurde für $\alpha = 5\%$ und $\alpha = 10\%$ geschätzt. Wenn $\bigcap = 1$, dann ist der Unterschied zwischen den beiden Durchschnittslohnindices bei entsprechendem α signifikant. Aufgrund fehlender Indexwerten fehlen einige Kombinationen.

d_1		d_2		d_1	d_2					
Abschnitt	**Geschlecht**	**Abschnitt**	**Geschlecht**	$\hat{I}^D_{d_1,99,93,\hat{\pi}}$	$\hat{I}^D_{d_2,99,93,\hat{\pi}}$	$\Lambda^D_{d_1,d_2,99}$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 5\%$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
C, E	M	D	M	105.4	104.7	0.7	4.0	3.3	0	0
C, E	M	D	F	105.4	107.7	-2.3	4.1	3.5	0	0
C, E	M	F	M	105.4	104.5	0.9	4.3	3.6	0	0
C, E	M	G, H	M	105.4	106.0	-0.6	4.2	3.5	0	0
C, E	M	G, H	F	105.4	105.7	-0.3	4.3	3.6	0	0
C, E	M	I	M	105.4	104.9	0.5	4.0	3.3	0	0
C, E	M	I	F	105.4	101.3	4.1	4.2	3.5	0	1
C, E	M	J, K	M	105.4	108.3	-2.9	4.0	3.4	0	0
C, E	M	J, K	F	105.4	111.2	-5.8	4.0	3.4	1	1
C, E	M	L	M	105.4	103.7	1.7	3.9	3.3	0	0
C, E	M	L	F	105.4	104.6	0.8	3.9	3.3	0	0
C, E	M	M,N,O	M	105.4	104.8	0.6	4.0	3.3	0	0
C, E	M	M,N,O	F	105.4	105.2	0.2	4.0	3.3	0	0
D	M	D	F	104.7	107.7	-3.0	1.7	1.5	1	1
D	M	F	M	104.7	104.5	0.2	2.1	1.8	0	0
D	M	G, H	M	104.7	106.0	-1.3	1.8	1.5	0	0
D	M	G, H	F	104.7	105.7	-1.0	2.1	1.8	0	0
D	M	I	M	104.7	104.9	-0.2	1.3	1.1	0	0
D	M	I	F	104.7	101.3	3.4	1.8	1.5	1	1
D	M	J, K	M	104.7	108.3	-3.6	1.4	1.2	1	1

d_1		d_2		d_1	d_2					
Ab-schnitt	**Ge-schlecht**	**Ab-schnitt**	**Ge-schlecht**	$\hat{I}^D_{d_1,99,93,\hat{\pi}}$	$\hat{I}^D_{d_2,99,93,\hat{\pi}}$	$\Lambda^D_{d_1,d_2,99}$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 5\%$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
D	M	J, K	F	104.7	111.2	-6.5	1.4	1.2	1	1
D	M	L	M	104.7	103.7	1.0	1.1	0.9	0	1
D	M	L	F	104.7	104.6	0.1	1.1	0.9	0	0
D	M	M,N,O	M	104.7	104.8	-0.1	1.3	1.1	0	0
D	M	M,N,O	F	104.7	105.2	-0.5	1.3	1.1	0	0
D	F	F	M	107.7	104.5	3.2	2.5	2.1	1	1
D	F	G, H	M	107.7	106.0	1.7	2.2	1.8	0	0
D	F	G, H	F	107.7	105.7	2.0	2.5	2.1	0	0
D	F	I	M	107.7	104.9	2.8	1.8	1.5	1	1
D	F	I	F	107.7	101.3	6.4	2.2	1.8	1	1
D	F	J, K	M	107.7	108.3	-0.6	1.9	1.6	0	0
D	F	J, K	F	107.7	111.2	-3.5	1.9	1.6	1	1
D	F	L	M	107.7	103.7	4.0	1.6	1.3	1	1
D	F	L	F	107.7	104.6	3.1	1.6	1.4	1	1
D	F	M,N,O	M	107.7	104.8	2.9	1.8	1.5	1	1
D	F	M,N,O	F	107.7	105.2	2.5	1.8	1.5	1	1
F	M	G, H	M	104.5	106.0	-1.5	2.5	2.1	0	0
F	M	G, H	F	104.5	105.7	-1.2	2.7	2.3	0	0
F	M	I	M	104.5	104.9	-0.4	2.1	1.8	0	0
F	M	I	F	104.5	101.3	3.2	2.5	2.1	1	1
F	M	J, K	M	104.5	108.3	-3.8	2.2	1.9	1	1
F	M	J, K	F	104.5	111.2	-6.7	2.2	1.9	1	1
F	M	L	M	104.5	103.7	0.8	2.0	1.7	0	0
F	M	L	F	104.5	104.6	-0.1	2.1	1.7	0	0

d_1		d_2		d_1	d_2					
Ab-schnitt	**Ge-schlecht**	**Ab-schnitt**	**Ge-schlecht**	$\hat{I}^D_{d_1,99,93,\hat{\pi}}$	$\hat{I}^D_{d_2,99,93,\hat{\pi}}$	$\Lambda^D_{d_1,d_2,99}$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 5\%$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
F	M	M,N,O	M	104.5	104.8	-0.3	2.2	1.8	0	0
F	M	M,N,O	F	104.5	105.2	-0.7	2.2	1.8	0	0
G, H	M	G, H	F	106.0	105.7	0.3	2.5	2.1	0	0
G, H	M	I	M	106.0	104.9	1.1	1.8	1.5	0	0
G, H	M	I	F	106.0	101.3	4.7	2.2	1.9	1	1
G, H	M	J, K	M	106.0	108.3	-2.3	1.9	1.6	1	1
G, H	M	J, K	F	106.0	111.2	-5.2	1.9	1.6	1	1
G, H	M	L	M	106.0	103.7	2.3	1.7	1.4	1	1
G, H	M	L	F	106.0	104.6	1.4	1.7	1.4	0	0
G, H	M	M,N,O	M	106.0	104.8	1.2	1.8	1.5	0	0
G, H	M	M,N,O	F	106.0	105.2	0.8	1.8	1.5	0	0
G, H	F	I	M	105.7	104.9	0.8	2.1	1.8	0	0
G, H	F	I	F	105.7	101.3	4.4	2.5	2.1	1	1
G, H	F	J, K	M	105.7	108.3	-2.6	2.2	1.9	1	1
G, H	F	J, K	F	105.7	111.2	-5.5	2.2	1.9	1	1
G, H	F	L	M	105.7	103.7	2.0	2.0	1.7	0	1
G, H	F	L	F	105.7	104.6	1.1	2.1	1.7	0	0
G, H	F	M,N,O	M	105.7	104.8	0.9	2.2	1.8	0	0
G, H	F	M,N,O	F	105.7	105.2	0.5	2.2	1.8	0	0
I	M	I	F	104.9	101.3	3.6	1.8	1.5	1	1
I	M	J, K	M	104.9	108.3	-3.4	1.4	1.2	1	1
I	M	J, K	F	104.9	111.2	-6.3	1.4	1.2	1	1
I	M	L	M	104.9	103.7	1.2	1.1	0.9	1	1
I	M	L	F	104.9	104.6	0.3	1.1	1.0	0	0

d_1		d_2		d_1	d_2					
Ab-schnitt	**Ge-schlecht**	**Ab-schnitt**	**Ge-schlecht**	$\hat{I}^D_{d_1,99,93,\hat{\pi}}$	$\hat{I}^D_{d_2,99,93,\hat{\pi}}$	$\Lambda^D_{d_1,d_2,99}$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 5\%$	$\hat{\lambda}^D_{d_1,d_2,99}$ $\alpha = 10\%$	$\bigcap$ $\alpha = 5\%$	$\bigcap$ $\alpha = 10\%$
I	M	M,N,O	M	104.9	104.8	0.1	1.3	1.1	0	0
I	M	M,N,O	F	104.9	105.2	-0.3	1.3	1.1	0	0
I	F	J, K	M	101.3	108.3	-7.0	1.9	1.6	1	1
I	F	J, K	F	101.3	111.2	-9.9	1.9	1.6	1	1
I	F	L	M	101.3	103.7	-2.4	1.7	1.4	1	1
I	F	L	F	101.3	104.6	-3.3	1.7	1.4	1	1
I	F	M,N,O	M	101.3	104.8	-3.5	1.8	1.6	1	1
I	F	M,N,O	F	101.3	105.2	-3.9	1.8	1.5	1	1
J, K	M	J, K	F	108.3	111.2	-2.9	1.6	1.3	1	1
J, K	M	L	M	108.3	103.7	4.6	1.2	1.0	1	1
J, K	M	L	F	108.3	104.6	3.7	1.3	1.1	1	1
J, K	M	M,N,O	M	108.3	104.8	3.5	1.5	1.2	1	1
J, K	M	M,N,O	F	108.3	105.2	3.1	1.5	1.2	1	1
J, K	F	L	M	111.2	103.7	7.5	1.2	1.0	1	1
J, K	F	L	F	111.2	104.6	6.6	1.3	1.1	1	1
J, K	F	M,N,O	M	111.2	104.8	6.4	1.5	1.2	1	1
J, K	F	M,N,O	F	111.2	105.2	6.0	1.5	1.2	1	1
L	M	L	F	103.7	104.6	-0.9	0.9	0.7	1	1
L	M	M,N,O	M	103.7	104.8	-1.1	1.1	0.9	0	1
L	M	M,N,O	F	103.7	105.2	-1.5	1.1	0.9	1	1
L	F	M,N,O	M	104.6	104.8	-0.2	1.2	1.0	0	0
L	F	M,N,O	F	104.6	105.2	-0.6	1.2	1.0	0	0
M,N,O	M	M,N,O	F	104.8	105.2	-0.4	1.4	1.2	0	0
Total									44	48

H SAS Programme

Folgend geben wir das Programm ’all.sas’ wieder, welches als Hauptprogramm dient.

```
/*############################################################################*/
/* Hauptprogramm für die Schätzungen der Lohnindices und deren Genauigkeit. */
/* Die gesamte Arbeit innerhalb der Revision wird von diesem Programm aus   */
/* gesteuert.                                                                */
/*                                                                           */
/* Autor: Daniel Kilchmann, METH/BFS.                          9.9.2004 */
/*___________________________________________________________________________*/

/* Setzen der Librefs und laden der Makros. */
%include ’h:\daten\sas\projets\iss\progs\final\include.sas’;

/* Aufarbeitung der SAKE Daten. */
%include ’h:\daten\sas\projets\iss\progs\final\sake.sas’;

/* Aufarbeitung der SSUV Daten. */
/* Benötigt ssaa.z1998 und ssaa.z1999.      */
%include ’h:\daten\sas\projets\iss\progs\final\ssaa.sas’;

/* Aufarbeitung der BESTA Daten.            */
%include ’h:\daten\sas\projets\iss\progs\final\besta.sas’;
/* Varianzschätzung der BESTA Daten 2000IV. */
/* Benötigt besta2.trim004a.                */
%include ’h:\daten\sas\projets\iss\progs\final\varBESTAdki.sas’;

/* Modell der Einschlusswahrscheinlichkeit. */
%include ’h:\daten\sas\projets\iss\progs\final\ssaaSake.sas’;

/* Schätzung der Indices und der Genauigkeit. */
%include ’h:\daten\sas\projets\iss\progs\final\iss.sas’;

/* Simulation des Lohnmassenindex und Schätzung der Genauigkeit. */
/* Benötigt BESTA Daten von 1993 bis 2000 in Excel Dateien und   */
/* die SAS data sets ssaa.z1993 bis ssaa.z2000                   */
%include ’h:\daten\sas\projets\iss\progs\final\simu_lm_index.sas’;
```

Literatur

[1] Bender, A., Hulliger, B. (1997). Enquête suisse sur la populatino active: rapport de pondération pour 1996. Rapport de méthodes, BFS, Neuchâtel.

[2] BFS (1999). Pressemitteilung: Erwerbstätigkeit und Beschäftigung im 3. Quartal 1999. Fachbereich 3, Arbeit und Erwerb. BFS, Neuchâtel.

[3] BFS (2000). Révision de la statistique de l'évolution des salaires. Cenception générale, BFS, Neuchâtel.

[4] BFS (2000). Révision de la statistique de l'évolution des salaires. Cenception détaillée, BFS, Neuchâtel.

[5] BFS (2004). Schweizerischer Lohnindex 2003: Tabellenauszug. Fachbereich 3, Arbeit und Erwerb. BFS, Neuchâtel.

[6] Capezzalli, E., Farine, A., Moresi, E., Murier, T., Vuille, A. (2002). Arbeitsmarktindikatoren 2002. BFS, Neuchâtel.

[7] Chambers, M., Hastie, T. (1992). Statistical Models in S. Wadsworth & Brooks, Pacific Grove (USA).

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

[9] Comment, T. (1994). Enquête suisse sur la populatino active (ESPA): Intervalles de confiance. SAKE-NEWS 2/94, BFS, Neuchâtel.

[10] Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995). Methodenbericht, BFS, Neuchâtel.

[11] Encyclopedia of statistical sciences (1983). Vol. 4, pp. 54 - 62, Wiley & Sons, New York.

[12] Greene, W. H. (1981). Sample Selection bias as a specification error: comment. Econometrica, Vol. 49, No. 3, pp. 795 - 798.

[13] Hastie T., Tibshirani R. (1997). Modern Regression and Classification, Kursunterlage, Zürich.

[14] Heckman, J. J. (1979). Sample Selection bias as a specification error, Econometrica, Vol. 47, No. 1, pp. 153 - 161.

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

[16] Saucy, F. et al. (2002). Die Beschäftigungsstatistik (BESTA): Methodische Grundlagen 2000. BFS, Neuchâtel.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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