

EVALUATION DER KONTROLLEN IM BESCHAFFUNGSWESEN

Schlussbericht

Erstellt im Auftrag des Eidg. Büros für die Gleichstellung für Frau und Mann EBG
sowie der Beschaffungskommission des Bundes BKB

Zürich, 15. August 2011

Judith Trageser, Susanne Stern, Rolf Iten



INFRAS

INFRAS

**BINZSTRASSE 23
POSTFACH
CH-8045 ZÜRICH
t +41 44 205 95 95
f +41 44 205 95 99
ZUERICH@INFRAS.CH**

**MÜHLEMATTSTRASSE 45
CH-3007 BERN**

WWW.INFRAS.CH

INHALT

Zusammenfassung	3
1. Einleitung	7
1.1. Ausgangslage	7
1.2. Ziele	7
2. Evaluationskonzept	8
2.1. Wirkungsmodell	8
2.2. Evaluationsgegenstand und -Fragen	9
2.3. Methodik	10
3. Praktische Anwendung der Methode	12
3.1. Umgang mit der Datenbasis	12
3.2. Analysemethoden	15
3.3. Dokumentation der Kontrollen	18
4. Ergebnisse der Kontrollen	20
5. Beurteilung	26
5.1. Konsistenz in der Anwendung der Methode	26
5.2. Qualität der Datengrundlagen	29
5.3. Genauigkeit des Messinstruments	32
5.4. Vergleichbarkeit der Ergebnisse	35
5.5. Angemessenheit der Toleranzschwelle	38
6. Empfehlungen	41
Literatur	44

ZUSAMMENFASSUNG

Das Eidg. Büro für die Gleichstellung von Frau und Mann (EBG) und die Beschaffungskommission des Bundes (BKB) haben INFRAS beauftragt, eine Evaluation der bisher durchgeführten Kontrollen zur Lohnungleichheit im Beschaffungswesen durchzuführen. Anlass dazu gab die Motion Carobbio Guscetti (10.3420 Effiziente Bekämpfung der Lohndiskriminierung in Betrieben), welche den Bundesrat u.a. auffordert, die Toleranzschwelle des Kontrollinstruments zu überprüfen. Zudem geben nun seit 5 Jahren gemachten Erfahrungen Gelegenheit, eine Auswertung der durchgeführten Kontrollen vorzunehmen.

Die Ziele der Evaluation sind die Prüfung der praktischen Anwendung der Methoden in 14 kontrollierten Unternehmen bis Ende 2010, die Beurteilung der Resultate der durchgeführten Kontrollen und die Beurteilung der Toleranzschwelle von 5% vor dem Hintergrund der Resultate der bisherigen Kontrollen. Dazu wurde im Rahmen der Evaluation eine systematische Analyse der durchgeführten Kontrollen hinsichtlich der Ergebnisse der Kontrollen, der Qualität der Datenbasis, der Merkmale der kontrollierten Unternehmen und der Anwendung der Methode durchgeführt. Darüber hinaus wurden telefonische Interviews mit vier Fachpersonen, welche bisher die Expertisen erstellt haben, sowie ein vertieftes persönliches Interview mit einer statistischen Fachperson geführt. Ergänzend wurde statistische Fachliteratur herangezogen.

Ergebnisse der Kontrollen

Von den 14 durchgeführten Kontrollen weisen insgesamt 10 Unternehmen eine statistisch signifikante Lohnungleichheit auf. Das bedeutet, dass diese Unternehmen das Gebot der Lohnungleichheit gemäss Art. 3, Abs. 2 des Gleichstellungsgesetzes bzw. Art. 8, Abs. 3 der Bundesverfassung prinzipiell verletzen, weil die Lohndifferenz zwischen Frauen und Männern in diesen Unternehmen unter Berücksichtigung der Unterschiede im Ausbildungsniveau, im Anforderungsniveau, in den potenziellen Erwerbsjahren und den Dienstjahren beim Unternehmen signifikant grösser als 0 ist. Von diesen 10 Unternehmen überschreiten allerdings nur 2 Unternehmen die im Beschaffungswesen geltende Toleranzschwelle von 5%. Bei den anderen 8 Unternehmen ist das Ergebnis entweder zwar höher als 5% aber nicht statistisch signifikant höher als dieser Toleranzwert (4 Unternehmen) oder die gemessene Lohndiskriminierung liegt unter der Toleranzschwelle (4 Unternehmen). In den restlichen vier Unternehmen konnte keine geschlechtsspezifische Lohnungleichheit statistisch nachgewiesen werden.

Beurteilung der Kontrollen

Die durchgeführten Kontrollen wurden auf ihre Qualität und Vergleichbarkeit hin beurteilt. Zudem wurde eine Beurteilung der Angemessenheit der Toleranzschwelle vorgenommen. Kriterien für die Qualität und Vergleichbarkeit der Kontrollen sind die Konsistenz in der Anwendung der Methode, die Qualität der Daten der Unternehmen sowie die Genauigkeit des Messinstruments.

Qualität und Vergleichbarkeit der Kontrollen

- › Konsistenz in der Anwendung der Methode: Das Instrument zur Überprüfung der Lohngleichheit wurde bei den bisherigen Kontrollen insgesamt konsistent angewendet. Die Fachpersonen sind bei den Kontrollen sorgfältig vorgegangen und haben sich – soweit möglich – an die Vorgaben des Instruments gehalten. Trotzdem könnte die Konsistenz durch eine standardisierte Anleitung bei der Plausibilisierung, die Definition von Ausreissern, die Definition von Kriterien für vertiefende Analysen und eine detailliertere Darstellung der Ergebnisse weiter verbessert werden.
- › Qualität der Datengrundlagen: Die Qualität der Datengrundlagen ist mehrheitlich als gut zu bezeichnen. In Einzelfällen könnten die Ergebnisse der Analysen jedoch insbesondere durch fehlende Fälle, durch fehlende Variablen, durch die Handhabung der Einstufung der Mitarbeitenden in die Anforderungsniveaus und Kategorien der beruflichen Stellung sowie durch die geringe Grösse des Unternehmens beeinträchtigt sein. Die Qualität der Datenbasis hängt stark von der Verfügbarkeit und dem Goodwill der Unternehmen ab. Das EBG hat hier wenig juristische Mittel in der Hand, um bei den Unternehmen Druck aufzusetzen.
- › Genauigkeit des Messinstruments: Das Standardmodell zur Überprüfung der Lohngleichheit eignet sich gut, um die geschlechtsspezifische Lohndiskriminierung relativ genau zu schätzen. Die Genauigkeit der Schätzung liesse sich jedoch über alternative Spezifikationen der einbezogenen Variablen und einem standardisierten Verfahren in Bezug auf die Ausreisser weiter verbessern. Darüber hinaus können vertiefende Analysen, die über das Standardmodell hinausgehen, dazu beitragen, die Genauigkeit der Schätzung zu verbessern.
- › Insgesamt wird die Vergleichbarkeit der Ergebnisse auf der einen Seite durch die standardisierte Methode sichergestellt, wobei sie durch eine noch konsistentere Anwendung und bessere Datenqualität weiter optimiert werden könnte. Auf der anderen Seite ist die Vergleichbarkeit durch die einheitlichen Vorgaben des Instruments auch eingeschränkt, weil es nicht auf betriebsspezifische oder branchenspezifische Besonderheiten eingeht. Ein davon abhängiges Problem ist die geringere statistische Aussagekraft bei kleinen Fallzahlen, d.h. bei kleinen Un-

ternehmen, durch die diese Unternehmen bevorteilt sein können, weil eine Verletzung des Gebots der Lohnleichheit schwerer nachzuweisen ist.

Angemessenheit der Toleranzschwelle

Die Toleranzschwelle ermöglicht es, mit einem pragmatischen Verfahren stärkere Verletzungen des Gebots der Lohnleichheit aufzudecken und dabei die Wahrscheinlichkeit gering zu halten, dass Unternehmen fälschlicherweise der Lohndiskriminierung beschuldigt werden. Unter diesem Gesichtspunkt ist die Toleranzschwelle positiv zu würdigen, wobei auch die Höhe von 5% eine angemessene Grössenordnung darstellt. Auf der anderen Seite könnten durch die Toleranzschwelle bei den bisherigen Kontrollen auch einzelne Unternehmen fälschlicherweise frei gesprochen worden sein. Um dies zu verhindern müssten das Standardinstrument und die Toleranzschwelle aus unserer Sicht angepasst werden.

Empfehlungen

Aufgrund der Beurteilung der bisher durchgeführten Kontrollen kommen wir zu den folgenden Empfehlungen:

Empfehlungen zum Standardinstrument

1. Zur Verbesserung der Genauigkeit und Belastbarkeit des Standardmodells könnten verschiedene kleine Optimierungen vorgenommen werden. Dazu gehören standardisierte Verfahren bei der Plausibilisierung sowie bei der Bestimmung von Ausreissern und eine optimierte Spezifikation einzelner Modellvariablen (Ausbildung und Dienstjahre).
2. Bei fehlenden Daten sollte stärker Druck auf die Unternehmen ausgeübt werden. Dazu ist eine bessere juristische Grundlage notwendig.
3. Die Überprüfung der Gleichwertigkeit der Tätigkeiten von einzelnen Mitarbeitenden sollte durch bessere Grundlagen wie Dokumentationen der Unternehmen und Arbeitsbewertungstools verbessert werden.

Das Signifikanzniveau von 5%, welches evtl. kleine Unternehmen bevorteilt, sollte aufgrund verschiedener Überlegungen beibehalten werden. Erstens wäre dies aus juristischer Perspektive kaum rechtmässig, zweitens würde die Akzeptanz und Verständlichkeit des Instruments bei den Unternehmen reduziert und drittens ist eine „faire“ Festlegung der Signifikanzniveaus nach Grösse des Unternehmens in der Praxis kaum umsetzbar. Bei kleinen Unternehmen sollten stattdessen andere Verfahren wie z.B. Matching-Verfahren (Verfahren bei dem für einzelne Frauen statistische Zwillinge gesucht und miteinander verglichen werden) als komplementäre Varianten geprüft werden.

Empfehlung zur Toleranzschwelle

4. Unter der Bedingung, dass die vorgeschlagenen kleineren Optimierungen des Standardmodells umgesetzt und vertiefte Analysen mit Empfehlungen zu diesen zum Standard gemacht werden, empfehlen wir die Toleranzschwelle auf 2%-3% herabzusetzen. Die Herabsetzung der Toleranzschwelle lässt sich folgendermassen rechtfertigen: Erstens erhöht sich durch die Anpassungen die Genauigkeit der Ergebnisse, zweitens zeigen die gemachten Erfahrungen mit den durchgeführten Kontrollen, dass die Toleranzschwelle von 2-3% durchaus von einigen Unternehmen eingehalten wird und drittens sollte die Sensibilisierung der Unternehmen in Bezug auf die Lohnungleichheit inzwischen gewachsen sein. Der Vorteil der Senkung der Toleranzschwelle ist, dass weniger Unternehmen fälschlicherweise von der Diskriminierung freigesprochen würden. Der Nachteil besteht darin, dass sich der Aufwand für die Kontrollen etwas erhöht (wobei bereits bei den durchgeführten Kontrollen häufig vertiefte Analysen vorgenommen wurden) und Einbussen in der Objektivität der Methode entstehen, da die Fachpersonen mehr Handlungsspielraum bei der Analyse haben. Wichtig wäre daher, Kriterien zu definieren, unter welchen Bedingungen vertiefte Analysen durchgeführt werden.

1. EINLEITUNG

1.1. AUSGANGSLAGE

In der Antwort zur Motion Carobbio Guscetti (10.3420 Effiziente Bekämpfung der Lohndiskriminierung in Betrieben) haben die Beschaffungskommission des Bundes (BKB) und das Eidg. Büro für die Gleichstellung von Frau und Mann (EBG) angekündigt, eine Evaluation der bisher durchgeführten Kontrollen zur Lohngleichheit im Beschaffungswesen durchzuführen. Seit Einführung der Kontrollen im Jahr 2006 wurden 14 Unternehmen bis Ende 2010 überprüft. Der Abschluss der fünfjährigen Pilotphase im Jahr 2011 bietet Gelegenheit, eine Auswertung der Resultate und des Kontrollverfahrens durchzuführen.

In der Motion wird der Bundesrat u.a. aufgefordert, die Toleranzschwelle von 5% zu überprüfen. Dies wird damit begründet, dass einige der untersuchten Betriebe an der Grenze des Toleranzwerts lagen. Einige Betriebe lagen auch darüber, jedoch mit einem nicht signifikant von der Toleranzschwelle verschiedenen Ergebnis. Trotz negativem Ergebnis konnten bei diesen Unternehmen gemäss Art. 8, Abs., 2 lit.c des Bundesgesetzes über das öffentliche Beschaffungswesen (BöB) keine Massnahmen ergriffen werden.

Die Toleranzschwelle wurde bei der Erarbeitung der Methodik zur Überprüfung der Lohngleichheit von BKB und EBG festgelegt, weil der Lohn von weiteren objektiven Erklärungsfaktoren beeinflusst werden kann, die in der standardisierten Analyse nicht berücksichtigt werden (BASS 2005). Mit der Toleranzschwelle sollte der Kontrollaufwand in einem vernünftigen Rahmen gehalten und gleichzeitig verhindert werden, dass ein Unternehmen ungerechtfertigt zur Rechenschaft gezogen wird. Bei Einführung dieser neuen Kontrollen im Jahre 2006 wurde diese Schwelle bewusst grosszügig gewählt, um den Unternehmen Zeit zu geben, sich auf die neue Situation einzustellen.

INFRAS wurde vom EBG, in Absprache mit der BKB, beauftragt, die Evaluation der durchgeführten Kontrollen zur Lohngleichheit im Beschaffungswesen durchzuführen.

1.2. ZIELE

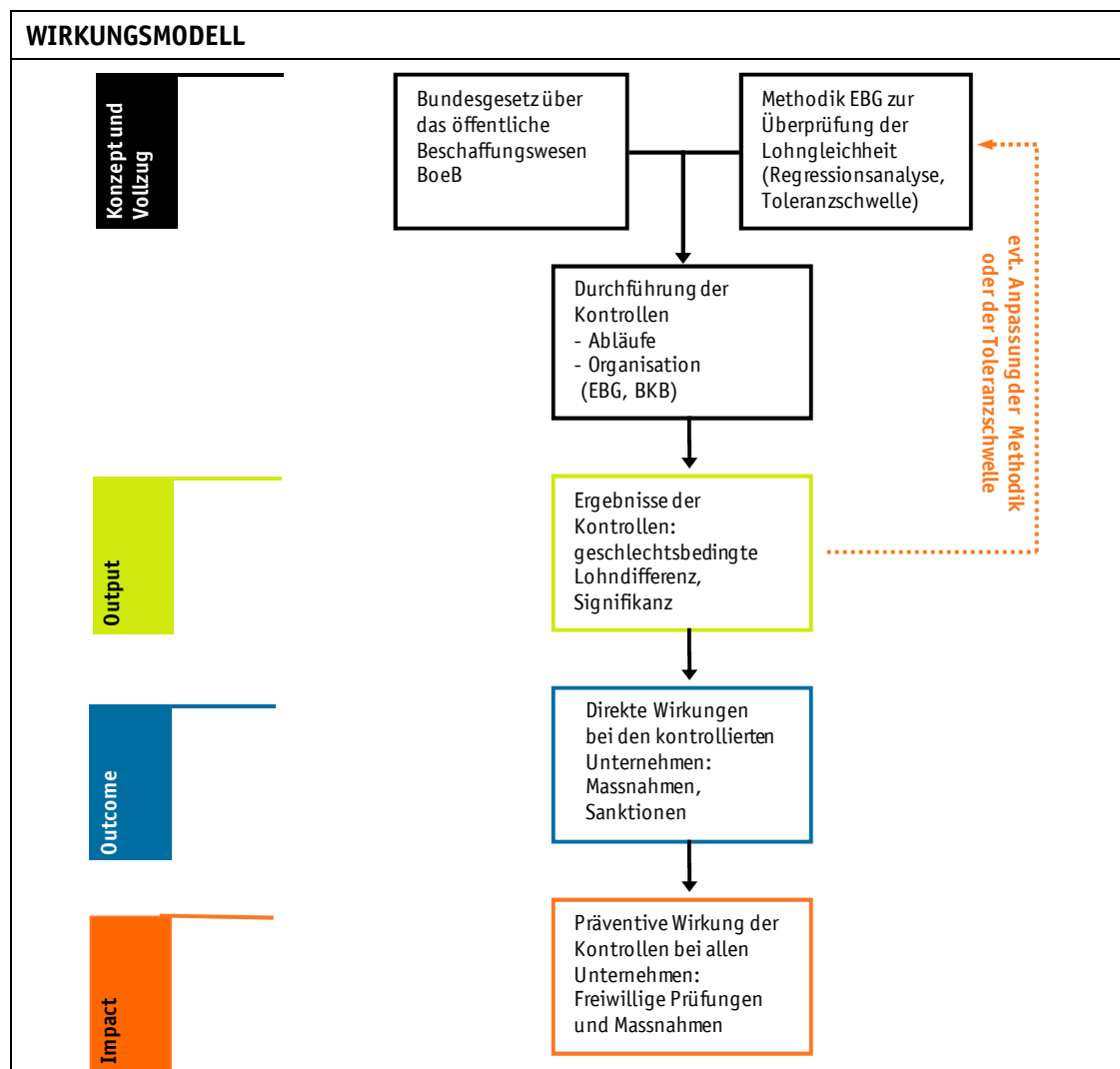
Die ökonomisch-statistische Evaluation verfolgt im Wesentlichen drei Ziele:

1. Prüfung der praktischen Anwendung der Methode in den 14 bisher kontrollierten Unternehmen,
2. Beurteilung der Resultate der bisherigen Kontrollen,
3. Beurteilung der Toleranzschwelle von 5% vor dem Hintergrund der Resultate der bisherigen Kontrollen.

2. EVALUATIONSKONZEPT

2.1. WIRKUNGSMODELL

Das folgende Wirkungsmodell bildet die Basis für die Strukturierung des Evaluationsgegenstandes und der Evaluationsfragen (siehe Figur 1). Der Fokus der ökonomisch-statistischen Evaluation liegt auf der Ebene des Outputs. Auf der Grundlage der Evaluation der Kontrollergebnisse können Rückschlüsse gezogen werden für die Ebene des Konzepts (evtl. Präzisierung der Methodik oder Anpassung der Toleranzschwelle). Die Ebene des Vollzugs ist Gegenstand von separaten Arbeiten der BKB. Die Ebene der Wirkungen wird zum aktuellen Zeitpunkt nicht analysiert.



Figur 1 Wirkungsmodell

2.2. EVALUATIONS-GEGENSTAND UND -FRAGEN

In Tabelle 1 sind die detaillierten Evaluationsfragen im Überblick dargestellt. Wir unterteilen den Evaluationsgegenstand in die drei Bereiche praktische Anwendung der Methodik, Evaluationsergebnisse und Toleranzschwelle.

EVALUATIONS-GEGENSTÄNDE, -KRITERIEN UND -FRAGEN		
Evaluationsgegenstand	Evaluationskriterien	Evaluationsfragen
Praktische Anwendung der Methodik	Angemessenheit, Vergleichbarkeit	› Wie wurde bei den bisher durchgeführten Analysen konkret vorgegangen: Methode (standardisierte, vertiefte Analyse), verwendete Variablen, verwendete statistische Tests, unternehmensspezifische Vorgehensweise, Behandlung von Ausreissern, Fallausschlüsse, etc.? › Ist das von den externen Fachpersonen angewendete Vorgehen konsistent? Welche methodischen Abweichungen/Unterschiede können festgestellt werden? › Sind die angewendeten Vorgehensweisen angemessen? Gibt es evtl. „blinde Flecken“?
Ergebnisse der Kontrollen	Qualität, Vergleichbarkeit	› Was sind die Ergebnisse der bisherigen Kontrollen (deskriptive Analyse)? › Wie ist die Qualität der Kontrollergebnisse zu beurteilen? › Ist die angewandte Methode konsistent, so dass die Ergebnisse vergleichbar sind? › Wie sensitiv sind die Ergebnisse (z.B. beim Ausschluss einzelner Variablen oder Fälle, hinsichtlich Ausreisser, etc.)? › Sind Unterschiede in den Ergebnissen auf Unterschiede in der Anwendung der Methode zurück zu führen?
Toleranzschwelle	Angemessenheit, Zweckmässigkeit	› Wie stark über- oder unterschreiten die Ergebnisse die Toleranzschwelle? › Wie beeinflusst die verwendete Analyse-methode (standardisiert, vertiefte Methode) und die Art und Anzahl der verwendeten Variablen das Über- bzw. Unterschreiten der Toleranzschwelle bzw. die Signifikanz des Ergebnisses? › Wie beeinflusst die Grösse des untersuchten Unternehmens die Signifikanz des Über- bzw. Unterschreitens der Toleranzschwelle. › Wie beeinflusst der Anteil der erklärten Varianz (R^2) das Über- bzw. Unterschreiten der Toleranzschwelle? › Ist die einheitliche Toleranzschwelle von 5% vor dem Hintergrund der bisherigen Kontrollergebnisse angemessen und zweckmässig? Wie könnte die Toleranzschwelle allenfalls angepasst oder abgestuft werden (z.B. unterschiedliche Toleranzschwellen je nach Grösse des Unternehmens, Art der verwendeten Methode, etc.) und wie hätten sich diese Varianten auf die bisher kontrollierten Unternehmen ausgewirkt?

Tabelle 1

2.3. METHODIK

Zur Klärung der Evaluationsfragen haben wir einen Mix an verschiedenen Methoden eingesetzt:

1) Systematische Analyse der durchgeführten Kontrollen

› Diese dient der Beurteilung der praktischen Anwendung der Methode sowie der Vergleichbarkeit der Ergebnisse. Bei der systematischen Analyse wurden die einzelnen Kontrollen im Hinblick auf relevante Kriterien, welche Einfluss auf die Vergleichbarkeit der Ergebnisse haben könnten (angewandte Methode, Grösse des Unternehmens, Art und Anzahl verwendeter Variablen, Anteil erklärter Varianz, etc.), ausgewertet. Die Auswertung erfolgt deskriptiv und strukturiert entlang der folgenden Kriterien:

- › Ergebnisse Diskriminierungskoeffizient und Signifikanz,
- › Erklärungskraft des Modells (R^2),
- › Grösse des Betriebs (N), Anteil Frauen,
- › Anwendung Analysemethode,
- › Anwendung statistische Tests,
- › Umgang mit Ausreissern,
- › Durchführung von Zusatzanalysen,
- › Darstellung und Interpretation der Ergebnisse,

Auf Basis dieser Auswertung wurde die Konsistenz der 14 Kontrollen überprüft und qualitativ analysiert, welchen Einfluss Unterschiede zwischen den Kontrollen auf die Ergebnisse haben könnten.

2) Interviews

- › Es wurden vier Interviews mit Fachpersonen geführt, welche bisher die Kontrollen für das EBG durchgeführt haben. Die Interviews mit den Fachpersonen dienten erstens dazu, ergänzende Informationen zur praktischen Anwendung der Kontrollmethode zu erhalten. Zweitens konnten die Fachpersonen Auskunft über die Sensitivität der Ergebnisse geben¹. Drittens wurden bei den Fachpersonen, basierend auf ihren Erfahrungen mit den kontrollierten Unternehmen, Informationen über die Stärken und Schwächen der Methode, Angemessenheit der Toleranzschwelle sowie Verbesserungsmöglichkeiten eingeholt.
- › Ein persönliches Interview mit der Projektleiterin diente dazu, einen Überblick über die Hintergründe und den Gesamtverlauf der bisherigen Kontrollen zu erhalten.

¹ Eigentliche Sensitivitätsanalysen können nicht durchgeführt werden, da wir aus Datenschutzgründen keinen Zugriff auf die Einzeldaten der kontrollierten Unternehmen haben.

- › Darüber hinaus haben wir ein vertieftes persönliches Interview mit einer statistischen Fachperson geführt, die sich im Speziellen mit der Methode der multiplen logistischen Regression im sozialwissenschaftlichen Bereich auseinandergesetzt hat. Mit der Fachperson wurden die Ergebnisse und die Angemessenheit der Toleranzschwelle in statistischer Hinsicht diskutiert.

3) Literaturanalyse

Zur Beurteilung der Angemessenheit der Toleranzschwelle haben wir uns neben dem eigenen Know-How auf statistische Literatur abgestützt. Gesucht wurde auch nach Literatur zu Erfahrungen mit Kontrollinstrumenten aus anderen Bereichen, die auf der Regressionsmethode basieren.

3. PRAKTISCHE ANWENDUNG DER METHODE

In diesem Kapitel sind die Ergebnisse zur praktischen Anwendung des Instruments bei den bisher durchgeführten Kontrollen aufgeführt. Im Vordergrund stehen dabei der Umgang mit der Datenbasis, die verwendeten Analysemethoden und die Darstellung der Ergebnisse durch die Personen, die die Kontrollen durchgeführt haben.

3.1. UMGANG MIT DER DATENBASIS

Die von den Betrieben zur Verfügung gestellten Daten wiesen eine unterschiedliche Qualität auf. In den wenigsten Fällen waren die gelieferten Daten lückenlos. Die befragten Fachpersonen, welche die Expertisen bisher erstellt haben, berichten, dass der Prozess der Plausibilisierung teilweise sehr aufwändig ist. In einzelnen Fällen dauerte es Wochen bis Monate, bis die korrekten Daten geliefert wurden. In einem Fall wurden die Daten in der ersten Lieferung sogar auf Papierbasis geschickt. Gemäss einer Fachperson macht die Beschaffung der korrekten Daten jeweils ein Drittel des Aufwands aus. Dass die Unternehmen einfach die Daten der Lohnstrukturerhebung des BFS (LSE) übernehmen könnten, sei eher die Ausnahme.

Umgang mit fehlenden Daten

In zwei Unternehmen wurden die Daten von Tochterfirmen nicht geliefert und bei der Hälfte der kontrollierten Unternehmen fehlten Daten zu einzelnen Variablen. Meistens betraf dies die Variable Ausbildung. Mehrmals kam es auch vor, dass die Löhne der obersten Geschäftsleitungsebene nicht ausgewiesen waren. Die einzelnen Fälle mit fehlenden Angaben zu Variablen sind im Folgenden zusammenfassend aufgelistet:

- › Bei zwei Fällen wurden wie erwähnt ganze Divisionen nicht berücksichtigt. In beiden Fällen lässt sich dies jedoch gemäss EBG vertreten. In einem Fall sind die Divisionen rechtlich unabhängig von dem Mutterunternehmen, in dem anderen Fall waren die Divisionen vor der Kontrolle an ein anderes Unternehmen verkauft worden.
- › Bei einem grossen Unternehmen mit knapp 5'000 Mitarbeitenden (MA) fehlten die Daten zur Ausbildung von allen Mitarbeitenden. Diese Daten konnten nicht nacherhoben werden. In einer Firma mit rund 500 Mitarbeitenden trat der gleiche Fall auf. Hier wurden von einem Teilkollektiv von 166 Mitarbeitenden die Daten nacherhoben. In einem weiteren Unternehmen fehlten nur bei 18 von knapp 4'000 Mitarbeitenden die Ausbildungsdaten. Diese 18 Fälle wurden aus der Analyse ausgeschlossen. Schliesslich lieferte ein Unternehmen vier statt neun Ausbildungsstufen, wobei dies durch das Tätigkeitsgebiet des Unternehmens begründet ist.

- › In einem Fall fehlten die Variablen Ausbildung, berufliche Stellung und Anforderungsniveau. Diese konnten jedoch nacherhoben werden.
- › In zwei Fällen wurden die Löhne der obersten Führungsebene nicht mitgeliefert. In beiden Kontrollen wurden die Fälle ausgeschlossen bzw. bei einer Kontrolle mit Proxy-Werten ein worst-case-Szenario berechnet. In einer weiteren Kontrolle fehlten bei drei von 35 Männern der zwei obersten Funktionsstufen die Löhne. Diese drei Männer wurden ausgeschlossen.
- › In einem Fall wurde bei 214 von 3'153 Mitarbeitenden das Dienstalter nicht mitgeliefert. Diese Fälle wurden aus der Analyse ausgeschlossen.

Eine Fachperson weist darauf hin, dass sich fehlenden Daten auch aufgrund des Stichmonats ergeben können. So würden teilweise einzelne Mitarbeitende nicht erfasst, weil sie nur einen Teil des Jahres gearbeitet haben. Ein anderes Problem bildeten die Zulagen, die z.T. nur am Ende des Jahres gezahlt würden. Um diese korrekt zu erfassen, müsste der Dezember als Stichmonat genommen werden.

Plausibilisierung und Bereinigung

Vorgehen bei der Plausibilisierung

Zur Plausibilisierung der Daten setzten die Fachpersonen verschiedene Methoden ein. Die Befragten gingen dabei ähnlich vor:

- › Zunächst machten die Fachpersonen einzelne *deskriptive Auswertungen zu den einzelnen Verteilungen der Variablen*. Aufgrund dessen wurden erste optische Kontrollen gemacht und erste Ausreisser identifiziert. Dabei kontrollierten die Befragten z.B. auch plausible Minimal- und Maximalwerte der Löhne.
- › Des Weiteren wurden *einfache Plausibilitätskontrollen* wie z.B. der Vergleich von Alter und Dienstalter oder die Übereinstimmung der Lohnbestandteile (z.B. 13. Gehalt sollte 6–8% vom Gesamtlohn betragen) durchgeführt.
- › Die Fachpersonen setzten auch verschiedene Methoden ein, um mögliche *abgeleitete Variablen* zu kontrollieren. Damit wird ausgeschlossen, dass z.B. die Einstufung der Anforderungsniveaus auf Basis des Ausbildungsniveaus oder der beruflichen Stellung und nicht aufgrund der Vorgaben der Lohnstrukturerhebung getroffen wurde. Dies birgt die Gefahr, dass nicht sauber unterschieden werden kann, wodurch Lohnunterschiede im Unternehmen erklärbar sind. Eine weitere Gefahr besteht, wenn Unternehmen das Anforderungsniveau aufgrund von bereits gegebenen Stufen (z.B. Funktionsstufen oder Lohnstufen) zuteilen, da diese Einstufungen bereits Diskriminierungen beinhalten und diese dann durch die Analyse nicht aufgedeckt werden

können. Solchen möglichen abgeleiteten Variablen sind die Fachpersonen gemäss den Berichten und Interviews entweder durch *Korrelationsanalysen zwischen den Variablen*, *einfachen Regressionsanalysen* oder *im Gespräch mit den Unternehmen* näher auf den Grund gegangen.

- › Nach diesen Plausibilisierungsschritten wurden die Daten an die Unternehmen zurück gespielt und Korrekturen eingefordert.
- › Ein weiterer Plausibilisierungsschritt besteht in der Residuenanalyse im Rahmen der Regressionsanalyse. Diese dient einerseits der Prüfung der Modellvoraussetzungen für die Regressionsanalyse. Andererseits zeigt sie einzelne Fälle auf, die vom durch die Regression geschätzten Lohn stark abweichen². Die Residuenanalyse wurde von den Fachpersonen teilweise bereits vor und teilweise nach der zweiten Datenlieferung (je nach Qualität der Daten) durchgeführt.
- › Schliesslich wurden gemäss einzelnen Interviews auch die korrigierten Daten auf Änderungen in plausiblen Daten überprüft. Dazu existieren laut einer Person spezielle Statistikprogramme.

Aufgrund dieser Plausibilisierungsschritte wurden folgende Schwachstellen in der Datenbasis aufgedeckt:

Ausreisser und Fallausschlüsse

Durch die Plausibilitätschecks konnten verschiedene Ausreisserfälle in den Datensätzen identifiziert werden. Dabei handelte es sich entweder um Fälle, die gemäss Kontrollinstrument nicht in die Analyse einfließen (z.B. Lehrlinge, Praktikanten, IV-Fälle), Fehler in der Datenbasis oder erklärbare bzw. plausible Ausreisser. Die Fehler in der Datenbasis wurden jeweils in Absprache mit den Unternehmen korrigiert. Bei den plausiblen Ausreissern handelte es sich beispielsweise um eine Person, die nicht mehr in der Geschäftsleitung war, jedoch einen lebenslangen Bonus erhielt, ein Student, welcher wertvolle Informatikdienstleistungen erbringt, im Ausland angestellte Mitarbeitende, etc.). Bei diesen Fällen wurde in Absprache mit dem EBG entschieden, ob sie aus der Analyse ausgeschlossen werden sollen oder nicht. Die plausiblen Ausreisser wurden jeweils deklariert.

Unplausible Besetzung einzelner Anforderungsniveaus und Kategorien beruflicher Stellungen

In vier der kontrollierten Unternehmen zeigte sich, dass die Anforderungsniveaus und die Kategorien der beruflichen Stellung nicht vollständig besetzt waren. So waren z.B. in einem Unternehmen 100% der Frauen und 90% der Männer derselben beruflichen Stellung (ohne Kaderfunk-

² Die Regression berechnet den Lohn, der aufgrund der Lohnstruktur des Unternehmens und den Qualifikationsmerkmalen der Mitarbeitenden resultieren sollte. Die Residuenanalyse stellt die Ausreisser von diesem berechneten Lohn dar.

tion) zugewiesen. Die Einstufungen waren laut der Fachpersonen nicht plausibel. Sie wurden jedoch nur in einem Unternehmen korrigiert.

Abgeleitete Variablen

Bei drei der kontrollierten Unternehmen zeigte sich, dass diese ihre Mitarbeitenden nicht auf Basis der Vorgaben der Lohnstrukturerhebung (LSE) in die Anforderungsniveaus oder Stufen der Variable berufliche Stellung eingeteilt haben:

- › Im ersten Fall wurden die Einstufungen der Mitarbeitenden in die Anforderungsniveaus und Kategorien der beruflichen Stellung direkt aus den Lohnstufen (nach Landold-Mächler) abgeleitet. Dies wurde in einer zweiten Version durch das Unternehmen korrigiert.
- › Im zweiten Fall wiesen das Anforderungsniveau und die berufliche Stellung eine starke Korrelation mit der Funktionsstufe im Unternehmen auf. Die damit verbundenen Probleme wurden im Bericht deklariert, die Daten jedoch nicht durch das Unternehmen korrigiert.
- › Im dritten Fall waren ebenfalls Anforderungsniveau und berufliche Stellung aufgrund der Funktionsstufen der Firma festgelegt. Dies wurde mit der HR-Leiterin überprüft und daraufhin einzelne Anpassungen gemacht.

3.2. ANALYSEMETHODEN

Bei allen durchgeführten Kontrollen wurde die Standardmethode gemäss Kontrollinstrument verwendet. Laut den Fachpersonen war dies immer möglich. Die Fachpersonen haben sich dabei an die Vorgaben des Instruments gehalten. Obwohl die Anwendung der Standardmethode immer möglich war, sind die Fachpersonen dennoch auf verschiedene Schwierigkeiten und Grenzen gestossen, aufgrund derer sie vertiefte oder erweiterte Analysen verwendet haben:

Einbezug weiterer Variablen

In drei Kontrollen wurden zusätzliche Variablen in die Regressionsanalyse aufgenommen. Hintergrund dieser erweiterten Analyse waren jeweils betriebsspezifische Besonderheiten aufgrund derer die Firmen argumentierten, dass sie für das Lohnsystem relevant sind und daher das Ergebnis beeinflussen könnten:

- › In einem schweizweit tätigen Unternehmen wurde der Arbeitsort in die erweiterte Regressionsanalyse als arbeitsplatzbezogenes Merkmal einbezogen (13 Standorte). Dies hat das Ergebnis der Lohndiskriminierung zu Ungunsten der Frauen verstärkt. Auch in einer anderen Kontrolle wurde der Standort mit einer Variablen (3 Ausprägungen) berücksichtigt.

- › In einer weiteren Kontrolle wurde eine Schicht-Arbeit-Variable einbezogen.

Getrennte Regressionsanalyse mit Teilkollektiven

Getrennte Regressionsanalysen bei verschiedenen Teilkollektiven können laut den Fachpersonen in Einzelfällen sinnvoll sein, wenn sich die Lohnstruktur nach einem Merkmal unterscheidet. Dieses Verfahren kam bei den durchgeführten Kontrollen in zwei Fällen zum Zug:

- › In einer Kontrolle stellte sich heraus, dass Angestellte im Monatslohn und Angestellte im Stundenlohn unterschiedlich entlohnt werden. Mitarbeitende im Monatslohn wiesen aufgrund von besser entschädigter Nachtarbeit einen höheren standardisierten Monatslohn auf. Bei der getrennten Regressionsanalyse ergab sich eine tiefere Lohnungleichheit zwischen Frauen und Männern.
- › In einer anderen Kontrolle wurden getrennte Analysen für das Mutter- und Tochterunternehmen berechnet. Bei der Standardanalyse hatte sich eine nicht durch die anderen Merkmale erklärbare Lohnungleich zwischen Frauen und Männern von über 5% ergeben. Die getrennten Analysen ergaben, dass die Lohnungleichheit nach Definition des Instruments im Mutterunternehmen eingehalten ist, nicht aber im Tochterunternehmen (wobei das Ergebnis für das Tochterunternehmen nicht signifikant gegenüber der Toleranzschwelle war).

Getrennte Regressionsanalysen nach Männern und Frauen

Getrennte Regressionsanalysen für Männer und Frauen im Unternehmen können helfen, die Lohnpraxis nach Geschlecht genauer zu analysieren. Sie zeigen auf, welchen Einfluss die einzelnen Qualifikationsmerkmale oder arbeitsplatzbezogenen Merkmale auf den Lohn der Frauen und der Männer haben. So wird gemäss einer Fachperson häufig die potenzielle Erwerbserfahrung von Frauen und Männern unterschiedlich von Arbeitgebern bewertet. Einem Mutterschaftsurlaub wird dabei zum Beispiel weniger Gewicht gegeben als eine Erwerbserfahrung an einem anderen Ort.

Getrennte Analysen für Männer und Frauen wurden bei zwei Kontrollen durchgeführt. In beiden Fällen zeigte sich, dass die Lohnpraxis für Männer und Frauen unterschiedlich war. In einem Fall war für die Frauen nur die potenzielle Erwerbserfahrung für den Lohn ausschlaggebend, während bei den Männern alle Faktoren mitspielten. In dem anderen Fall profitierten die Frauen im Vergleich zu den Männern weniger von einer besseren Ausbildung und einer höheren Anzahl Dienstjahre.

Matching-Verfahren

In zwei Kontrollen führten die Fachpersonen ein sogenanntes Matching-Verfahren durch. Hintergrund für dieses Verfahren war, dass in den entsprechenden Unternehmen nur wenige Frauen beschäftigt waren (4 und 13) und daher das Ergebnis der Standardanalyse besser abgetüzt werden sollte. Beim Matching-Verfahren werden einzelne Frauen oder homogenen Gruppen von Frauen mit einzelnen Männern oder Gruppen von Männern verglichen, die in den lohnrelevanten Merkmalen gleiche oder ähnliche Ausprägungen aufweisen. Das Verfahren wurde in den zwei Fällen etwas unterschiedlich gehandhabt. In einem Fall wurden gematchte Gruppen miteinander verglichen und die Mittelwertunterschiede anhand eines statistischen Tests untersucht. Im zweiten Fall wurden keine Gruppen gematcht, sondern den 13 Frauen manuell einzelne oder mehrere Männer mit ähnlichen Voraussetzungen zugewiesen und die Lohnunterschiede dieser gematchten Paare deskriptiv ausgewiesen. Im ersten Fall ergab sich in einer von zwei Gruppen eine signifikante Lohnungleichheit zwischen Männern und Frauen mit gleichen Voraussetzungen zu Ungunsten der Frauen. Im zweiten Fall schwankten die Lohnungleichheiten der gematchten Paare zwischen 0.4% und 13.6% zu Ungunsten der Frauen. In beiden Fällen deuten die Ergebnisse darauf hin, dass eine stärkere Lohnungleichheit zu Ungunsten der Frauen in den Unternehmen vorherrscht als durch das Standardinstrument als das Standardinstrument ausweist. Ob die Kriterien des Beschaffungswesens verletzt werden, lässt sich durch die Matching-Verfahren jedoch nicht nachweisen, da die Methodik des Matching-Verfahrens von der Standardanalyse abweicht und die Ergebnisse nicht direkt miteinander vergleichbar sind.

Weitere spezielle Analysen

In einzelnen Fällen wurden darüber hinaus weitere abweichende Analysen durchgeführt. Diese hatten alle den Hintergrund, dass die Datenbasis lückenhaft oder die Verteilung der Variablen untypisch waren:

- › In einem Unternehmen verteilten sich die weiblichen Arbeitskräfte zu 90% auf die höchsten Ausbildungsstufen und zu 10% auf tiefere Niveaus (Lehrniveau). Weil die Variable im Modell ordinal skaliert ist, führte dies gemäss der Fachperson, die die Expertise durchführt, dazu, dass die Ausbildungsprämien (zusätzlicher Lohn, der aufgrund eines höheren Ausbildungsniveaus anfallen müsste) auf den unteren Ausbildungsniveaus überschätzt wurden. Dadurch sei es auch zu einer Überschätzung der geschlechtsspezifischen Lohndiskriminierung zu Ungunsten der Frauen gekommen. Um diese Verzerrung zu kontrollieren, wurde das Modell alternativ

mit einer Dummy-Variablen³ berechnet. Diese alternative Modellspezifikation führte zu einer tieferen geschlechtsspezifischen Lohnungleichheit.

- › Bei zwei Kontrollen wurde aufgrund von fehlenden Daten eine „*Worst-Case-Szenario*“-Analyse gemacht. In einem Fall fehlten die Daten zu den Ausbildungsniveaus von der Mehrheit der Mitarbeitenden. Von den Frauen und einem Teilkollektiv der Männer waren die Ausbildungsniveaus bekannt. Auf Basis des Teilkollektivs wurde bei den Männern mit fehlenden Angaben Ausbildungsniveaus eingesetzt, die schlechter als alle bekannten Ausbildungsniveaus waren. Dadurch wurde die Lohndiskriminierung für den Fall bestimmt, dass alle nicht erfassten Männer über eine schlechtere Ausbildung als die Frauen verfügen. Daraus resultiert die grösstmögliche Lohnungleichheit zwischen Männern und Frauen. Diese lag höher als die ohne Ausbildungsniveau berechnete Lohndiskriminierung. Sie lag knapp über der Signifikanzschwelle, jedoch nicht signifikant. In einem zweiten Fall wurden für fehlende Lohndaten von vier geschäftsleitenden Männern Proxy-Löhne eingefügt, die höher als die erwarteten Löhne waren. Dadurch veränderte sich das Ergebnis gegenüber der Analyse ohne Geschäftsleitende nicht.
- › In einer Kontrolle wurde zusätzlich zum Standardmodell eine Analyse ohne die Löhne der drei männlichen Geschäftsleitungsmitglieder durchgeführt, da diese gegenüber der restlichen Belegschaft sehr stark nach oben abwichen. Dies führte zu dem Ergebnis, dass der Diskriminierungseffekt leicht stieg und damit signifikant über der Toleranzschwelle lag, während der Effekt vorher nicht signifikant über der Schwelle lag.

3.3. DOKUMENTATION DER KONTROLLEN

Die Durchführung und Ergebnisse der Kontrollen wurden in allen Fällen ausführliche dokumentiert. Die Darstellung erfolgte mit einzelnen kleineren Abweichungen relativ standardisiert. Die Kontrollberichte erstrecken sich in der Regel auf ca. 20 Seiten und sind nach den folgenden Inhalten strukturiert:

- › Zusammenfassung der wichtigsten Resultate,
- › Ausgangslage, grober Ablauf und Besonderheiten bei den Kontrollen,
- › Datenbasis (Stichmonat, fehlende Daten, verwendete Lohndaten, etc.),
- › Summarische Beurteilung der Lohnungleichheit:
 - › Tabellarische Darstellung der Durchschnittswerte und Verteilungen der Variablen des Modells nach Geschlecht (inklusive durchschnittliche Abweichung der Frauen),

³ Statt ordinal, wurde die Variable nominal skaliert und 9 Ausbildungsabschlüsse als Dummies eingefügt. Dadurch wird der zusätzliche Wertzuwachs eines Ausbildungsniveaus nicht wie bei der ordinalen Variable vordefiniert, sondern die Abstände zwischen den Ausbildungsstufen durch das Modell optimiert.

- › Grafische, tabellarische und schriftliche Darstellung der Durchschnittslöhne der Frauen und Männer nach den lohnrelevanten Merkmalen. Dadurch lässt sich in einem ersten Schritt erkennen, wie die Löhne von den einzelnen Variablen abhängen.
- › Analyse der Einflussfaktoren: Tabellarische und schriftliche Darstellung der Ergebnisse der Basisregression und erweiterten Regression, Interpretation und Fazit in Bezug auf die Einhaltung der Lohnungleichheit.
- › Eventuelle erweiterte Analysen und deren Ergebnisse.

In einigen Fällen weichen die Berichte von der oben beschriebenen Struktur und Darstellung ab:

- › Bei der summarischen Beurteilung der Lohnungleichheit fehlt teilweise die Anzahl der Männer und Frauen je Ausprägung.
- › Bei einer Kontrolle fehlen die Ergebnisse der Basisregression in der tabellarischen Übersicht.
- › Eine Fachperson stellt die Ergebnisse der Basisregression in der tabellarischen Übersicht etwas anders dar. Statt die durch das Geschlecht erklärte Lohndiskriminierung unter Berücksichtigung der anderen Faktoren auszuweisen, wird der durch die anderen Faktoren erklärte Anteil ausgewiesen.
- › Eine Fachperson stellt die Ergebnisse der Regressionsanalysen anschaulich grafisch dar. Diese Darstellung zeigt die durch die lohnrelevanten Faktoren erklärten Anteile der Lohndifferenz.
- › Eine Fachperson stellt die Ergebnisse der Regressionsanalysen und Residuenanalysen im Anhang ausführlich dar. Die Darstellung der Ergebnisse der Residuenanalyse erlaubt es zu kontrollieren, inwieweit die Modellvoraussetzungen erfüllt sind.

Standardmässig nicht dokumentiert sind in den Berichten die genauen Ergebnisse zur Signifikanz der Lohnungleichheit. Aus den Berichten ist meistens nicht ersichtlich, auf welchem Niveau die Lohnungleichheit signifikant ist bzw. wäre.

4. ERGEBNISSE DER KONTROLLEN

Die Ergebnisse der bisherigen Kontrolle sind in der folgenden Tabelle zusammenfassend dargestellt. Vorgängig sind die in der Tabelle aufgeführten Kennzahlen erläutert:

Erläuterungen zu den aufgeführten Kennzahlen:

- › Brutto-Lohndifferenz: Differenz der Durchschnittslöhne der Männer und Frauen ohne Kontrolle von erklärenden Faktoren.
- › Geschlechtsspezifische Lohndifferenz nach Basisregression: Anteil der Brutto-Lohndifferenz, der nicht durch *persönliche Qualifikationsmerkmale* (*Ausbildungsjahre, potenzielle Erwerbsjahre, Dienstjahre*) sondern durch das Geschlecht erklärt werden kann.
- › Geschlechtsspezifische Lohndifferenz nach erweiterter Regression: Anteil der Brutto-Lohndifferenz, der nicht auf persönliche Qualifikationsmerkmale sowie *arbeitsplatzbezogene Merkmale* (*Anforderungsniveau und berufliche Stellung*), sondern rein auf das Geschlecht zurück zu führen ist.
- › Signifikanz von 0: Gibt an, ob die geschlechtsspezifische Lohndifferenz signifikant grösser als 0 ist. Ja bedeutet, dass eine statistisch gesicherte geschlechtsspezifische Lohnungleichheit zwischen Frauen und Männern besteht (mit einer Irrtumswahrscheinlichkeit von 5%). Nein bedeutet umgekehrt, dass die geschlechtsspezifische Lohndifferenz zwischen Frauen und Männern statistisch nicht gesichert ist.
- › Signifikanz von Toleranzschwelle: Ja bedeutet, dass die gemessene geschlechtsspezifische Lohndifferenz signifikant grösser als die im Beschaffungswesen geltende Toleranzschwelle von 5% ist (mit Irrtumswahrscheinlichkeit von 5%). Nein bedeutet umgekehrt, dass die Abweichung von der Toleranzschwelle statistisch nicht gesichert ist.
- › R^2 : Das R^2 ist ein Bestimmtheitsmass, das die Güte des Modells misst. Es bildet ab, wie gut die Lohndifferenzen durch die einbezogenen Variablen erklärt werden können (Vorhersagekraft/Erklärungsgehalt des Modells). Die Werte des R^2 liegen zwischen 0% und 100%. Ein R^2 von 73% bedeutet zum Beispiel, dass die Lohnungleichheit zu 73% durch Unterschiede in den Qualifikationsmerkmalen und arbeitsplatzbezogenen Merkmalen erklärt werden können.

ERGEBNISSE DER KONTROLLEN LOHNGLEICHHEIT										
Betrieb Nr.	Anzahl Mitarbeitende (MA)	Anzahl und Anteil Frauen	Brutto-Lohndifferenz in %	Geschlechts-spezifische Lohndifferenz nach Basisregression	Geschlechts-spezifische Lohndifferenz nach erweiterter Regression	signifikant von 0 verschieden	signifikant von 5% verschieden	R ²	Lohn-diskriminierung ja/nein	Kommentar
1	82	13 16%	-25.6%	-13.2%	-10.5%	ja	nein	73%	nein	Die Fallzahlen (insbesondere der Frauen) sind hier relativ klein. Es besteht der Verdacht der statistisch bedingten Unschuldsvermutung von kleinen Firmen.
2	107	13 12%	-9.7%	+4.5%	+6.3%	nein	nein	77%	nein	In diesem Fall ergibt sich eine (statistisch nicht signifikante) Lohndiskriminierung zu Ungunsten der Männer. Aufgrund der unterschiedlichen Qualifikations- und Arbeitsplatzmerkmale müssten die Frauen noch weniger als durchschnittlich 10% verdienen als die Männer. Gemäss detaillierteren Berechnungen ist dieses Ergebnis durch eine unterschiedliche Lohnstruktur von im Stundenlohn und im Monatslohn arbeitenden Personen erklärbar (Frauen arbeiten mehr auf besser entschädigter Stundenbasis). Bei getrennten Analysen der Lohngruppen ergibt sich eine Lohnungleichheit zu Ungunsten der Frauen, jedoch unterhalb der Toleranzschwelle.
3	187	77 41%	-26.6%	-22%	-9.2%	ja	ja	80%	ja	Getrennte Regressionsanalyse des Stammhauses und der Tochterfirma ergaben Einhaltung der Lohngleichheit im Stammhaus und eine Lohnungleichheit in der Tochterfirma, die statistisch von 0 aber nicht von der Toleranzschwelle verschieden ist.
4	3'941	772 20%	-24.4%	-8.1%	-2.2%	ja	n.a.	90%	nein	Eine Tochterfirma mit 1'200 MA konnte mangels Daten nicht in die Analyse einbezogen werden.
5	1'511	559 34%	-21.2%	-10.6%	-2.5%	ja	n.a.	87%	nein	Zu diesem Unternehmen wurden zwei Analysen durchgeführt. Die erste Analyse ergab keine Einhaltung der Lohngleichheit. Die zweite Analyse basierte auf einem veränderten Gesamtvergütungsmodell und einem veränderten Personalstamm.

6	233	67 29%	-28.6%	-12.2%	-1.0%	nein	n.a.	85%	nein	Starker Unterschied zwischen Basisregression und erweiterter Regression deutet auf Beschäftigungsdiskriminierung hin.
7	452	114 25%	-23.9%	-17.5%	-6.7%	ja	nein	83%	nein	Die Anforderungsniveaus wurden aus den Lohnstufen hergeleitet. Bei dieser Einstufung kann nicht ausgeschlossen werden, dass sie diskriminierungsfrei ist.
8	289	179 62%	-32.9%	-14.1%	-8.4%	ja	nein	82%	nein	Bei einer separaten Analyse ohne Geschäftsleitende ergibt sich eine höhere Lohndiskriminierung von 9.6% statistisch signifikant verschieden von 5%.
9	3'153	531 17%	-22.5%	-13.4%	-4.2%	ja	n.a.	88%	nein	Das Merkmal Anforderungsniveau korreliert stark mit dem Lohn und der Funktionsstufe. Wie im Fall 7 kann daher nicht garantiert werden, dass das Anforderungsniveau diskriminierungsfrei bestimmt wurde.
10	508	29 6%	-13.8%	(-4.9%)	-1.2%	nein	n.a.	81%	nein	Es fehlten Angaben zum Ausbildungsniveau der MA. Zum Teil wurden diese nacherhoben und eine erneute Analyse damit durchgeführt. Diese zeigte eine etwas stärkere Lohndiskriminierung, jedoch tiefer als die Toleranzschwelle auf (4.2%). Dies lässt vermuten, dass Frauen in Bezug auf ihr Bildungsniveau relativ zu den Männern tiefere Löhne beziehen.
11	2'780	1'117 40%	-36.7%	-19.2%	-5.0%	ja	n.a.	85%	nein	Der Arbeitsort ist zusätzlich in der Analyse eingeschlossen. Dadurch ist die Lohnungleichheit grösser als vorher ausgefallen.
12	143	29 20%	-27.2%	-2.4%	1.5%	nein	n.a.	92%	nein	Unter Berücksichtigung der lohnrelevanten Merkmale verdienen Frauen 1.5% mehr als Männer. Der starke Anteil erklärter Lohndifferenz zeigt, dass Frauen trotz ihrer Qualifizierung nicht in höhere Stellungen gelangen
13	4'912	385 8%	-8.6%	(-4.9%)	-10.6%	ja	ja	66%	ja	Das Merkmal Ausbildung wurde mangels verfügbarer Daten nicht einbezogen. Die Lohndifferenz wird bei Kontrolle der übrigen lohnrelevanten Merkmale grösser. Erklärung: Frauen verfügen bei den kontrollierten lohnrelevanten Merkmalen über eine bessere Ausstattung als die Männer.
14	78	4 5%	-17.4%	-17.1%	-8.2%	ja	nein	91%	nein	Anzahl der Frauen ist für die Analyse eigentlich zu klein. Ein deskriptives Matching von Einzelfällen deutet darauf hin, dass die Lohnungleichheit nicht erfüllt ist.

Tabelle 2: Legende: k.A. = Keine Angabe; n.a. = nicht anwendbar. Wenn das Ergebnis bereits unter der Toleranzschwelle liegt, macht es keinen Sinn, die signifikante Abweichung zur Toleranzschwelle auszuweisen.

Die Ergebnisse aus den Standardanalysen lassen sich folgendermassen kategorisieren:

Unternehmen mit geschlechtsspezifischer Lohnungleichheit über der Toleranzschwelle

a) Lohndifferenz statistisch signifikant grösser als die Toleranzschwelle

- › Zwei von 14 Kontrollen (Unternehmen 3 und 13) kommen zu dem Ergebnis, dass das Gebot der Lohnungleichheit gemäss Art. 3, Abs. 2 des Gleichstellungsgesetzes bzw. Art. 8, Abs. 3 der Bundesverfassung (gleicher Lohn für gleichwertige Arbeit) nicht eingehalten ist und die im Rahmen des Beschaffungswesens des Bundes geltende Toleranzschwelle überschritten wird. Die geschlechtsspezifische Lohnungleichheit nach erweiterter Regression liegt in diesen Fällen signifikant höher als die Toleranzschwelle von 5% (U3: -9,2%; U13: -10,6%).

b) Lohndifferenz statistisch nicht signifikant grösser als die Toleranzschwelle aber grösser als 0

- › Bei vier Unternehmen (Unternehmen 1, 7, 8 und 14) ergab sich eine geschlechtsspezifische Lohndiskriminierung von über 5%, welche statistisch signifikant (gegenüber 0) ist, jedoch nicht statistisch signifikant gegenüber der im Beschaffungswesen geltenden Toleranzschwelle (U1: -10,5%; U7: -6,7%; U8: -8,4%; U14: -8,2%). Das heisst, dass in diesen Fällen zwar eine geschlechtsspezifische Lohnungleichheit zu Ungunsten der Frauen vorliegt, die im Rahmen des Beschaffungswesens geltende Toleranzschwelle jedoch eingehalten wird. Mit anderen Worten ist gemäss der heute geltenden Praxis das Ergebnis aus Sicht des Vollzugs von Art. 8 Abs. 2 lit.c BöB tolerierbar, wenn auch nicht befriedigend (vgl. Strub 2005).

c) Lohndifferenz weder statistisch grösser als die Toleranzschwelle noch grösser als 0

- › In einem Fall (Unternehmen 2) übersteigt die geschlechtsspezifische Lohnungleichheit (in diesem Fall zu Ungunsten der Männer) die Toleranzschwelle von 5% (+6,3%), diese Differenz ist jedoch nicht signifikant. Das bedeutet, dass die aufgrund der Regressionsanalyse keine Lohndiskriminierung (zu Gunsten oder zu Ungunsten der Frauen) festgestellt werden kann.

Kontrollen mit geschlechtsspezifischer Lohnungleichheit unter der Toleranzschwelle

- › Bei sieben Unternehmen ist die geschlechtsspezifische Lohnungleichheit kleiner oder gleich die geltende Toleranzschwelle von 5%, davon bei einem Unternehmen kleiner als 1%, bei zwei Unternehmen kleiner als 2%, bei zwei Unternehmen kleiner als 3% und bei zwei Unternehmen kleiner oder gleich 5%. In einem Unternehmen fiel die Lohndifferenz zu Ungunsten der Männer aus (1.5%).

a) Lohndifferenz statistisch signifikant grösser als 0

Bei vier dieser Unternehmen (Unternehmen 4, 5, 9, 11) mit einem Ergebnis unter der Toleranzschwelle ist die geschlechtsspezifische Lohndifferenz signifikant von 0 verschieden. Das heisst, dass diese vier Unternehmen den Lohnungleichheitsgrundsatz zwar mit einer Wahrscheinlichkeit über 95% (Signifikanzniveau 5%) verletzen, die im Beschaffungswesen geltende Toleranzschwelle jedoch unterschritten wird und damit die Vorgaben des Beschaffungswesens eingehalten werden.

b) Lohndifferenz nicht statistisch signifikant

Bei den übrigen drei Unternehmen (Unternehmen 6, 10, 12) kann die geschlechtsspezifische Lohnungleichheit statistisch nicht nachgewiesen werden, da sie nicht signifikant ist. Das heisst, mit einer Wahrscheinlichkeit von über 5% könnte dieses Ergebnis auch durch Zufall zustande gekommen sein.

Fazit:

Von den 14 durchgeführten Kontrollen weisen insgesamt 10 Unternehmen eine statistisch signifikante Lohnungleichheit auf. Von diesen 10 Unternehmen überschreiten allerdings nur zwei Unternehmen die im Beschaffungswesen geltende Toleranzschwelle. Die anderen 8 Unternehmen halten die Vorgaben im Beschaffungswesen ein, wobei vier Unternehmen unter der Toleranzschwelle liegen und vier Unternehmen nicht signifikant darüber. In den restlichen vier Fällen konnte keine statistisch signifikante geschlechtsspezifische Lohnungleichheit nachgewiesen werden.

Die teilweise durchgeführten vertieften Analysen bei einzelnen Kontrollen lassen einige Ergebnisse der Standardanalyse in Bezug auf die Einhaltung des Lohnungleichheitsgebots hinterfragen. Zu nennen sind hier folgende drei Kontrollen:

- › Im Unternehmen Nr. 2 kamen die Standardanalysen zu dem (nicht signifikanten) Ergebnis, dass Lohnungleichheit zu Ungunsten der Männer besteht. Vertiefte Analysen zeigten, dass dies auf unterschiedliche Lohnpraktiken bei im Stundenlohn und im Monatslohn angestellten Mitarbeitenden zurück zu führen ist. Durch separate Regressionsanalysen bei diesen Gruppen kehrte sich das Ergebnis um. Die Ergebnisse waren jedoch nicht höher als die Toleranzschwelle bzw. ebenfalls nicht signifikant. Ein durchgeführtes Matching-Verfahren zeigte dabei, dass bei den 13 Frauen eine Lohndifferenz zu Männern mit vergleichbaren Voraussetzungen von zwischen -0,4% und -13,6% bestehen. Dies deutet darauf hin, dass das Gebot der Lohnungleichheit in diesem Unternehmen teilweise nicht eingehalten ist (vgl. Kap. 3.2).
- › Im Unternehmen Nr. 14 konnte eine geschlechtsspezifische Lohndiskriminierung von über 5% mit dem Standardmodell nicht nachgewiesen werden. Die Fallzahlen der Frauen waren hier

sehr klein. Das daraufhin verwendete Matching-Verfahren deutet darauf hin, dass die Lohndifferenzen zwischen ähnlich qualifizierten und ähnlich im Unternehmen positionierten Männern und Frauen tatsächlich relativ gross sind. Die Ergebnisse des Matching-Verfahrens reichen allerdings nicht aus, um eine Verletzung des Lohnleichheitsgebots statistisch nachzuweisen. Das Verfahren wurde daher als Argumentationsunterstützung verwendet, um die Unternehmen mit konkreteren Ergebnissen auf das Ungleichgewicht in ihrem Unternehmen hinzuweisen.

- › Im Unternehmen Nr. 8 wurde zusätzlich kontrolliert, inwiefern das Ergebnis durch die überdurchschnittlich hohen Löhne der drei männlichen Geschäftsleitenden beeinflusst wurde. Es zeigte sich dadurch, dass in der restlichen Belegschaft das Lohnleichheitsgebot nach den Kriterien des Beschaffungswesens nicht mehr eingehalten würde. Allerdings gilt gemäss Gleichstellungsgesetz die Lohnleichheit aller Mitarbeitenden im Unternehmen, so dass das Unternehmen nur mündlich auf dieses separate Ergebnis hingewiesen wurde.

Bei einem Teil der kontrollierten Unternehmen konnte zwar keine Lohndiskriminierung der Frauen im engeren Sinne festgestellt werden. Die Ergebnisse der Basisregression und der erweiterten Regression lassen jedoch vermuten, dass in den Unternehmen Beschäftigungsdiskriminierung vorherrscht. Wenn sich im Rahmen der erweiterten Analyse die in der Basisregression festgestellte geschlechtsspezifische Diskriminierung stark abschwächt, deutet dies darauf hin, dass Frauen mit sonst gleichen Qualifikationsmerkmalen weniger guten Zugang zu höheren Positionen und Tätigkeiten mit höherem Anforderungsniveau haben. Ein solcher Verdacht ergibt sich aufgrund der Ergebnisse bei der Hälfte der Unternehmen, namentlich den Unternehmen 3, 5, 6, 7, 9, 11 und 14. Bei dieser Feststellung handelt es sich allerdings um eine Beobachtung, welche für die Kontrollen der Lohnleichheit keine Relevanz hat, da die Beschäftigungsdiskriminierung nicht Gegenstand der Bestimmungen im Rahmen des Beschaffungswesens ist.

5. BEURTEILUNG

In diesem Kapitel werden die Qualität und Vergleichbarkeit der Kontrollen sowie die Angemessenheit der Toleranzschwelle auf Basis der deskriptiven Auswertung der Kontrollen und der Interviews beurteilt. Voraussetzungen für die Vergleichbarkeit der Ergebnisse sind die Konsistenz in der Anwendung der Methode und die Qualität der Datengrundlagen. Daher erfolgt zunächst eine Beurteilung für diese beiden Aspekte. Anschliessend wird eine Beurteilung der Genauigkeit des Messinstruments getroffen und schliesslich die Angemessenheit der Toleranzschwelle diskutiert.

5.1. KONSISTENZ IN DER ANWENDUNG DER METHODE

Insgesamt kann bei den bisher durchgeführten Kontrollen von einer konsistenten Anwendung der Methode gesprochen werden. Die Fachpersonen hielten sich jeweils an die Vorgaben des Instruments und wendeten die Standardmethode konsequent an. Im Folgenden wird die konsistente Anwendung der Methode nach den verschiedenen Schritten beurteilt:

- › Umgang mit fehlenden Daten: Laut den Interviews haben sich die Fachpersonen bemüht, fehlende Daten bei den Unternehmen nachzuerheben. In den meisten Fällen ist es gelungen, die fehlenden Daten zu erhalten. Um die Konsistenz der Datenbasis zu erhöhen, könnte standardmässig der Dezember als Stichmonat eingeführt werden. Dadurch kann sichergestellt werden, dass alle Zulagen richtig erfasst werden.
- › Plausibilisierung: Aufgrund der Interviews und der Kontrollberichte scheint die Plausibilisierung sorgfältig durchgeführt worden zu sein. Das Instrument schreibt hier kein bestimmtes Vorgehen vor. Eventuell könnte die Datenplausibilisierung durch eine Checkliste (o.ä.) weiter standardisiert und damit hinsichtlich der Konsistenz verbessert werden. Als wichtiger Punkt bei der Plausibilisierung hat sich die Kontrolle der Zuteilung der Mitarbeitenden zu den Anforderungsniveaus bzw. die Überprüfung von abgeleiteten Variablen heraus gestellt. Inwiefern dies von den Fachpersonen systematisch überprüft wurde, lässt sich nicht hundertprozentig nachvollziehen.
- › Bereinigung: Die Bereinigung der Datensätze erfolgte im Grossen und Ganzen auf konsistente Art und Weise. Positiv zu würdigen ist, dass die Bereinigung in enger Absprache mit den Unternehmen und dem EBG erfolgte und die Ausschlüsse gut dokumentiert sind. Nicht ganz konsistent erfolgte das Vorgehen beim Umgang mit plausiblen Ausreissern. Hier wurde jeweils in Absprache mit dem EBG entschieden, ob diese in die Analyse einfließen sollen oder nicht. Eine Definition, was ein Ausreisser ist und was nicht, besteht für das Kontrollinstrument nicht. Das

kann dazu führen, dass das Ergebnis in einem Fall von Ausreissern beeinflusst wird und in einem anderen Fall – in dem die Ausreisser entfernt wurden – nicht. Gemäss einer Fachperson, die Expertisen erstellt und der statistischen Fachperson könnte hier ein standardisiertes Verfahren zum Zug kommen, welches einen einheitlichen Umgang mit Ausreissern gewährleistet (z.B. Cook-Distance-Test oder Festlegung „Verhältnis vom Ausreisser zum Standardfehler sollte nicht grösser als z.B. 3 betragen“). Nicht ganz einheitlich wurde bei unplausiblen Einstufungen der Anforderungsniveaus/Kategorien beruflicher Stellung sowie bei identifizierten abgeleiteten Variablen vorgegangen. In einzelnen Fällen wurden diese korrigiert, in anderen nicht, weil der Aufwand für die Unternehmen zu gross gewesen wäre (vgl. dazu Kap. 5.2).

- › Prüfung der Modellvoraussetzungen⁴: Inwiefern die Fachpersonen die Modellvoraussetzungen für die Regressionsanalysen systematisch geprüft haben, liess sich im Rahmen der Evaluation nicht hundertprozentig nachvollziehen. Laut den Interviews war dies der Fall. Die Ergebnisse dieser Prüfung sind jedoch nur bei einer Fachperson dokumentiert. Verletzungen der Modellannahmen können zur Folge haben, dass entweder der Schätzwert der geschlechtsspezifischen Lohndiskriminierung verzerrt oder das Ergebnis zur Signifikanz der Ergebnisse ungenau ist. Zeigen die Residuen z.B. systematische Tendenzen, so bedeuten die Regressionskoeffizienten nichts, auch wenn sie signifikant sein sollten. Wenn die Prüfung der Modellvoraussetzungen ergibt, dass die Annahmen verletzt sind, kann dies teilweise durch Transformationen der Variablen behoben werden. Solche Transformationen wurden bei den bisherigen Kontrollen nicht durchgeführt.
- › Anwendung der Standardmethode: In beinahe allen Fällen wurde das Verfahren der Standardmethode konsequent umgesetzt. Ausnahmen bilden zwei Unternehmen, bei denen für alle Mitarbeitenden die Variable Ausbildung fehlte und nicht nacherhoben werden konnte. In einem Fall (Unternehmen 10) wurde aus unserer Sicht das bestmögliche Vorgehen gewählt, um zu einem Ergebnis in Bezug auf die Einhaltung der Lohngleichheit zu kommen. Es konnte mit Hilfe der Angaben aus einem Teilkollektiv aus unserer Sicht nachvollziehbar begründet werden, dass die Lohngleichheit im Unternehmen eingehalten wird. Die notwendige Abweichung von der Methode hatte dennoch zur Folge, dass das Ergebnis in Bezug auf die geschlechtsspezifische Lohndiskriminierung nicht dem entspricht, welches bei Anwendung der Standardanalyse resultiert hätte. In dem anderen Fall (Unternehmen 13) kann das Ergebnis aufgrund der feh-

⁴ Bei Regressionsanalysen müssen verschiedene Annahmen überprüft werden, damit die Ergebnisse der Regression zuverlässig interpretiert werden können: Linearität der Beziehungen zwischen den unabhängigen Variablen und der abhängigen Variablen, Homoskedastizität (konstante Streuung der Residuen, Unabhängigkeit der Residuen, keine lineare Abhängigkeit zwischen den unabhängigen Variablen und Normalverteilung der Residuen).

lenden Angaben angezweifelt werden. Es ist denkbar, dass durch Einbezug der Ausbildungsvariable die festgestellte signifikante Lohndiskriminierung abnehmen würde.

- › Anwendung vertiefter Analysen: Durch das Instrument sind keine Kriterien definiert, nach denen vertiefte Analysen veranlasst werden. In der Beschreibung des Instruments heisst es: „Je nach Ergebnis der standardisierten Analysen bzw. um zusätzliche Erkenntnisse über die Lohnsituation in einer Firma zu gewinnen, kann es sinnvoll sein, die Analysen zu vertiefen“. Wenn diese offene Formulierung zur Folge hat, dass in einigen Fällen vertiefte Analysen durchgeführt werden und in anderen nicht, dann kann dies hinsichtlich der Gleichbehandlung der Unternehmen problematisch sein. Denn durch vertiefte Analysen kann das Ergebnis der Lohndiskriminierung verändert werden (z.B. durch Einbezug von Variablen, getrennte Analysen bei Mutter- und Tochterfirma etc.). Dabei kann das Ergebnis in beide Richtungen abweichen, wie zum Beispiel der Fall 11 beim Einbezug der Variablen Arbeitsort zeigt. Um die Konsistenz bei der Anwendung vertiefter Analysen zu verbessern, wären Kriterien sinnvoll, die definieren, unter welchen Bedingungen, welche vertiefte Analyse in Betracht zu ziehen ist (z.B. Einbezug einer Variable „Tätigkeit“ bei speziellen Branchen, Berücksichtigung von „Leistungs-löhnen“ durch eine Variable „Leistung“, separate Regressionsanalysen bei grossen Unternehmen, Analysen ohne Geschäftsleitungsebene bei stark abweichenden Löhnen etc.). Bei den bisher durchgeführten Kontrollen hätten aus unserer Sicht in Einzelfällen vertiefende Analysen durchgeführt werden können, um die Ergebnisse der Standardanalyse zu erhärten bzw. näher zu beleuchten. Ob dies das Ergebnis verändert hätte, ist schwer einzuschätzen.
- › Dokumentation der Ergebnisse: Die Ergebnisse der Regressionsanalysen wurden – einen Fall ausgenommen – standardmässig ausgewiesen. Die Dokumentation könnte durch die Darstellung von Detailanalysen und -ergebnissen im Anhang noch verbessert werden. Wünschenswert wäre, dass – wie in einem Fall gemacht – die Ergebnisse zur Prüfung der Modellvoraussetzungen sowie das detaillierte Ergebnis zur Signifikanz des Geschlechtskoeffizienten zusätzlich im Anhang aufgeführt werden.

Fazit:

Das Instrument zur Überprüfung der Lohngleichheit wurde bei den bisherigen Kontrollen insgesamt konsistent angewendet. Die Fachpersonen sind bei den Überprüfungen sorgfältig vorgegangen und haben sich – soweit möglich – an die Vorgaben des Instruments gehalten. Trotzdem könnte die Konsistenz durch eine standardisierte Anleitung bei der Plausibilisierung, die Definition von Ausreissern, die Definition von Kriterien für vertiefende Analysen und eine detailliertere Darstellung der Ergebnisse weiter verbessert werden.

5.2. QUALITÄT DER DATENGRUNDLAGEN

Die Qualität der Datengrundlagen bei den durchgeführten Kontrollen ist mehrheitlich als gut zu bezeichnen. In einigen Fällen war die Qualität jedoch aufgrund von fehlenden Daten, kleinen Fallzahlen oder unplausiblen Zuteilungen der Mitarbeitenden beeinträchtigt. Im Folgenden treffen wir eine Einschätzung darüber, welche Auswirkungen die teilweise mangelnde Qualität der Daten bei den durchgeführten Kontrollen gehabt haben könnte.

Fehlende Fälle

Bei den durchgeführten Kontrollen kam es selten vor, dass einzelne Mitarbeitende ausgeschlossen werden mussten, da ihre Daten falsch oder lückenhaft waren. In Einzelfällen wurde bei grossen Unternehmen ein verschwindend kleiner Teil von Mitarbeitenden ausgeschlossen, bei denen jedoch der Lohn bekannt war. Dies dürfte – auch gemäss der konsultierten statistischen Fachperson – keinen Einfluss auf das Ergebnis gehabt haben.

Kritischer zu beurteilen sind zwei Fälle, bei denen die Löhne des Topkaders fehlten. Solche fehlenden Fälle könnten – je nach Höhe des Lohnes und der Grösse des Unternehmens – das Ergebnis beeinflussen. Dies wurde in einem Fall überprüft, indem sehr hohe Löhne für diese Personen als Proxy eingesetzt wurden. Es zeigte sich, dass diese fehlenden Fälle das Ergebnis nicht beeinflusst hätten. In dem anderen Fall ist unklar, inwiefern sich die Lohnungleichheit zwischen Frauen und Männern bei Einbezug der obersten Führungsebene verändert hätte. Das Vorgehen im ersten Fall ist aus unserer Sicht geeignet, um den Einfluss fehlender Daten zu kontrollieren. Allerdings sollten gemäss der konsultierten statistischen Fachperson bei einem solchen Verfahren jeweils verschiedene Werte eingesetzt werden, um einerseits sicherzustellen, dass das Modell durch die Extremwerte nicht verzerrt wird. Andererseits soll der Einfluss verschiedener Werte auf die geschlechtsspezifische Lohndiskriminierung geprüft werden.

Fehlende Variablen

Bei zwei Kontrollen fehlten für alle Mitarbeitenden die Angaben zum Ausbildungsniveau. Dies beeinflusst das Ergebnis in Richtung einer höheren Lohndiskriminierung der Frauen, wenn die Frauen im Unternehmen im Durchschnitt schlechter ausgebildet sind. In einem Fall resultierte denn auch eine signifikante Lohndiskriminierung von über 5% der Toleranzschwelle (Unternehmen 13). Je nach Verteilung der Ausbildung nach Geschlecht könnte eine vollständige Analyse bei diesem Unternehmen nach unserer Einschätzung auch zum Ergebnis kommen, dass die Lohnungleichheit eingehalten ist. Dass das Ergebnis auch in die andere Richtung gehen kann, zeigt der andere Fall (Unternehmen 10), bei dem ohne das Merkmal Ausbildung eine tiefere geschlechtsspezifische Lohnungleichheit resultierte als bei einer späteren Überprüfung mit einem Teilkollektiv.

Nach unserer Einschätzung sind fehlende Variablen des Standardmodells kritisch zu bewerten, da sie einen grossen Erklärungsanteil haben und das Ergebnis in beide Richtungen beeinflussen können.

Unplausible Abgrenzungen der Anforderungsniveaus oder beruflichen Stellungen/abgeleitete Variablen

Gemäss den Fachpersonen werden die Anforderungsniveaus und berufliche Stellung in den Unternehmen teilweise nach unterschiedlichen Logiken eingeteilt. Insbesondere stelle sich bei den Unternehmen häufig die Frage, wo die Grenze zwischen Stufe 3 und 4 beim Anforderungsniveau gezogen werden soll. Die Schwierigkeiten und teilweise Willkür bei der Einteilung bergen gemäss den Fachpersonen die Gefahr, dass die Zuteilung der Mitarbeitenden bereits eine Diskriminierung beinhaltet. Daher sei es wichtig, sich die Systematik durch die Unternehmen erklären zu lassen. Laut einer Fachperson kann die Überprüfung der Einstufung sehr anspruchsvoll sein, insbesondere wenn die Tätigkeiten im Unternehmen stark segregiert sind. Bei solchen Fällen stelle sich die Frage, wie die Gleichwertigkeit dieser Tätigkeiten überprüft werden könne. Ein besonderes Problem bilden laut Fachpersonen auch Einteilungen der Anforderungsniveaus und der beruflichen Stellungen aufgrund von anderen Kategorisierungen der Unternehmen (abgeleitete Variablen, vgl. Kap. 3.1.). Bei diesen kann ebenfalls nicht ausgeschlossen werden, dass sie diskriminierungsfrei sind.

Unserer Einschätzung nach können die Abgrenzungen des Anforderungsniveaus oder der beruflichen Stellung die Ergebnisse unter Umständen stark beeinflussen. Wichtig ist daher erstens sicherzustellen, dass die Einteilungen diskriminierungsfrei sind und zweitens, dass die Gleichwertigkeit der unterschiedlichen Tätigkeiten auf einem Anforderungsniveau gegeben ist.

Bei den bisher durchgeführten Kontrollen könnte in einem Fall das Ergebnis durch die Einteilung des Anforderungsniveaus verwässert worden sein (Unternehmen 9). In diesem Fall wurde die Einteilung vermutlich aus den Funktionsstufen im Unternehmen abgeleitet, welche erstens bereits eine Geschlechterdiskriminierung beinhalten können und in die zweitens bereits andere Merkmale wie die Ausbildung eingeflossen sein können. Bei einzelnen anderen Unternehmen wurde dieser mögliche Bias im Vorhinein oder später (zum Teil) ausgeräumt (vgl. Kap. 3.1.). Bei den verbleibenden Unternehmen ist jedoch nicht klar, inwiefern diese auf diskriminierungsfreie Stufeneinteilungen kontrolliert wurden.

Kleine Fallzahlen

Einen weiteren Qualitätsmangel bilden kleine Fallzahlen bei einzelnen kontrollierten Unternehmen. Die Daumenregel von mehr als 50 Mitarbeitenden wurde zwar bei der Auswahl der Unternehmen in allen Fällen eingehalten, die Vorgabe der Mindestanzahl von zehn weiblichen Arbeitskräften jedoch in einem Fall nicht (Unternehmen 14) und in zwei Fällen nur knapp (Unternehmen 1 und 2). Zwei dieser drei Unternehmen bilden Grenzfälle, weil sie eine geschlechtsspezifische Lohnungleichheit von über 5% aufweisen und diese zwar signifikant von 0 abweicht, nicht aber von der Toleranzschwelle. Hier drängt sich der Verdacht auf, dass die fehlende Signifikanz aufgrund der kleinen Fallzahl resultiert. Bei kleinen Fallzahlen muss die effektive Lohnungleichheit zwischen Frauen und Männern umso grösser sein, damit diese statistisch signifikant nachgewiesen werden kann. Man spricht bei diesem Phänomen von der „statistisch bedingten Unschuldsvermutung kleiner Unternehmen“ (Strub 2005).

Möglichkeiten der Manipulation der Daten

Aus Sicht einiger Fachpersonen besteht prinzipiell immer die Möglichkeit, dass die Daten durch die Unternehmen so manipuliert werden, dass sie eine geringere Lohnungleichheit aufweisen. Eine Person schätzt die Gefahr eher gering ein, da in der Regel mehrere Personen (meist eine Person aus der HR-Abteilung und eine Person aus der Geschäftsleitung) an der Datenlieferung beteiligt seien. Gemäss dieser Person ist es jedoch heikel, die Daten, insbesondere das Anforderungsniveau, nochmals durch das Unternehmen ändern zu lassen, wenn unplausible Daten identifiziert werden. Zwei andere Fachpersonen sind der Meinung, dass die Methode ein gewisses Vertrauen in die Unternehmen voraussetze. Ein konkreter Manipulationsverdacht wurde durch die Fachpersonen nicht geäussert. Unserer Einschätzung nach kann die Manipulation der Daten nicht ausgeschlossen werden, die Gefahr dürfte jedoch nicht sehr gravierend sein, da die Manipulation auch einer intensiven Auseinandersetzung mit der Methode bedarf.

Fehlende Dokumentation der Daten

Eine Fachperson weist darauf hin, dass eine Dokumentation der Daten durch die Unternehmen von Vorteil sein könnte. Ohne Dokumentation habe man wenige Informationen zur Lohnpolitik der Unternehmen, z.B. unter welchen Bedingungen Boni gezahlt werden und ob es auch nicht-monetäre Lohnbestandteile gibt. Zudem habe man keine Angaben über die Grösse und Struktur des Unternehmens mit denen man die Datenbasis plausibilisieren könnte.

Fazit:

Die Qualität der Datengrundlagen ist mehrheitlich als gut zu bezeichnen. In Einzelfällen könnten die Ergebnisse der Analysen jedoch insbesondere durch fehlende Fälle, durch fehlende Variablen, durch die Handhabung der Einstufung der Mitarbeitenden in die Anforderungsniveaus und berufliche Stellung sowie durch die kleine Grösse des Unternehmens beeinträchtigt sein. Die Qualität der Datenbasis hängt stark von der Verfügbarkeit und dem Goodwill der Unternehmen ab. Teilweise führen aber auch Schwierigkeiten bei der Zuteilung der Mitarbeitenden zu den Anforderungsniveaus und der Kategorien der beruflichen Stellung zu Mängeln in der Datenqualität.

5.3. GENAUIGKEIT DES MESSINSTRUMENTS

In diesem Kapitel beurteilen wir, wie genau und zuverlässig das Instrument die geschlechtsspezifische Lohndiskriminierung in den Unternehmen abbilden kann. Damit sollen Aussagen über die Belastbarkeit der Ergebnisse der durchgeführten Kontrollen getroffen werden.

In der Statistik wird in Bezug auf die Genauigkeit einer Messung von „Güte“ und andererseits von der „Zuverlässigkeit“ der Messung gesprochen. Mit der „Güte“ der Messung ist die Genauigkeit gemeint, mit der das Instrument den Effekt des Geschlechts auf den Lohn misst bzw. wie gut das Modell die Realität abbildet. In dem Zusammenhang spricht man auch von interner Validität einer Messung. Mit der „Zuverlässigkeit“ der Messung ist die Freiheit von Messfehlern gemeint, mit der das Instrument den geschätzten Effekt erfasst.

Güte des Modells (Interne Validität)

Ein Indikator für die Güte des Instruments ist das Bestimmtheitsmass R^2 . Dieses misst das Verhältnis der durch die berücksichtigten Variablen erklärte Streuung der Löhne zur gesamten Streuung der Löhne und zeigt auf, wie gut das gesamte Modell die Lohnunterschiede im Unternehmen erklären kann. Ist das R^2 niedrig, so kann dies ein Indikator dafür sein, dass noch wei-

tere Erklärungen für die Lohnunterschiede existieren, das Modell also unvollständig ist. Ein unvollständiges Modell kann zu einer Verzerrung der Schätzwerte führen (Backhaus et al. 2000). Dies ist damit zu erklären, dass die Schätzwerte, beispielsweise der Einfluss des Geschlechts auf den Lohn, möglicherweise einen Teil des Effekts einer anderen Variable absorbieren, welche im Modell nicht berücksichtigt ist. Die Höhe des R^2 liefert einen Hinweis auf die Vollständigkeit des Modells. Aufgrund der Höhe kann jedoch nicht darauf geschlossen werden, ob noch weitere Erklärungsfaktoren eine Rolle spielen würden. So kann beispielsweise ein R^2 von 70% bereits das bestmögliche Modell sein und der Restanteil der unerklärten Varianz ist auf eine unsystematische Festlegung der Löhne durch die Unternehmen zurückzuführen. Umgekehrt ist es möglich, dass ein Modell mit einem R^2 von 90% durch weitere Variablen noch verbessert werden kann.

Da durch das Instrument die lohnbestimmenden Faktoren vorgegeben sind, kann es die Lohnpolitik im Unternehmen nicht immer hundertprozentig abbilden. Je nach Unternehmen oder Branche entspricht das Modell daher mehr oder weniger gut der Realität. Bei den durchgeführten Kontrollen ergaben sich für das R^2 Werte zwischen 66% und 92%. Dies sind im Vergleich zu sonstigen multivariaten Regressionsanalysen in den Sozialwissenschaften sehr hohe Werte, die für ein gutes Abbild der Realität durch das Modell und eine relativ hohe Vollständigkeit des Modells sprechen. Dennoch muss davon ausgegangen werden, dass das Modell eine Restunschärfe beinhaltet, die je nach Unternehmen unterschiedlich hoch ist.

Die Güte des Modells und damit die Genauigkeit des Ergebnisses könnte also in einigen Fällen allenfalls durch den Einbezug weiterer Variablen verbessert werden. Laut Fachpersonen ist dies durchaus vorgekommen. Variablen, die die Erklärungskraft des Modells zum Teil verbessern, seien beispielsweise der Arbeitsort, die Tätigkeit im Unternehmen, der Bezug von Schichtzulagen etc. Wichtig sei hierbei jeweils zu überlegen, ob die Faktoren bereits Lohndiskriminierungen beinhalten können. Kritisch sei es beispielsweise, firmeninterne Funktionsstufen einzubeziehen, da sich die Löhne häufig aus diesen Stufen herleiten. Ein Grenzfall stellt auch der Einbezug einer Variable „Tätigkeitsbereich“ dar. So werden beispielsweise Sekretariatsmitarbeitende systematisch schlechter bezahlt, weil sie z.B. auf dem Markt nicht denselben Wert haben wie Mitarbeitende aus anderen Tätigkeitsbereichen, obwohl die Tätigkeiten ein gleiches Anforderungsniveau haben.

Eine weitere Möglichkeit, die „Güte“ des Modells bzw. die Genauigkeit der Schätzergebnisse zu verbessern, besteht darin, das Modell über die Spezifikation der einzelnen Variablen zu optimieren. Auch wenn die Modellvoraussetzungen erfüllt sind (vgl. Kap. 5.1), kann es vorkommen, dass durch die Transformation von Variablen die Messgenauigkeit des Modells noch erhöht wer-

den kann. Laut der statistischen Fachperson sollte bei der Überprüfung der Modellannahmen immer auch geschaut werden, ob durch eine Transformation der Variablen das Modell optimiert werden kann. Gemäss der statistischen Fachperson und einer Fachperson, die Expertisen durchführt, besteht bei dem gegebenen Instrument noch zweierlei Optimierungspotenzial hinsichtlich der Messgenauigkeit:

- › Erstens sei die Ausbildungsvariable durch die ordinale Skalierung nicht optimal spezifiziert, da sie den Wertzuwachs einer Ausbildung vordefiniere. Mit einer Dummy-Variablen könnten mögliche Verzerrungen aufgrund dieser Spezifikation ausgeräumt werden.
- › Zweitens weise der vom Dienstalter abhängige Lohn nicht immer einen linearen Verlauf auf, sondern nehme mit höherem Dienstalter wieder ab. Dies könne durch einen zusätzlichen quadratischen Term des Dienstalters – wie bei der Variable potenzielle Erwerbsjahre – besser berücksichtigt werden.

Zuverlässigkeit der Messung

Ein Indikator für die Zuverlässigkeit der Messung ist die Signifikanz des Schätzwerts. Diese gibt an, mit welcher Wahrscheinlichkeit der Diskriminierungseffekt effektiv und nicht nur zufällig grösser als die Toleranzschwelle ist. Sie ist ein Mass für die Beweiskraft der Daten. Im Instrument ist eine Irrtumswahrscheinlichkeit von 5% (Signifikanzniveau) festgelegt, die als Entscheidungskriterium dient. Das heisst, dass ein Ergebnis nur als zuverlässig erachtet wird, wenn die Irrtumswahrscheinlichkeit weniger oder gleich 5% beträgt.

Die Signifikanz in Bezug auf die Toleranzschwelle wurde in den durchgeführten Kontrollen jeweils überprüft. Ein Problem dieses Ansatzes ist, dass auch signifikante Ergebnisse einen geringen Aussagewert haben können. So nimmt mit grossen Fallzahlen die statistische Power (Teststärke) zu, was zu signifikanteren Ergebnissen führt. Umgekehrt nimmt die Signifikanz bei kleineren Fallzahlen automatisch ab (Problem der statistisch bedingten Unschuldsvermutung kleiner Unternehmen, vgl. Kap. 5.2).

Laut der statistischen Fachperson könnte die Genauigkeit des Instruments durch die Verwendung eines Heteroskedastie-konsistenten Standardfehlers⁵ verbessert werden. Dadurch liesse

⁵ Heteroskedastie (auch Heteroskedastizität genannt) bedeutet, dass die Residuen der Regression bzw. die Abweichung der Werte von der geschätzten Regressionsgerade in Abhängigkeit der erklärenden Variable variieren. Damit ist eine Grundvoraussetzung der Regressionsanalyse, die Homoskedastizität (konstante Streuung der Residuen) verletzt. Dies führt einerseits zu Verzerrungen bei den Schätzwerten für die Regressionskoeffizienten und andererseits zu einer Verzerrung des Standardfehlers, welcher die Signifikanz des Schätzwertes misst.

sich standardmässig vermeiden, dass eine mögliche Verletzung der Homoskedastizitätsannahme zu Ungenauigkeiten bei der Signifikanz führt.

Ein weiterer Indikator für die Zuverlässigkeit der Messung ist die Stabilität des Modells in Bezug auf Ausreisser. Gemäss den Fachpersonen war die Sensitivität der Ergebnisse hier zum Teil hoch, insbesondere bei kleinen Fallzahlen. Die statistische Fachperson weist darauf hin, dass beim Thema Ausreisser immer ein Trade-off bestehe, da sie einerseits Informationsgehalt haben und andererseits das Ergebnis verzerren können. Die statistische Fachperson sowie eine Fachperson, die Expertisen durchführt, regen daher an, zu definieren, was als Ausreisser gilt und ein standardisiertes Verfahren beim Ausschuss von Ausreissern zu verwenden. Dazu können bestimmte Tests wie der Cook's Distance-Test dienen (vgl. Kap. 5.1).

Fazit:

Durch das Standardmodell kann die geschlechtsspezifische Lohndiskriminierung bereits relativ genau geschätzt werden. Die Genauigkeit der Schätzung liesse sich jedoch über alternative Spezifikationen der einbezogenen Variablen und einem standardisierten Verfahren in Bezug auf Ausreisser weiter verbessern. Darüber hinaus können vertiefte Analysen, die über das Standardmodell hinausgehen, dazu beitragen, die Genauigkeit der Schätzung zu verbessern.

5.4. VERGLEICHBARKEIT DER ERGEBNISSE

Die Einschätzung der Vergleichbarkeit der Ergebnisse ist wichtig, um zu beurteilen, ob alle Unternehmen gleich behandelt wurden. Die Vergleichbarkeit der Ergebnisse der Kontrollen hängt von der Konsistenz in der Anwendung der Methode (vgl. Kap. 5.1), der unterschiedlichen Qualität der Datenbasis (vgl. Kap. 5.2) sowie der Genauigkeit der Messung (vgl. Kap. 5.3) ab.

Die Vergleichbarkeit der Ergebnisse wird aus unserer Sicht auf der einen Seite durch die standardisierte Methode gewährleistet. Es wird bei allen Unternehmen der gleiche Massstab angewendet (gleicher Lohn für gleichwertige Arbeit). Auf der anderen Seite haben die Ausführungen in den vorigen Kapiteln gezeigt, dass die Vergleichbarkeit der Ergebnisse teilweise aufgrund von Datenmängeln und leichten Unterschieden in der Handhabung des Instruments eingeschränkt sein kann. Während solche Inkonsistenzen unserer Einschätzung nach durch die vollständige Erhebung der Daten oder durch weitere standardisierten Vorgaben ausgeräumt werden können, erachten wir – basierend auf den Interviews – die drei folgenden Aspekte bei der Vergleichbarkeit der Ergebnisse als gravierender:

1) Grösse des Unternehmens bzw. statistisch bedingte Unschuldsvermutung kleiner Unternehmen

Die Grösse des Unternehmens hat einen Einfluss auf die statistische Unschärfe (Signifikanz) des Ergebnisses. Laut allen Fachpersonen bildet dies ein Problem für die Vergleichbarkeit der Ergebnisse, da kleine Unternehmen eher von der Verletzung des Lohnungleichheitsprinzips freigesprochen werden als grosse Unternehmen, wenn die geschlechtsspezifische Lohnungleichheit knapp über der Toleranzschwelle liegt. Diesem Problem war man sich auch bei der Entwicklung der Methode bewusst und das Signifikanzniveau wurde auf Basis der folgenden Überlegungen festgelegt: Es darf nicht zu klein gewählt werden, da sonst nur grosse Firmen verdächtigt werden; andererseits darf es nicht zu gross sein, weil dann sehr viele Firmen falsch verdächtigt würden. Um die Gleichbehandlung von grossen Firmen gegenüber kleinen Firmen zu gewährleisten, wurde daher das Signifikanzniveau auf 5%⁶ festgelegt (Strub 2004).

Bei der Frage, ob das Signifikanzniveau aus heutiger Sicht angepasst werden müsste, gehen die Meinungen der Fachpersonen auseinander:

- › Eine Fachperson ist der Meinung, dass eine flexible Anpassung nach Grösse des Unternehmens ein schlechtes Signal für die Unternehmen wäre. Im Extremfall könne dies dazu führen, dass Mitarbeitenden entlassen werden, um auf ein höheres Signifikanzniveau zu kommen. Das Signifikanzniveau habe also einen politischen Aspekt.
- › Eine Fachperson ist der Meinung, dass man über eine Erhöhung des Signifikanzniveaus bei kleinen Unternehmen (bis zu 100 oder 120 Mitarbeitende) nachdenken könne.
- › Eine Fachperson ist der Meinung, dass man bei einer flexiblen Anpassung des Signifikanzniveaus zunächst (z.B. anhand der Lohnstrukturerhebung) untersuchen müsste, wie sich die Signifikanzen nach Grösse verändern. Auch eine andere Fachperson spricht sich dafür aus, dass es in einem solchen Fall klare Vorgaben bräuchte und die flexible Anpassung nicht in der Hand der Fachpersonen liegen sollte. Aus Sicht dieser Fachperson haben klare Regelungen Vorrang, auch wenn dadurch die Gefahr bestehe, dass einzelne Unternehmen knapp durch die Kontrolle kommen.

⁶ Das Signifikanzniveau von 5% ist nicht zu verwechseln mit der Toleranzschwelle von 5%. Das Signifikanzniveau stellt die Irrtumswahrscheinlichkeit dar, d.h. mit einem Signifikanzniveau von 5% wird eine Wahrscheinlichkeit, dass das Ergebnis durch Zufall zustande gekommen ist, von 5% zugelassen. Die Toleranzschwelle betrifft das Ergebnis selbst. D.h., dass mit einer Toleranzschwelle von 5% eine geschlechtsspezifische Lohnungleichheit von weniger als 5% toleriert wird.

- › Die statistische Fachperson warnt vor einer flexiblen Anpassung des Signifikanzniveaus. Das Signifikanzniveau habe eine starke juristische Komponente, da es festlegt, bei wie viel Prozent der durchgeführten Kontrollen in Kauf genommen wird, dass die Unternehmen aufgrund der Kontrolle fälschlicherweise „verurteilt“ werden. Vor diesem Hintergrund erscheine 5% bereits sehr hoch.

2) Branchen- und betriebsspezifische Besonderheiten

Die strikten Vorgaben des Modells können gemäss mehreren Fachpersonen auch limitierend auf die Vergleichbarkeit der Ergebnisse wirken. So nehme mit der Grösse des Unternehmens auch die Heterogenität im Unternehmen (verschiedenen Standorte, Märkte etc.) zu. Mit der Heterogenität steige die Gefahr, dass das Modell die Lohnstrukturen im Unternehmen nicht mehr gut abbilde und das Ergebnis dadurch verzerrt sei. Bezüglich der Berücksichtigung der betriebsspezifischen Besonderheiten warnen einzelne Fachpersonen davor, zu stark vom Gesetz bzw. vom Lohnleichheitsgrundsatz abzuweichen. Der Einbezug weiterer Variablen oder die Einführung von mehreren Funktionsstufen müssten gut begründet sein.

Gemäss einer Fachperson ist durch die Standardmethode insbesondere auch die Vergleichbarkeit zwischen Branchen eingeschränkt. In heterogenen Branchen könnte die Lohnstruktur z.B. durch vier Anforderungsniveaus schlechter abgebildet werden als in homogenen Branchen. Berücksichtige man dies nicht, so würde bei inhomogenen Branchen eine stärkere Diskriminierung gemessen als bei homogenen Branchen.

3) Unterschiedliche Handhabung bei der Optimierung des Modells sowie bei der Anwendung vertiefter Methoden

Zur Optimierung des Modells (vgl. Kap. 5.3) und bei der Anwendung vertiefter Analysen lässt das Instrument relativ viel Freiraum, d.h. es existieren keine klaren Vorgaben. Aus statistischer Sicht kann die Genauigkeit der Schätzergebnisse jedoch beeinflusst werden, je nachdem wie gut die einbezogenen Variablen spezifiziert sind und ob vertiefte Analysen durchgeführt werden (z.B. separate Regressionsanalysen verschiedener Divisionen, Einbezug von Variablen, etc.). Die Vergleichbarkeit der Ergebnisse ist höher, je genauer das Modell an die Gegebenheiten des Unternehmens angepasst ist. Das bedeutet, dass die Vergleichbarkeit steigt, je tiefer man bei den Analysen geht. Bei den bisher durchgeführten Kontrollen wurden in vielen Fällen vertiefte Analysen durchgeführt. In keinem Fall haben diese vertieften Analysen zu einem belastbaren Ergebnis geführt, welches die Entscheidung, ob Lohndiskriminierung vorherrscht oder nicht, beein-

flusst hätte. Es ist jedoch durchaus vorstellbar, dass es in Zukunft zu einem solchen Fall kommen kann (vgl. Kap. 5.1).

Fazit:

Die Vergleichbarkeit der Ergebnisse wird auf der einen Seite durch die standardisierte Methode zum Teil sichergestellt, wobei sie durch eine noch konsistentere Anwendung und bessere Datenqualität weiter optimiert werden könnte. Auf der anderen Seite ist die Vergleichbarkeit durch die einheitlichen Vorgaben des Instruments auch eingeschränkt, weil es nicht auf betriebsspezifische oder branchenspezifische Besonderheiten eingeht. Ein davon abhängiges Problem ist die geringere statistische Aussagekraft bei kleinen Unternehmen, durch die kleinere Unternehmen bevorteilt sein können.

5.5. ANGEMESSENHEIT DER TOLERANZSCHWELLE

Die Toleranzschwelle wurde bei der Entwicklung des Modells mit der Begründung eingeführt, dass neben den Qualifikationsmerkmalen und arbeitsplatzbezogenen Merkmalen, die in die standardisierte Regression einfließen, weitere Unterschiede in der Qualifikation von Frauen und Männern oder andere objektive Erklärungsfaktoren zu Lohnunterschieden zwischen den Geschlechtern führen könnten. Es mache deshalb wenig Sinn, kleine Lohnungleichheiten weiter zu verfolgen, da diese möglicherweise durch weitere Erklärungsfaktoren gerechtfertigt werden könnten. Auf der andern Seite verlange das Gesetz, dass der Grundsatz gleicher Lohn für gleichwertige Arbeit eingehalten wird. Die Pilotphase bei der Entwicklung des Instruments kam zu dem Ergebnis, dass sich eine Toleranzschwelle von 5% bewährt hat (vgl. Strub 2004).

Nach den bisher durchgeführten Kontrollen beurteilen die Fachpersonen die Toleranzschwelle bei Anwendung des Standardmodells prinzipiell immer noch als angemessen, da mit dem Standardmodell nur eine grobe Messung gemacht würde. Nach Meinung mehrerer Fachpersonen könnte die Toleranzschwelle dennoch unter bestimmten Umständen angepasst werden:

- › Aus Sicht einer Fachperson könnte man die Höhe der Toleranzschwelle evtl. auf 3% senken mit der Begründung, dass das Instrument inzwischen bekannt sei. Die bisherigen Kontrollen würden zudem zeigen, dass die Resultate bei grossen Firmen zwischen 2% und 3% lägen. Falls bei den Analysen in die Tiefe gegangen werden soll, könne überlegt werden, ob die Toleranzschwelle weggelassen werden kann. Es stellt sich für diese Fachperson jedoch die Frage, wie bei kleinen Unternehmen mit einer heterogenen Mitarbeiterstruktur vorgegangen werden soll.

- › Auch aus Sicht einer anderen Fachperson könnte die Toleranzschwelle herunter gesetzt werden, wenn vertiefte Analysen durchgeführt werden. Weil immer eine gewisse Unschärfe bestehen bliebe, solle die Toleranzschwelle jedoch nicht ganz abgeschafft werden.
- › Eine Fachperson ist der Meinung, dass die Höhe der Toleranzschwelle evtl. nach Branche angepasst werden könnte. Damit wären Branchen mit inhomogenen Tätigkeitsgebieten weniger der Gefahr der Diskriminierung ausgesetzt. Ideal wäre auch, Referenzmärkte mit individuellen Toleranzschwellen zu kreieren, an denen man die Unternehmen messen würde. Dies ist aus Sicht der Fachperson jedoch schwer umsetzbar.
- › Eine andere Fachperson vertritt schliesslich die Ansicht, dass man über die Höhe der Toleranzschwelle streiten könne. Nach seinem Gefühl sind die 5% immer noch ein gutes Mass, evtl. könne man die Toleranzschwelle leicht auf 4% herunter setzen. Tiefer würde die Fachperson beim Standardmodell nicht gehen, da es Variablen geben könne, die 2-3% erklären können (z.B. innerbetriebliche Aus- und Weiterbildungen, Leistungslöhne etc.). Zudem beinhalte das Modell mit den geltenden Spezifikationen immer noch Unschärfen.

Einzelne Fachpersonen merken daneben an, dass sich die Toleranzschwelle aus strategischen Gesichtspunkten bewährt habe. Die Unternehmen würden das Instrument dadurch als fair empfinden und sich kooperativ zeigen.

Die statistische Fachperson hält die Toleranzschwelle für problematisch. So könne bei Unternehmen, die unter der Toleranzschwelle liegen, nicht ausgeschlossen werden, dass darunter Unternehmen sind, bei denen keine statistischen Unsicherheiten mehr bestehen bzw. das Modell bereits die Realität sehr gut abdecke. In solchen Fällen stelle die Toleranzschwelle eine politische Festlegung bzw. einen Kulanzwert dar, d.h. bei diesen Unternehmen wird eine statistisch gesicherte Lohndiskriminierung akzeptiert. Das Argument für die Einführung der Toleranzschwelle (s.o.) ist für die Fachperson nicht stichhaltig. Der Einbezug weiterer Variablen könne auch in die andere Richtung gehen, d.h. bei einem Modell, das die Realität nicht ganz abbildet, muss es nicht unbedingt heissen, dass in der Realität weniger Lohndiskriminierung vorhanden ist.

Fazit:

Die Toleranzschwelle ermöglicht es, mit einem pragmatischen Verfahren stärkere Verletzungen des Gebots der Lohngleichheit aufzudecken und dabei die Wahrscheinlichkeit gering zu halten, dass Unternehmen fälschlicherweise der Lohndiskriminierung beschuldigt werden. Unter diesem Gesichtspunkt ist die Toleranzschwelle positiv zu würdigen, wobei auch die Höhe von 5% eine angemessene Grössenordnung darstellte. Auf der anderen Seite könnten durch die Toleranzschwelle bei den bisherigen Kontrollen auch einzelne Unternehmen fälschlicherweise frei gesprochen worden sein. Um dies zu verhindern müssten das Standardinstrument und die Toleranzschwelle aus unserer Sicht angepasst werden.

Toleranzschwellen in anderen Anwendungsgebieten

Zu der Anwendung von Toleranzschwellen in anderen Bereichen wurde eine Web-Recherche durchgeführt. Es wurden dazu wenige Informationen gefunden. Insbesondere liessen sich keine Informationen zu Toleranzschwellen bei Instrumenten finden, welche auf Regressionsanalysen beruhen.

Bereiche, in denen Toleranzschwellen gelten, sind meist Messungen von Geräten, wie beispielsweise Geschwindigkeitsmessungen, akustische Messungen oder die Messung radioaktiver Strahlen bei nuklearmedizinischen Geräten. Diese Toleranzschwellen dienen der Kompensation von evtl. Messfehlern. Die Festlegung der Toleranzschwellen erfolgt auf Basis von Erfahrungswerten zu den Messungenauigkeiten, beispielsweise aus den statistischen Schwankungen einer Messreihe. Bei Geschwindigkeitsmessungen können die möglichen Messungenauigkeiten geräteintern vorkommen, sie können aber auch auf Fehlern des Bedienungspersonals beruhen. Lasermessungen sind beispielsweise genauer und daher die Toleranzen geringer als bei Radarmessungen oder beim Nach- oder Vorfahren. So liegt die erwartete Messungenauigkeit eines Radars bei 3%. In der Schweiz gelten für Radarmessungen Toleranzschwellen von 5km/h bis 100 km/h, 6km/h zwischen 101 und 150 km/h und 7 km/h ab 151 km/h (METAS, ASTRA 2010). Damit liegen diese Toleranzschwellen im Schnitt etwas höher als die erwartete Messungenauigkeit.

6. EMPFEHLUNGEN

Aufgrund der ausgeführten Beurteilung kommen wir zu den folgenden Empfehlungen:

1. Zur Verbesserung der Genauigkeit des Standardmodells bzw. der Belastbarkeit des Standardmodells, könnten folgende Massnahmen umgesetzt werden:
 - › Einheitliche Verwendung des Stichmonats Dezember,
 - › Erarbeitung einer standardisierten Anleitung für die Plausibilisierung (Checkliste o.ä.),
 - › Einheitliches Verfahren zur Bestimmung von Ausreissern,
 - › Standardmässige Verwendung einer Dummy-Variable statt einer ordinal skalierten Ausbildungsvariable, standardmässige Verwendung eines zusätzlichen quadratischen Terms bei der Variable Dienstalder; Oder: Pro Kontrolle optimierte Spezifikation der einzelnen Modellvariablen. Die zweite Variante würde die Genauigkeit gegenüber der ersten Variante noch etwas erhöhen, wäre jedoch mit einem zusätzlichen Aufwand verbunden. Es müsste dazu auch ein standardmässiges Vorgehen (z.B. in Form einer Checkliste) definiert werden,
 - › Standardmässige Verwendung eines Heteroskedastie-konsistenten Standardfehlers.
2. Zum Umgang mit fehlenden Daten empfehlen wir, bei fehlenden Variablen stärkeren Druck auf die Unternehmen auszuüben. Dazu fehlen dem EBG bisher die juristische Grundlagen und Sanktionsmöglichkeiten. Bisher ist die Einhaltung der Lohngleichheit in der Selbstdeklaration und den AGB's der Unternehmen festgehalten. Die gesetzlichen Grundlagen könnten jedoch bezüglich Pflicht zur Datenlieferung präzisiert werden. Als second-best-Lösung sollte zumindest für einen gematchten Teil der Mitarbeitenden die Variable erhoben werden, um zu kontrollieren, in welche Richtung die Verzerrung aufgrund der fehlenden Variable geht. Dieses Verfahren kann allerdings nicht vergleichbare Ergebnisse in Bezug auf die Lohndiskriminierung liefern. Bei fehlenden Managerlöhnen sollte ebenfalls auf die Daten beharrt werden. Im Notfall können hohe Schätzwerte eingesetzt werden, wobei dies ebenfalls mit Verzerrungen verbunden sein kann. Daher sollte bei einem solchen Vorgehen immer verschiedene Schätzwerte mit deren Auswirkungen auf das Ergebnis getestet werden.
3. Die Überprüfung der Gleichwertigkeit bei der Zuteilung der Mitarbeitenden zu Anforderungsniveaus und beruflichen Stellungen ist ein ernst zu nehmender Punkt. Hier sollte standardmässig eine Dokumentation bei den Unternehmen angefordert werden oder die Logik bei der Zuteilung mit den Unternehmen diskutiert werden. Eine weitere Möglichkeit ist, ein Instrument zur geschlechtsunabhängigen Arbeitsbewertung/Funktionsanalyse

einzuführen (z.B. Abakaba oder VIWIV-Tool, vgl. Katz, Baitsch 1997). Dies wäre zwar mit einem erheblichen Mehraufwand verbunden, sie ist jedoch aus Gründen der Vergleichbarkeit der Zulassung von zusätzlichen Funktionsstufen vorzuziehen. Das EBG prüft zurzeit die Machbarkeit eines einfachen Arbeitsbewertungssystems für Unternehmen mit weniger als 200 Mitarbeitenden. Erste Ergebnisse liegen Ende 2011 vor.

4. Eine flexible Anpassung des Signifikanzniveaus nach Grösse des Unternehmens erachten wir nicht als empfehlenswert, da dies erstens aus juristischer Perspektive kaum rechtmässig wäre, zweitens die Akzeptanz und Verständlichkeit des Instruments bei den Unternehmen reduzieren würde und drittens eine „faire“ Festlegung der Signifikanzniveaus nach Grösse des Unternehmens in der Praxis kaum umsetzbar ist. Das Problem der statistisch bedingten Unschuldsvermutung kleiner Firmen muss unseres Erachtens in Kauf genommen werden. Kleinere Unternehmen, die davon „profitieren“ könnten, sollten trotzdem nicht von den Kontrollen ausgeschlossen werden, da das Instrument erstens immer noch Lohnungleichheiten aufdecken kann und zweitens zu einer Sensibilisierung bei diesen Unternehmen führt und dadurch ebenfalls positive Wirkungen auslösen kann. Zusätzlich sollten bei kleinen Unternehmen andere Instrumente eingesetzt werden. Dazu ist z.B. die Eignung des Matching-Verfahrens zu prüfen. Auch das Arbeitsbewertungsinstrument könnte bei kleineren Unternehmen ein geeigneter Ansatz sein, die Einhaltung der Lohngleichheit zu überprüfen.
5. Die Toleranzschwelle sollte aus unserer Sicht herabgesetzt werden. Erstens sollte die Sensibilisierung der Unternehmen in Bezug auf Lohngleichheit inzwischen gestiegen sein. Zweitens zeigen die zusätzlichen Erfahrungen zur Lohnungleichheit in Unternehmen, dass die Toleranzschwelle auch tiefer angesetzt werden kann. Drittens wurden bei den bisher durchgeführten Kontrollen bereits vertiefte Analysen durchgeführt, wodurch eine bessere Genauigkeit des Ergebnisses erzielt wird. Wir schlagen vor, die Toleranzschwelle auf 2%-3% herabzusetzen, unter der Bedingung, dass das Standardmodell mit den kleineren Verbesserungsmöglichkeiten angepasst wird und zusätzliche (vertiefte) Analysen als Standard (Einschluss zusätzlicher Variablen, getrennte Regressionsanalysen etc.) eingeführt werden. In begründeten Fällen können die Ausprägungen der Variablen des Modells besser auf das Unternehmen angepasst werden (z.B. mehr Stufen beim Anforderungsniveau). Falls möglich, sollte dies innerhalb von Branchen einheitlich gehandhabt werden. Bei den vertieften Analysen muss sichergestellt werden, dass die Anpassungen diskriminierungsfrei sind. Zudem sollte auch jeweils geprüft werden, ob sich die Erklärungskraft des Modells (adjustiertes R^2) durch diese Anpassungen verbessert.

Die Herabsetzung der Toleranzschwelle hätte erstens den Vorteil, dass weniger Unternehmen fälschlicherweise von der Diskriminierung freigesprochen werden. Zweitens würden die Ergebnisse mit dem Grundsatz, das bestmögliche Modell für die Unternehmen anzustreben, besser vergleichbar. Der Nachteil dieses Vorgehens ist, dass sich der Aufwand für die Kontrollen erhöht. Da jedoch bisher bereits ein relativ hoher Aufwand für die Standardmethode angefallen ist und sich zeigte, dass teilweise bereits ohnehin in die Tiefe gegangen werden musste, halten wir den Zusatzaufwand im Verhältnis zum Zusatznutzen für erträglich. Ein weiterer Nachteil ist, dass Einbussen in der Objektivität der Methode entstehen, da die Fachpersonen vermehrt eigene Handhabe haben. Daher erachten wir es als wichtig, Empfehlungen für die vertieften Analysen zu erarbeiten. Diese sollten aufführen, unter welchen Bedingungen welche vertiefenden Analysen durchgeführt werden sollten und welche möglichen Erklärungsfaktoren dabei diskriminierungsfrei bzw. kritisch sind. Zudem muss bei der Anwendung von vertieften Analysen festgelegt werden, welches Ergebnis bei der Entscheidung, ob Lohndiskriminierung vorliegt oder nicht, gültig ist. Grundsätzlich sollte jeweils das präziseste Ergebnis massgebend sein.

LITERATUR

Backhaus et al. 2003: Multivariate Analysemethoden. 10. Auflage, Springer Verlag.

Katz, Baitsch 1997: Lohngleichheit für die Praxis. Zwei Instrumente zur geschlechtsunabhängigen Arbeitsbewertung.

METAS, ASTRA 2010: Verkehrsmesstechnik. Fachbroschüre des Bundesamts für Metrologie METAS.

http://www.metas.ch/metasweb/Dokumentation/Publikationen/Download/Verkehrsmesstechnik_d.pdf

Strub 2004: Überprüfung der Einhaltung von Lohngleichheit zwischen Frauen und Männern bei Beschaffungen des Bundes. Bericht über die Pilotphase zur Umsetzung von Art. 8 Abs. 1 Bst.c des Bundesgesetzes über das öffentliche Beschaffungswesen.