



Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 2

Methodische Anmerkungen zur Revision des Gleichstellungsgesetzes

Autoren: Matthias Moelleney / Fritz Dolder

Beitragsarten: Beiträge

Rechtsgebiete: Arbeitsrecht

Zitiervorschlag: Matthias Moelleney / Fritz Dolder, Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 2, in: Jusletter 3. Juli 2017

Die Regressionsanalyse OLS weist bei der Beurteilung des Gender Wage Gap auf Betriebsebene neben den in Teil 1 besprochenen Problembereichen auch methodisch begründete Einfallstore für Unsicherheit auf. Für einen praktischen Einsatz auf Betriebsebene zeigen sich weitere Problembereiche wie beschränkter Anwendungsbereich für KMUs, Verständnisprobleme der Rechtsanwender, Notwendigkeit einer manuellen Nachprüfung von individuellen Verstößen gegen Art. 3 GIG. Für die individuelle Lohnfindung und Lohnvergleichung und eine prima facie Beurteilung ganzer Betriebe werden alternative quantitative Methoden untersucht, welche keine vergleichbaren Problembereiche wie die Regressionsanalyse aufweisen.

Inhaltsverzeichnis

- 2. Probleme des Einsatzes der multiplen Regression zur betrieblichen Lohnanalyse im Rahmen des GIG (Fortsetzung)
 - 2.4. Einfallstore für Unsicherheit
 - 2.4.1. Auswahl der Beurteilungskriterien (unabhängige Variablen x_i)
 - 2.4.2. Festlegen von Entscheidungsgrenzen (Cut-off values)
 - 2.4.3. EXKURS: Entscheidungsgrenzen auf der individuellen Ebene? Umsetzung von Art. 3 GIG im Einzelfall
- 3. Praktische Schwierigkeiten bei der Anwendung der Regressionsanalyse auf Betriebsebene
- 4. Alternative Lösungsansätze
 - 4.1. Grundregeln: Konsens und Transparenz
 - 4.2. Stufe 1: Auswahl und Inhalt der Beurteilungskriterien
 - 4.3. Stufe 1: Verfahren zur Auswertung der Lohndaten
 - 4.4. Kollektive Beurteilung des Gender Wage Gap auf Betriebsebene: Trotz allem?
 - 4.5. Stufe 2: Lohnbänder & Positionieren innerhalb der Lohnbänder
 - 4.6. Regressionsanalyse und alternative Lösungsvorschläge im Vergleich

2. Probleme des Einsatzes der multiplen Regression zur betrieblichen Lohnanalyse im Rahmen des GIG (Fortsetzung)

2.4. Einfallstore für Unsicherheit

[Rz 1] Wer auch immer in der empirischen Sozialforschung die multiple Regressionsanalyse zur Untersuchung multiattributiver Phänomene einsetzt, sollte sich bewusst sein, dass er mit einem *Modell* arbeitet, dessen Resultate in vielfacher Hinsicht durch die Autoren des Modells beeinflusst, im schlimmsten Fall: *manipuliert* werden können. Der Benutzer der Methode der multiplen Regressionsanalyse sollte nicht übersehen, dass die Methode auf einer Modellbeziehung zwischen einer Vielzahl unabhängiger Variablen x_i und der abhängigen Variablen y (hier: Lohn) beruht und dass diese Modellbeziehung von aussenstehenden Modellbauern / Autoren und nicht etwa von der Natur, der sozialen Wirklichkeit oder ähnlichen Mächten geschaffen worden ist. In besonders ausgeprägtem Mass ist der Benutzer eines derartigen Regressionsmodells bei der Auswahl der unabhängigen Variablen (Kriterien) x_i dem fachlichen Verständnis und den rechtspolitischen Vorstellungen und Problemlösungsansätzen des Modellbauers *ausgeliefert*.¹

Einfach ausgedrückt bedeutet dies, dass ein Modell zur multiplen Regressionsanalyse nur das misst, was der Modellbauer wissen bzw. demonstrieren möchte: Sie gibt nur «*Resultate im Sinne des Erfinders*». ² Daraus ergeben sich auch in der Untersuchung des *wage gender gap* durch Regressionsanalyse Einfallstore für die rechtspolitischen und ideologischen Prioritäten des Modellbauers. Diese Einfallstore liegen einerseits in der Auswahl der unabhängigen Variablen x_i und andererseits in der Festlegung von Entscheidungsgrenzen (*cut-off values*) für Entscheidungen über die Resultate der Regressionsanalyse.

Zudem sollte nicht übersehen werden, dass die multiple Regressionsanalyse auf der *kollektiven Ebene* ganzer Betriebe stattfindet. Sie sucht eine Entscheidungsgrenze festzulegen, anhand derer «verdächtige» von unverdächtigen Betrieben unterschieden werden und «verdächtige» Betriebe «aussortiert» werden (und gegebenenfalls mit rechtlichen Sanktionen belegt werden) können. Wie bereits in Abschnitt 2.2 erläutert, kann indessen dieses Aussortieren lediglich eine (\pm unverbindliche) *Vorstufe* bilden für die Entdeckung und Beurteilung von Verstössen gegen Art. 3 des Gleichstellungsgesetzes (GIG; SR 151.1) auf der *individuellen Ebene*.

2.4.1. Auswahl der Beurteilungskriterien (unabhängige Variablen x_i)

[Rz 2] Seit dem Beginn systematischer Studien zur Arbeits(platz)bewertung im deutschsprachigen Raum vor rund Hundert Jahren wurde mehr oder weniger stillschweigend davon ausgegangen, dass strukturierte Verfahren mit einer Mehrzahl von Beurteilungskriterien (multi-criteria Verfahren) zur Lösung dieser Aufgabe *unentbehrlich* sind. Dagegen war die Auswahl, Zahl, Inhalt und Gewichtung der zu bewertenden Attribute und damit der Kriterien für die Beurteilung Gegenstand von Diskussionen und Kontroversen in der Fachwelt, welche bis heute nicht abgeschlossen worden sind. In methodischer Hinsicht wird dabei unterschieden zwischen einer Bewertung des *Arbeitsplatzes* (Stellenbewertung) einerseits und einer Bewertung der *Person* des Arbeitnehmers/in andererseits, also der Ebene des sog. *Humankapitals*. Bei der Bewertung des Arbeitsplatzes wird zudem diskutiert, ob diese marktunabhängig oder abhängig vom Arbeitsmarkt ausgestaltet werden solle, d.h. ob Einflüssen des Arbeitsmarktes berücksichtigt werden sollen. Seit Gründung des REFA (Reichsausschuss für Arbeitsstudien) in Deutschland (1924) und dem Genfer

Schema (1950) hat sich immerhin ein fachlicher Konsens herausgebildet, die Beurteilungskriterien in fünf bis sechs Gruppen zu gliedern, welche teilweise dem Humankapital, teilweise der personenunabhängigen Arbeitsplatzbewertung (Stellenbewertung) zuzuordnen sind.³

[Rz 3] Im vorliegenden Zusammenhang des *wage gender gap* bilden *Zahl und Inhalt* der als Beurteilungskriterien verwendeten unabhängigen Variablen x_i das zentrale *Einfallstor für Unsicherheit, Willkür* und ideologische Präferenzen bei der multiplen Regressionsanalyse. Je nachdem, ob ein bestimmtes Attribut erfasst als unabhängige Variable x_i (Kriterium) verwendet wird oder im nicht erklärten bzw. nicht als erklärungs würdig eingestuften Rest des Regressionsmodells versinkt, zeigt das Resultat (zumindest optisch) einen unterschiedlichen Aspekt.

Auch unter rechtlichen Gesichtspunkten (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.3) erscheint die Zahl und der Inhalt der eingesetzten Variablen x_i alles andere als harmlos: Wird den unmittelbar an der Lohnfindung Beteiligten (Vertragsparteien, allenfalls Sozialpartner) die Auswahl von Zahl und Inhalt dieser unabhängigen Variablen x_i zugunsten eines vom Gesetzgeber einseitig erlassenen Standard-Analysemodells *entzogen*, so geht dies inhaltlich erheblich weiter als das bereits in Art. 3 GIG als zwingendes Gesetzesrecht ausgestaltete allgemeine Verbot der Lohndiskriminierung.

[Rz 4] Landesweite Harmonisierung der unabhängigen Variablen x_i für alle Betriebe?

Im Zuge der geplanten Revision des GIG (*Art. 13b* Methode der Lohnanalyse) wird angestrebt, das gleiche Set von Variablen x_i (Beurteilungskriterien) nicht nur im Anwendungsbereich des Bundesgesetzes über das öffentliche Beschaffungswesen (BöB; SR 172.056.1), sondern landesweit und flächendeckend als obligatorisch vorzugeben und damit auf *alle* untersuchten Betriebe anzuwenden.⁴ Dem steht der Wunsch der unmittelbar interessierten Vertragsparteien entgegen, die Variablen betriebsspezifisch auszuwählen und damit den Besonderheiten der verschiedenartigen Branchen, Regionen und Einzelbetriebe flexibel Rechnung zu tragen.

Eine uniforme Anwendbarkeit der gleichen Methode und des gleichen Regressionsmodells OLS auf alle Unternehmen erscheint schon aus einem formalen Grund *nicht ohne weiteres als wünschenswert*. Die geschlechtsspezifische Lohndiskriminierung ist ein inner-betriebliches Phänomen und daher nur im Einzelfall aufgrund der beobachteten tatsächlich ausbezahlten individuellen Löhne abschliessend zu beurteilen. Die gesamtwirtschaftliche Optik und eine uniforme landesweite Anwendung der gleichen Regressionsmethode OLS erscheint für die Lösung der praktischen normativen Aufgabe (Lohngleichheit nach Art. 3 GIG) deshalb von vornherein nicht als besonders effizient oder attraktiv.⁵ Als Alternative zu einem Anspruch auf *allgemeine* Anwendbarkeit eines einzigen Beurteilungsmodells bieten sich vielmehr *branchenspezifische* oder *regionsspezifische* Gruppen unabhängiger Variablen x_i für Regressionsmodelle an.⁶

[Rz 5] Unabhängigkeit der Variablen x_i :
Autokorrelation zwischen y und einzelnen x_i ?

Eine Übernähe zwischen zwei Variablen x_i (= Multikollinearität) soll vermieden werden, weil sie zu einer doppelten

Erfassung und deshalb zu einer Übergewichtung bestimmter Inhalte und Einflüsse führt und dadurch die Anwendung des Regressionsmodells verzerrt und dessen Resultate künstlich verändert. Aus dem gleichen Grund sollte eine Übernähe zwischen der abhängigen Variable y (Lohn) und *einer* der unabhängigen Variablen x_i (Autokorrelation) vermieden werden, um das Regressionsmodell nicht ineffizient zu machen: Eine einzige unabhängige Variable x_i würde in einer derartigen Situation genügen. So kann eine grosse inhaltliche Nähe einer unabhängigen Variablen x_i zu der abhängigen Variablen y (hier: Lohn) dazu dienen, den Wert des Determinationskoeffizienten R^2 wunschgemäss zu erhöhen. In der *Giftliste* von DIETER URBAN / JOCHEN MAYERL wird dies unter der zynischen *Empfehlung S5* zusammengefasst.⁷

Derartige Übernähen zwischen unabhängigen Variablen bzw. unabhängigen und abhängigen Variablen können auf einfache Weise diagnostiziert werden, indem die bivariaten Korrelationen (Korrelationskoeffizient R nach *Pearson*) zwischen den beiden beteiligten Variablen berechnet werden. Bewegen sich diese Korrelationen in einem Bereich von vielleicht $> + 0.7$, so sollte das Modell überdacht und allenfalls angepasst werden. Im Kontext des Standard-Analysemodells EBG besteht aufgrund der Umschreibung des Inhalts der betreffenden Variablen ein leichter Verdacht, dass zwischen der abhängigen Variablen (Lohn y) und der unabhängigen Variablen «*betriebliche Stellung*» eine schwache Autokorrelation bestehen könnte, welche im Einzelfall die Regressionsanalyse dominieren könnte. Ebenso besteht ein leichter Verdacht einer hohen Korrelation zwischen den beiden unabhängigen Variablen «*betriebliche Stellung*» und «*betriebliches Kompetenzniveau*» (ab 2012).⁸ Diese Verdachtsmomente sollten jeweils im betrieblichen Einzelfall vor einer Regressionsanalyse untersucht und durch die Berechnung der bivariaten Korrelationskoeffizienten bestätigt oder verworfen werden.

[Rz 6] Inhalt der unabhängigen Variablen x_i : Humankapital oder Arbeitsplatz?

Das Standard-Analysemodell EBG verwendet als unabhängige Variablen x_i fünf Variable, von denen deren drei der *Person* des Arbeitnehmers/in d.h. dem sog. *Humankapital* zuzuordnen sind, und deren zwei zur personenunabhängigen Bewertung des *Arbeitsplatzes* dienen.⁹ Die vom Bund finanzierte Studie FELFE et al. (2015) unterstellt ohne weiteres die Relevanz dieser fünf Variablen x_i unter Berufung auf die (nicht näher bezeichnete) *wissenschaftliche Literatur*, welche vermutlich eher soziologische Studien auf der nationalen, nicht der einzelbetrieblichen Ebene betraf.¹⁰ Diese Anerkennung ging ohne weitere Erörterung auch in die Antwort des Bundesrates vom 18. November 2015 zur Beantwortung des Postulats Noser ein.¹¹ Es sollte indessen nicht übersehen werden, dass diese Auswahl der unabhängigen Variablen x_i die Elemente des *Humankapitals* und dadurch die Gesichtspunkte der Arbeitnehmerseite ohne erkennbare Begründung stärker gewichtet als diejenigen der personenunabhängigen *Stellenbewertung* (Arbeitsplatzbewertung).

[Rz 7] Einzelne Variablen x_i : Ausbildung: absolvierte oder notwendige?

Wie bereits an anderer Stelle angemerkt,¹² liegt es beim Festlegen des Inhalts der Variablen *Ausbildung* nahe, die für einen bestimmten Arbeitsplatz funktionell «*geforderte*» (notwendige, erforderliche) Ausbildung zu erfassen und diese dementsprechend als Element der Arbeitsplatzbewertung, *nicht* als Element des Humankapitals zu behandeln. Es dürfte dabei nicht der betrieblichen Wirklichkeit entsprechen, die tatsächlich absolvierte, vom Arbeitnehmer/in «*mitgebrachte*» Ausbildung als Element des Humankapitals als unabhängige Variable x_i einzusetzen (so das

Standard-Analysemodell EBG und *Logib*).

Der Einsatz über- oder unterqualifizierter Arbeitnehmer/innen an einem gegebenen Arbeitsplatz stellt für sich noch keinen Verstoss gegen Art. 3 GIG dar, denn das Risiko einer Überqualifizierung für einen Arbeitsplatz mit niedrigen Anforderungen (und daher niedrigem Lohn) muss vom Arbeitnehmer/in getragen werden. Dem steht unter dem Blickwinkel der Gerechtigkeit die Chance des Arbeitnehmers/in gegenüber, trotz Unterqualifizierung aus situationsbedingten Gründen einen Arbeitsplatz mit höheren Anforderungen (und dadurch höherem Lohn) zu erhalten.¹³ Die «mitgebrachte» Ausbildung gemäss *Logib* als eine Variable des Humankapitals zu erfassen, kommt im Ergebnis einer *staatlichen Garantie eines ausbildungskonformen Arbeitslohnes bzw. Arbeitsplatzes* gleich. Und dies kann der Staat mit seiner Rechtsordnung schon heute weder für Frauen noch für Männer leisten und wird er in Zukunft noch viel weniger leisten können (aufgrund der Verbesserungen im Bildungs- und Ausbildungsbereich).

[Rz 8] Es entbehrt in diesem Zusammenhang nicht einer gewissen Ironie, dass die Gesetzgebung im Personalrecht des Bundes (und der meisten Kantone) diese inhaltliche Umschreibung der Ausbildung im öffentlich-rechtlichen Bereich seit Jahren problemlos verwendet: Bei der Funktionsbewertung (Beispiel: VFA) wird ausdrücklich auf die vom Arbeitsplatz her *geforderte und notwendige*, nicht auf die vom Arbeitnehmer/in *mitgebrachte* Ausbildung abgestellt.¹⁴

[Rz 9] Einzelne Variablen x_j : *Anforderungsniveau des Arbeitsplatzes ab 2012: Betriebliches Kompetenzniveau*

Die unabhängige Variable *Anforderungsniveau des Arbeitsplatzes* des Standard-Analysemodells EBG wird mit einer Skalierung in vier Stufen erfasst, deren Umschreibung 2012 geringfügig geändert worden ist. Die Umsetzung dieser Skalierung dürfte im Einzelfall nicht eben einfach sein und viel Raum für persönliches Ermessen eröffnen: Insbesondere dürfte die Abgrenzung der beiden oberen Stufen Schwierigkeiten bereiten: Wo liegt der Unterschied zwischen «höchst komplexen» und «komplexeren» Aufgaben? Zudem dürfte die Unterscheidung zwischen «Entscheidungsaufgaben» und «technischen oder praktischen Aufgaben» Schwierigkeiten bereiten: Alle betrieblichen Aktivitäten führen bei genauer Betrachtung zu Entscheidungen.

Im Rahmen der fünf unabhängigen Variablen des Standard-Analysemodells EBG besteht zudem eine leichte Vermutung hohen Korrelation und damit einer Abhängigkeit zwischen der Variablen *Betriebliches Kompetenzniveau* mit den Variablen für die *Ausbildung* und für die *berufliche Stellung*. Wiederum sollte diese Vermutung im Einzelfall vor der Regressionsanalyse der Lohndaten eines Betriebs durch Kontrolle der bivariaten Korrelationskoeffizienten zwischen den betreffenden Variablen ausgeräumt bzw. bestätigt werden.¹⁵

[Rz 10] Einzelne Variablen: *Berufliche Stellung*

Die unabhängige Variable *Berufliche Stellung* des Standard-Analysemodells EBG wird mit einer Skalierung in fünf Stufen erfasst. Auch bei dieser Variablen dürfte die Umsetzung der Skalierung im Einzelfall nicht gerade einfach sein: Insbesondere dürfte die Abgrenzung der unterschiedlichen Kaderfunktion Schwierigkeiten bereiten: Wo liegt der Unterschied zwischen oberstem, oberem, mittlerem, unterem und unterstem Kader?

Im Rahmen der fünf unabhängigen Variablen des Standard-Analysemodells EBG besteht zudem erneut eine schwache Vermutung, dass von Fall zu Fall zwischen der Variablen *Berufliche Stellung* und der abhängigen Variablen *y* (Lohn) eine (geringe) Autokorrelation sowie zwischen den beiden Variablen *Berufliche Stellung* und *Betriebliches Kompetenzniveau* eine (schwache) Abhängigkeit bestehen könnte. Auch diese beiden Vermutungen sollten jeweils vor der Auswertung der Lohndaten eines Betriebs durch Kontrolle der bivariaten Korrelationskoeffizienten zwischen den beiden fraglichen Variablen ausgeräumt oder bestätigt werden.¹⁶

[Rz 11] Einzelne Variablen: *Potenzielle Erwerbsjahre*

Es ist schon im Vernehmlassungsverfahren kritisiert worden, dass beim Standard- Analysemodell EBG die Variable der (potenziellen) Erwerbsjahre aufgrund einer theoretischen Überlegung ermittelt werden, wobei vom Lebensalter des betreffenden Arbeitnehmers/in die Ausbildungsjahre (auf dem höchsten «mitgebrachten» Ausbildungsabschluss basierend) sowie sechs Vorschuljahre subtrahiert werden. Dabei ist einmal daran zu erinnern, dass nur diejenigen Ausbildungsjahre berücksichtigt werden sollten, welche für den betreffenden Arbeitsplatz tatsächlich notwendig sind, nicht die mitgebrachten Ausbildungsjahre des betreffenden Arbeitnehmers/in.¹⁷

Zudem behandelt die Definition dieser potenziellen Erwerbsjahre Frauen anders als Männer, da die *tatsächlichen* Erwerbsjahre nach wie vor bei Frauen tendenziell niedriger ausfallen, weil sie nach der Geburt von Kindern häufig eine Familienpause einschalten. Nach der Definition des Standard-Analysemodells EBG würden derartige erwerbsbiografische Unterschiede (Erziehungsjahre) *voll als Erwerbsjahre* angerechnet. Dies dürfte indessen die Realität des Arbeitsmarktes nicht korrekt abbilden, da für die Lohnfindung in den Betrieben (bis heute) *ausschliesslich die tatsächlich erworbenen*, nicht aber irgendwelche potenziellen Berufserfahrungen entscheiden.¹⁸

[Rz 12] Einzelne Variablen: Dienstjahre: Interne oder (auch) externe?

Bei der Variablen Dienstjahre ist ebenfalls bereits im Vernehmlassungsverfahren kritisiert worden, dass beim Standard-Analysemodell EBG für die Messung nur die *innerhalb* des Unternehmens erarbeiteten Dienstjahre zählen, die Erfahrung bei *anderen* Arbeitgeber/innen an dieser Stelle aber nicht berücksichtigt werden sollen.¹⁹ Dies dürfte insoweit an der betrieblichen Realität vorbeigehen, als die Arbeitgeber/innen unter anderem *gerade daran interessiert* sind, *externe Erfahrung* aus anderen Firmen zu erwerben («einzukaufen»), wenn sie neue Mitarbeiter rekrutieren.

Beispiele: Flugstunden von Piloten, Fahrkilometer im Transportgewerbe, Zahl von Operationen bei spezialisierten Chirurgen, Zahl von Publikationen und Lehrerfahrung (Zahl der gehaltenen Vorlesungen, Seminare usw.) im wissenschaftlichen Betrieb.

Nach dem Regressionsmodell des EBG dürften derartige ausser-betriebliche Berufserfahrungen bei der Lohnfindung *nicht* berücksichtigt werden. Demzufolge würden neue Mitarbeiter/innen, welche *gerade* aufgrund ihrer *externen Berufserfahrung* rekrutiert («eingekauft») worden sind, nach dem Modell EBG für die betreffende Zeit keine anrechenbaren Dienstjahre haben.

[Rz 13] Einzelne Variablen: Die Vergessenen

Bei den unabhängigen Variablen des Regressionsmodells EBG werden die Kriterien der *psychischen Belastung und körperlichen Anforderungen* des Arbeitsplatzes als Elemente der Lohnfindung nicht berücksichtigt. In der analytischen Arbeits(platz)bewertung ist es spätestens seit dem *Genfer Schema* von 1950 unbestrittener Bestandteil des *Rumpf-Konsenses der Fachwelt*, dass derartige körperliche und psychische Anforderungen bei der Arbeits(platz)bewertung und damit in der Lohnfindung berücksichtigt werden. Diese Kriterien sind auch in das Beurteilungsraster von ABAKABA (CHRISTIAN KATZ / CHRISTOF BAITSCH 1996 / 2006) übernommen und verfeinert worden.²⁰

Selbst die vom Bund finanzierte Expertise FELFE et al. (2015) weist auf diese bisher nicht näher begründete Lücke im Kriterienraster des Regressionsmodells EBG hin und empfiehlt deren Eliminierung.²¹ In seiner Antwort Noser vom 18. November 2015 hat der Bundesrat zwar zunächst den verwendeten Raster der unabhängigen Variablen des Regressionsmodells EBG verteidigt.²² Die Anregung der Studien FELFE et al. (2015) zur Prüfung und *Aufnahme weiterer unabhängiger Variablen* in das Regressionsmodell EBG hat der Bundesrat in seiner Antwort Noser jedoch übernommen – wenn auch mit spürbarer Abneigung.²³

Es entbehrt unter diesen Umständen erneut nicht einer gewissen Ironie, dass die Gesetzgebung im Personalrecht des Bundes (und der meisten Kantone) derartige Überlegungen für die Lohnfindung im öffentlich-rechtlichen Bereich seit Jahren vollzogen und das Genfer Schema von 1950 insoweit übernommen und verfeinert hat. Bei der in diesem Bereich eingesetzten Funktionsbewertung (beispielsweise gemäss VFA) werden die *psychische Belastung und körperliche Anforderungen* ohne irgendwelche Einschränkungen als unabhängige Variable eingesetzt.²⁴

[Rz 14] Zusammenfassung der Ergebnisse

Die angeführten Beispiele dürften ausreichend deutlich darlegen, dass die Auswahl und der Inhalt der unabhängigen Variablen x_i für ein lineares Regressionsmodell keineswegs harmlos ist, sondern ein zentrales Einfallstor für *Unsicherheit, Willkür und persönliche Präferenzen des Modellbauers* bilden. Diese Auswahl wirkt sich unmittelbar auf die Resultate der mit dem Modell ausgeführten Regressionsanalyse aus: Phänomene (Dichotomien), welche bei der Auswahl der Variablen nicht berücksichtigt worden sind, werden durch den Einsatz des betreffenden Regressionsmodells nicht dargestellt und sind deshalb in den Resultaten der Untersuchung nach aussen auch nicht erkennbar: Diese vergessenen Dichotomien versinken vielmehr im Sumpf des anonymen Unerklärten. Ein Regressionsmodell erfasst und demonstriert nur das, was der Modellbauer erfassen und demonstrieren möchte.

Insbesondere dürfte die Aufteilung der Aufgabe auf personenorientierte Kriterien des *Humankapitals* und auf personen-unabhängige Kriterien des *Arbeitsplatzes* (Anforderungen) kontrovers beurteilt werden. Deshalb erscheint es als aussichtsreicher, Beurteilungskriterien für einzelne Branchen, Regionen oder Einzelbetriebe auszuwählen und inhaltlich zu formulieren (durch Konsens der Beteiligten) statt sich um eine landesweit flächendeckend auf alle Betriebe anwendbare Auswahl von Kriterien zu bemühen.

Aufgrund der angeführten Beispiele dürfte im Ergebnis deutlich geworden sein, dass die Auswahl und die Inhalte der eingesetzten Beurteilungskriterien nicht durch kognitive Vorgänge gültig (verifizierbar,

falsifizierbar) bestimmt werden können. Deshalb wird es auch *niemals einen abschliessenden Konsens über Auswahl und Inhalt der für eine Arbeitsbewertung relevanten (notwendigen, zweckmässigen, effizienten) Kriterien geben.*

2.4.2. Festlegen von Entscheidungsgrenzen (Cut-off values)

[Rz 15] Regressionsanalyse und Entscheidungsgrenzen (cut-off values)

Der Soziologe interessiert sich in erster Linie für den Determinationskoeffizienten R^2 , also dafür, wie präzise sein OLS-Modell die beobachteten Daten abbildet: *goodness of fit*. Der an der normativen Umsetzung von Art. 3 GIG orientierte Beobachter dagegen interessiert sich für die Regressionskoeffizienten als Indikatoren für Lohnungleichheit innerhalb eines Betriebes. Dazu wird die Technik der Dummyvariablen Frau/Mann 1/0 eingesetzt als methodische Alternative zu der klassischen Zerlegung nach OAXACA/BLINDER (1973) oder anderen Verfahren.

Aber: Für die normative Umsetzung der Resultate der Regressionsanalyse OLS sind *Entscheidungsgrenzen (cut-off values)* erforderlich. Und da ist als selbstverständlich vorauszusehen, dass überall dort, wo derartige normative Entscheidungsgrenzen zur Auswertung der Resultate einer multiplen Regressionsanalyse festzulegen sind, Kontroversen und Konflikte zwischen den Nutzern eines Regressionsmodells vorprogrammiert sind. Dementsprechend wird zu allererst zu diskutieren sein, welcher Betrag (Beispiele: $\beta_5 > 0.1$ oder $R^2 > 0.6?$) die Grenze zwischen diskriminierend und nicht-diskriminierend bzw. zwischen verdächtig und nicht verdächtig eines betrieblichen Lohnsystems bildet. Und dies alles zur Beantwortung der Frage: Welche Werte sind bei der Beurteilung des Lohnsystems eines Betriebes im Hinblick auf die Konformität mit Art. 3 GIG «verdächtig», welche nicht? Von welchem beobachteten Wert an darf oder muss eine Diskriminierung vermutet oder angenommen werden?²⁵

Aber: Nachdem sich die Beteiligten einmal über den gewünschten Betrag/die gewünschte Zahl einer Entscheidungsgrenze geeinigt haben, wird es unvermeidlich weitere Diskussionen und Streit über die Grösse der Abweichungen zwischen diesem *pro futuro* festgelegten abstrakten *cut-off value* und dem tatsächlich beobachteten Wert (z.B. von β_5) zu geben. Die Signifikanz von β_5 (einschliesslich Toleranzschwelle) können zwar, soweit die mathematischen Voraussetzungen erfüllt sind, mit der t-Teststatistik beurteilt werden. Aber dabei wird es neue Diskussionen darüber geben, welches Signifikanzniveau bzw. welcher p-Wert (z.B. $p = 0.01, 0.05, 0.10$) für die Beurteilung der Signifikanz massgebend sein soll.

[Rz 16] Festlegen von Entscheidungsgrenzen (Cut-off Values) auf der Betriebsebene:

t-Test eines Regressionskoeffizienten β_5 mit «Toleranzschwelle S»

Wie bereits dargelegt, operiert die normativ orientierte multiple Regressionsanalyse auf der kollektiven Ebene ganzer Betriebe. Sie sucht eine Entscheidungsgrenze festzulegen, anhand derer «verdächtige» von «unverdächtigen» Betrieben unterschieden und verdächtige Betriebe «aussortiert» werden können. Wie ebenfalls bereits erläutert, sind aufgrund des beschränkten Informationswertes der OLS-Regressionsanalyse *false positive* und *false negative* Ergebnisse im Hinblick auf die Konformität mit Art. 3 GIG nicht auszuschliessen (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt

2.2).

- a. *False positive result*: Trotz Einstufung des Betriebs als «verdächtig» gibt es *keine* Verstösse gegen Art. 3 GIG: Falsche Verdächtigung, Beispiel «Wäscherei» (siehe Teil 1, Abschnitt 2.2).
- b. *False negative result*: Trotz Einstufung eines Betriebs als «nicht verdächtig» kommen *dennoch* punktuelle Verstösse gegen Art. 3 GIG vor: Falscher Freispruch. Beispiel: Grossbetrieb mit ansonsten hoher Compliance; aber bei der Gebäudereinigung oder der Wäscherei von Berufskleidung gibt es punktuelle Verstösse gegen Art. 3 GIG.

Demzufolge bildet das Aussortieren von Betrieben aufgrund einer OLS-Regressionsanalyse *höchstens eine (mehr oder weniger unzuverlässige) Vorstufe* für die rechtlich verbindliche Entdeckung und Beurteilung von Verstössen gegen Art. 3 GIG auf der individuellen Ebene einzelner Arbeitsverhältnisse.

[Rz 17] Die Entscheidungsgrenze für das Aussortieren «verdächtiger» Betriebe wird im Standard-Analysemodell EBG aus dem berechneten Regressionskoeffizienten β_5 der Dummyvariablen Frau/Mann (1:0) gewonnen, welcher zu diesem Zweck mit einer «Toleranzschwelle S» versehen wird. Dazu wird folgender Zusammenhang eingesetzt

$$T = (D - S) \div SE,$$

wobei

T Entscheidungsgrenze

D beobachteter Regressionskoeffizient β_5 (sog. «Diskriminierungskoeffizient»), berechnet aufgrund von N betrieblichen Lohndaten mit der multiplen Regressionsanalyse,

S «Toleranzschwelle», festgelegt *a priori* auf 0.050,

SE Standardfehler (*standard error*) von D (= β_5)

Der derart berechnete Wert von T wird mit der Student-Teststatistik (t-Test) auf Signifikanz getestet, wobei sich die Freiheitsgrade aus der Zahl N der beobachteten Datensätze der Arbeitnehmer/innen des Betriebs ergeben. Für diesen Signifikanztest wird ein Signifikanzniveau von $\alpha = 5\%$ ($F(z) = 0.95$) im einseitigen Test festgelegt. Erscheint der Wert von T in diesem t-Test als *signifikant*, entsprechend

$|T| \geq z$, so wird der fragliche Betrieb im Hinblick auf den *wage gender gap* als «verdächtig» qualifiziert.²⁶

Weitgehend *unkritisch* bleibt bei dieser Berechnung die Zahl N der beobachteten Arbeitnehmer/innen (= Freiheitsgrade im t-Test); sie entfaltet nur minimale oder gar keine Wirkung, da im hier interessierenden Bereich (ab N = 50 bis etwa 200) die Werte von z (Tabelle) nur noch minimal variieren:

	N: Anzahl Freiheitsgrade (= Beobachtungen = Arbeitnehmer)			
F(z)	50	100	20	∞
0.95	1.68	1.66	1.65	1.65

Dagegen wirken sich bei dieser Berechnung drei andere Grössen *kritisch* auf das Resultat aus: Der Standardfehler SE von β_5 , die Toleranzschwelle S sowie das ausgewählte Signifikanzniveau $\alpha = 5\%$ ($F(z) = 0.95$) des t-Tests.²⁷

[Rz 18] Wozu eine «Toleranzschwelle» S?
Regressionsanalyse – ein stumpfes Schwert?

Es werden für die Berechnung der Entscheidungsgrenze (cut-off value) zwei *a priori* Festlegungen getroffen, welche *nicht* auf wissenschaftlichen (kognitiven) Überlegungen beruhen und deshalb ebenfalls wichtige *Einfallstore* für Unsicherheit und Willkür bilden: Sowohl die «Toleranzschwelle» S als auch das Signifikanzniveau $F(z)$ sind kritisch für das Resultat.

Die Festlegung einer Toleranzschwelle S wird mit der Rücksicht auf Unvollkommenheiten des Verfahrens, insbesondere auf ein mögliches Missachten weiterer relevanter Beurteilungskriterien (unabhängige Variablen x_i) begründet. Dass diese Toleranzschwelle S keine wissenschaftliche (kognitive) Grundlage hat, räumt sogar der Bericht des Bundesrates (2015) auf die Anfrage Noser ein.²⁸ Zugleich wird eine Senkung dieser Toleranzschwelle in Aussicht gestellt, falls sich die Unvollkommenheiten des Verfahrens beseitigen lassen sollten, d.h. weitere relevante Beurteilungskriterien in das Verfahren eingeführt werden sollten.²⁹ Trotz dieser Unsicherheiten soll der in der beschriebenen Art adaptierte und auf Signifikanz geprüfte Wert von β_5 die *Entscheidungsgrenze* für die Qualifizierung der Betriebe als «verdächtig» oder «nicht verdächtig» bilden.³⁰

Es bleibt dabei unbekannt, inwiefern sich dieses S – wie angeführt – in der Praxis «bewährt» hat; es bleibt ebenfalls unbekannt, ob auch noch andere Werte für S getestet worden sind, beispielsweise 1% ($S = .010$).³¹ Dieses $S = .050$ scheint dem externen Beobachter *relativ gross* im Verhältnis zum Betrag von β_5 (nämlich rund die Hälfte von β_5), wenn man sich auf nationaler Ebene in einem Bereich von $\beta_5 = 11.2\%$ (.112, LSE 2000) und 8.7% (.087, LSE 2012) bewegt.³²

Eine kritische Anmerkung zu dieser «Toleranzschwelle» S liegt deshalb nahe: Man berechnet also β_5 -Werte für die Dummyvariable F/M, welche erfahrungsgemäss in einem Bereich zwischen 0.10 und 0.15 liegen, in dem Rechenbeispiel des Methodenbeschriebs 0.138. Dazu verwendet man ein mathematisches Instrument (multiple OLS-Regressionsanalyse, Logib) von hoher Präzision. Dennoch *misstraut* man der Validität (Treffsicherheit) des auf diese Art berechneten β_5 als Entscheidungsgrenze so sehr, dass man eine «Toleranzschwelle» S von 0.050 einführt, welche fast der Hälfte des Betrages von β_5 entspricht. Man verkleinert den Betrag von β_5 also jeweils um knapp die Hälfte und bestimmt schliesslich mit einem t-Test, ob der verbleibende Betrag T (noch) statistisch signifikant gegen Zufall ist. Dies erfolgt anscheinend zu dem Zweck, eventuelle *false positive results* zu vermeiden und damit Betriebe vor «falscher Verdächtigung» zu bewahren.

Aber: Das Schwert der Regressionsanalyse wird durch dieses Vorgehen erheblich «entschärft»: Wenn die Validität

(Treffericherheit) des Regressionskoeffizienten β_5 derart bescheiden ist, dass sie eine derart grosse betragsmässige Korrektur mit einer Toleranzschwelle nahelegt, erscheint der *zusätzliche Nutzen* der OLS-Regressionsanalyse mit der Dummyvariablen F/M im Hinblick auf die rechtssichere Umsetzung von Art. 3 GIG als *gering*: Die Regressionsanalyse wird degradiert zum «*stumpfen Schwert*» mit geringer Trennschärfe.

[Rz 19] Signifikanzniveau $F(z) = 0.95$ ($\alpha = 5\%$)

Auch das *a priori* ausgewählte Signifikanzniveau $F(z)$ des t-Tests ist *hoch kritisch* für das Resultat der Signifikanzbeurteilung und damit für die Qualifizierung eines Betriebes als «verdächtig» oder «nicht verdächtig»:

F(z)	Anzahl Freiheitsgrade N (Beobachtungen ÷ Arbeitnehmer/innen)			
	50	100	200	∞
0.90	1.30	1.29	1.29	1.28
0.95	1.68	1.66	1.65	1.65
0.975	2.01	1.98	1.97	1.96
0.99	2.40	2.37	2.35	2.33

Es ist nicht bekannt, nach welchen Überlegungen für Logib EBG dieses Signifikanzniveau 0.95 ausgewählt worden ist: Vermutlich dürfte dabei der *Gesichtspunkt des «Üblichen»* seine omnipräsente Wirkung entfaltet haben. Indessen scheint uns dieses Niveau *eher etwas grosszügig* zu sein, da zuvor schon eine *relativ üppige Toleranzschwelle S* festgelegt wird. Würde alternativ dazu ein Signifikanzniveau $F(z)$ von 0.99 entsprechend $\alpha = 1\%$ ausgewählt, so würden die T-Werte von einer erheblich kleineren Zahl von Betrieben die Signifikanzgrenze überschreiten und daher wesentlich weniger Betriebe als «verdächtig» qualifiziert werden.

[Rz 20] Verändern der Entscheidungsgrenze von β_5 : Verkleinern der Toleranzschwelle S?

Es wird mehrfach darauf hingewiesen, dass die Toleranzschwelle S von 5% abgesenkt werden müsste, wenn ein *besseres* Regressionsmodell (oder ein anderes Modell?) mit einer *grösseren* Zahl von unabhängigen Variablen x_i verwendet würde.³³ Es ist allerdings nicht genau erkennbar, nach welchen Gesichtspunkten und Überlegungen diese Absenkung gegebenenfalls später einmal tatsächlich durchgeführt werden solle. Anscheinend spielt dabei die Veränderung der Zahl der Variablen x_i die entscheidende Rolle, es gilt offenbar der Zusammenhang: Mehr Variablen x_i – geringerer Betrag der Toleranzschwelle.³⁴

Aus den angeführten Beispielen ergibt sich, dass das Festlegen von *Entscheidungsgrenzen* (Grenzwerten, *cut-off values*) für den normativen Einsatz eines linearen Regressionsmodells OLS mit *kognitiven Methoden nicht vollständig begründbar* ist (verifizierbar / falsifizierbar). Dies gilt insbesondere im Hinblick auf den Einsatz der Regressionskoeffizienten von Dummyvariablen bei der Entdeckung von Verstössen gegen Art. 3 GIG. Dadurch bilden diese Entscheidungsgrenzen neben der *Auswahl* der unabhängigen Variablen (Beurteilungskriterien) ein zweites Einfallstor für *Unsicherheit* und *Willkür*.

2.4.3. EXKURS: Entscheidungsgrenzen auf der individuellen Ebene? Umsetzung von Art. 3 GIG im Einzelfall

[Rz 21] Vorstufe: Die Berechnung des Regressionskoeffizienten β_5 und der Signifikanz des mit SE «korrigierten» Betrages von β_5 dient dazu, einzelne Betriebe als Ganzes als «verdächtig» bezüglich Lohngleichheit zu qualifizieren und «auszusortieren». Diese Berechnung kann dagegen keine *individuellen* Verstösse gegen Art. 3 GIG in einem Betrieb entdecken. Wie bereits erwähnt stellt sie lediglich eine (\pm unverbindliche) «Vorstufe» zur Ermittlung von Verstössen gegen Art. 3 GIG in einem Einzelbetrieb dar. Auch *nach* der Qualifizierung eines «verdächtigen» Betriebs (mit Hilfe des korrigierten β_5) müssen *individuelle* Verstösse gegen Art. 3 GIG vor Ort «manuell» durch direkte Konfrontation von Frauen- und Männerlöhnen entdeckt und nachgewiesen werden. Diese Konfrontation kann sich dabei auf Einzelpersonen oder auf Personengruppen (Berufsgruppen, Funktionsgruppen) erstrecken.

[Rz 22] Differenzen gegenüber einem tatsächlich beobachteten Männerlohn trotz gleichem Punktwert des Arbeitsplatzes

Bei dieser Entscheidung im Einzelfall ist *keine* Entscheidungsgrenze und keine (zweite) Toleranzschwelle festzulegen. *Jede* unter Berücksichtigung der als relevant erkannten Beurteilungskriterien beobachtete Differenz $y(m) > y(f)$ trotz gleicher Arbeitsplatzbewertung ist formell ein Verstoss gegen Art. 3 GIG. Zur Feststellung eines solchen Verstosses bildet allerdings eine multiple Regressionsanalyse *keine notwendige Voraussetzung*.³⁵

Leider sollte dabei nicht übersehen werden, dass vermutlich nach wie vor eine nicht gerade geringe Zahl von Verstössen gegen Art. 3 GIG bestehen bleiben wird, da Frauen es nach wie vor vorziehen werden, lieber weiterhin in einem emotionell intakten Arbeitsverhältnis zu arbeiten, statt einen Verstoss gegen Art. 3 GIG mit rechtlichen Mitteln zu korrigieren versuchen und dadurch in aller Regel das Arbeitsverhältnis zu beenden. Diese Prognose entspricht der leider unausweichlichen Erkenntnis, dass das Arbeitsrecht in der Schweiz mindestens im privatrechtlichen Bereich in der grossen Mehrzahl aller Fälle ein «*Recht des beendeten Arbeitsverhältnisses*» bleibt, in dem rechtliche Konflikte (nicht nur nach GIG) regelmässig erst *nach* Beendigung des Arbeitsverhältnisses ausgetragen werden.³⁶

[Rz 23] Differenz zwischen beobachtetem Lohn y und modellkonformem Lohn \hat{y}

Wie bereits erwähnt (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.3) bildet eine Differenz zwischen einem beobachteten Lohn y_k einer Einzelperson k und einem mit einer Regressionsanalyse (beispielsweise Standard- Analysemodell EBG / Logib) auf der Grundlage der Lohndaten (y_k, x_{ik}) dieser Person berechneten modellkonformen Lohn \hat{y} nicht ohne weiteres eine Lohnungleichheit und einen Verstoss gegen Art. 3 GIG. Ein solcher Verstoss kann – wie mehrfach erwähnt – nur durch direkte Konfrontation von (mindestens) zwei tatsächlich beobachteten Löhnen (Einzelpersonen oder einzelne Personengruppen) entdeckt werden. Umgekehrt bildet eine Anpassung eines beobachteten Lohnes y_k an einen derart ermittelten modellkonformen Lohn \hat{y} noch keine Sicherheit dafür, dass eine gegebenenfalls bestehende Lohnungleichheit und ein Verstoss gegen Art. 3 GIG dadurch vollständig korrigiert und eliminiert wird.

Es gibt wegen fehlender Rechtsgrundlage (vorläufig) auch keinen Rechtsanspruch des einzelnen Arbeitnehmer/in k auf Zahlung eines derartigen modellkonformen Lohnes \hat{y} . Dies steht natürlich unter dem Vorbehalt, dass es den

beteiligten Parteien selbstverständlich frei steht, einen derartigen Rechtsanspruch durch Konsens in einem Betriebs- oder Bewertungsreglement zu *vereinbaren*. Für die Formulierung und Konstruktion einer derartigen Regelung dürfte allerdings wohl regelmässig die Hilfe von ausserbetrieblichen Experten erforderlich sein.³⁷

[Rz 24] Zusammenfassung der Ergebnisse

Die dargestellten Problembereiche der Regressionsanalyse dürften unter anderem auf den Gegensatz zwischen der Untersuchung des *Gender Wage Gap* auf Landesebene und derjenigen auf Betriebsebene zurückzuführen sein. Der Verstoss gegen Art. 3 GIG ist ein inner-betrieblicher, auf Einzelfälle zugeschnittener Sachverhalt, welcher die Bildung von Differenzen zwischen Frauen- bzw. Männerlöhnen bestimmter Einzelpersonen oder einzelner Personengruppen (Berufsgruppen, Funktionsgruppen) voraussetzt. Ob derartige Verstösse gegen Art. 3 GIG im Einzelfall mit Hilfe von statistischen, auf der *kollektiven Ebene* entwickelter, auf einer Vielzahl gleichartiger Sachverhalte aufbauenden Methoden (multiple Regressionsanalyse) effizient und zuverlässig entdeckt werden können, bleibt zweifelhaft. Es erscheint insgesamt als *keine besonders glückliche Idee*, die Methode der multiplen Regressionsanalyse OLS von der nationalen Ebene auf die Ebene des einzelnen Betriebs zu übertragen und ihr eine normative Funktion zur Entdeckung von individuellen Verstössen gegen Art. 3 GIG zu übertragen.

Zwar erscheint die OLS-Regressionsanalyse als effizientes Instrument für die wissenschaftliche Analyse des kollektiven *Gender Wage Gap* (aber auch anderer Dichotomien) auf der kollektiven, namentlich Landesebene, etwa zum Studium der historischen Entwicklung einzelner Dichotomien im Zeitablauf.³⁸ **Aber** für eine Lohnanalyse auf der *betrieblichen Ebene* im Hinblick auf die Konformität eines Lohnsystems mit Art. 3 GIG weist die multiple OLS-Regressionsanalyse nicht übersehbare Problembereiche sowohl auf der theoretischen als auch auf der praktischen Ebene auf.

Wie bereits in Teil 1 unserer Studie dargelegt, unterliegt die Methode der Regressionsanalyse OLS für die betriebliche Lohnanalyse einer Reihe von Vorgaben mathematischer (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.1) und institutioneller Art (Abschnitt 2.3) und sie verfügt nur über einen beschränkten Informationswert für eine rechtssichere Entdeckung von Verstössen gegen Art. 3 GIG (Abschnitt 2.2).

Zudem weist die Methode der Regressionsanalyse OLS für eine Beurteilung des *Gender Wage Gap* auf Betriebsebene mehrere Einfallstore für Unsicherheiten auf (Abschnitt 2.4). Ihr Nutzen für eine effiziente und rechtssichere Anwendung von Art. 3 GIG auf Betriebsebene erscheint dadurch insgesamt als beschränkt.

3. Praktische Schwierigkeiten bei der Anwendung der Regressionsanalyse auf Betriebsebene

[Rz 25] Beschränkter Anwendungsbereich

Ein besonders wichtiges Kriterium, um die Tauglichkeit eines Instruments zur Messung von Lohnungleichheit zu beurteilen, ist die Anwendbarkeit auf die Masse der Unternehmen. Der Vorschlag geht von einer Untergrenze von 50

Beschäftigten aus, ab der die statistische Regressionsanalyse sinnvoll eingesetzt werden kann. Alleine damit scheiden bereits 98.5% aller Schweizer Unternehmen aus der Betrachtung aus, wobei diese Kleinunternehmen mit weniger als 50 Mitarbeitenden ca. 47.5% aller Arbeitskräfte in der Schweiz beschäftigen.³⁹ Oder anders ausgedrückt: bei fast der Hälfte aller Arbeitnehmer/innen lässt sich mit dem Mittel der statistischen Regressionsanalyse überhaupt keine Aussage hinsichtlich Lohnungleichheit machen.

Hinzu kommt, dass für eine statistische Analyse weitere Voraussetzungen erfüllt werden müssen. Selbst wenn der betrachtete Betrieb mehr als 50 Arbeitskräfte beschäftigt, müssen für einen Vergleich beide Geschlechter über den gesamten Bereich annähernd *gleichmässig verteilt* sein und jeweils mindestens 10 Personen umfassen. Damit scheiden viele weitere Unternehmen, wie zum Beispiel grössere Handwerksbetriebe und Unternehmen aus den Bereichen Hoch- und Tiefbau für die Anwendung der Regressionsanalyse und damit des Standard-Analysemodells EBG und *Logib* aus, denn dort findet man in den allermeisten Fällen keine äquivalente Häufigkeitsverteilung der Löhne der beiden Geschlechter. Von daher dürfte der Anteil der Arbeitnehmer/innen, bei denen die Frage der Lohnungleichheit mit statistischen Analysen bearbeitet werden kann, bei deutlich unter 50% aller Angestellten liegen.

[Rz 26] Verständnisprobleme der Rechtsanwender

Schon in der Frühzeit der Entwicklung von multi-criteria Methoden in den Sozialwissenschaften ist darauf hingewiesen worden, dass multi-criteria Instrumente für die Anwender leicht *verständlich* sein sollten, um in der Praxis auch tatsächlich erfolgreich angewendet zu werden. Jenseits aller mathematischen Überlegungen müsse die Methode für den Anwender *unmittelbar* Sinn machen, sonst werde sie in der Praxis nicht benutzt.⁴⁰ Es ist zu befürchten, dass sich bei der praktischen Anwendung der multiplen Regressionsanalyse mit der – keineswegs selbstverständlichen – Technik der Dummyvariablen (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.2) auf Betriebsebene unter Umständen erhebliche Verständnisprobleme ergeben dürften. Im Ergebnis dürfte daher ein gewichtiger *praktischer Nachteil* der multiplen Regressionsanalyse bei der Untersuchung der Lohnungleichheit darin bestehen, dass diese Methode weder von den unmittelbar betroffenen Personen (Arbeitgeber/innen, Arbeitnehmer/innen) noch im Falle eines Rechtsstreits vom juristischen Entscheidungsträger (Gerichte) verstanden werden wird.⁴¹

[Rz 27] Beschränkter Informationswert der Resultate: «Lokalisierung» der Lohnungleichheit

Wenn man bereit ist zu akzeptieren, dass weniger als die Hälfte der Zielgruppe überhaupt betrachtet werden kann, stellt sich als nächstes die Frage nach dem Informationswert und der praktischen Verwertbarkeit der Analysedaten. Was *Logib* und ähnliche Instrumente leisten können, ist eine Vermutung einer möglichen Lohnungleichheit in einer Gruppe von mindestens 50 (besser: 100) Personen und bei gleichmässiger Verteilung der Geschlechter auszusprechen (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.2). Damit ist das vermutete Problem aber nicht ausreichend präzise auf die betroffenen Personen *lokalisierbar* und schon gar nicht gelöst. Es müssen in dieser Gruppe (Betrieb) mit dem vermuteten Lohnungleichheitsproblem zunächst alle Arbeitsplatzbewertungen auf eine mögliche Geschlechterbenachteiligung hin analysiert werden. Danach muss überprüft werden, ob die Kriterien zur Einstufung der Mitarbeitenden in das jeweils definierte Lohnband der Stelle benachteiligungsfrei gestaltet sind und schliesslich, ob

alle Mitarbeitenden den Kriterien entsprechend korrekt im Lohnband eingestuft wurden. Erst an diesem Punkt der Untersuchung ist es möglich, eine aufgefundene Lohnungleichheit zu korrigieren.

[Rz 28] Effizienz einer Beurteilung eines Einzelbetriebs

Eine kollektive Beurteilung eines ganzen Betriebs zur Beurteilung einer Rechtsstreitigkeit *im Zivilrecht* zwischen (meistens) einer Klägerin und wenigen männlichen Referenzpersonen ist nicht notwendig und auch nicht ausreichend für den rechtsgenügenden Nachweis eines Verstosses gegen Art. 3 GIG, wie in Teil 1 in Abschnitt 2.2 dargelegt worden ist. Im Gegensatz dazu reicht eine analytische Arbeitsplatzbewertung der unmittelbar betroffenen Personen für diesen Nachweis aus. Eine multiple Regressionsanalyse erscheint unter diesem Blickwinkel nicht als unentbehrlich und den anderen Methoden nicht überlegen. Falls dennoch (trotz allen methodischen Einwänden) eine kollektive Beurteilung einzelner Betriebe gewünscht werden sollte, stehen zudem effizientere und zuverlässigere quantitative Methoden als die multiple Regressionsanalyse zur Verfügung (unten Abschnitt 4.3).

[Rz 29] Kosten der Regressionsanalyse

Das Standard-Analysemodell EBG und die Software *Logib* wurde im Auftrag des Eidgenössischen Büros für Gleichstellung (EBG) entwickelt und mit öffentlichen Mitteln finanziert. Es steht interessierten Unternehmen kostenlos zum Download zur Verfügung. Auf Unternehmensseite fallen dementsprechend vor allem die Kosten für die Schulung der Anwender und die praktische Datenerhebung an. Falls die *Logib-Analyse keinen* Verdacht von Lohnungleichheit ergibt (vielleicht $\beta_5 < 0.10$), bleiben die Kosten im Rahmen und bewegen sich im Bereich von drei bis vier Arbeitstagen der Anwender. Anders sieht es aus, wenn das Instrument potenzielle Probleme (vielleicht $\beta_5 > 0.10$, Rechenbeispiel des «Methodenbeschriebs» $\beta_5 = 0.138$) anzeigt. In solchen Fällen müssen Detailanalysen bis auf die Stufe der einzelnen Arbeitsplätze und der individuellenlohneinstufungen gemacht werden.

Insbesondere wenn das Unternehmen weder eine anerkannte, diskriminierungsfreie Methode zur Bewertung aller Stellen einsetzt noch Kriterien zur Einstufung der Beschäftigten in die jeweiligen Lohnbänder etabliert hat, welche geschlechtsunabhängig sind, wird es ausserordentlich aufwendig. In diesem Fall müssen alle Stellen und alle individuellen Einstufungen des von einer potenziellen Lohnungleichheit betroffenen Bereichs aufgrund von Interviews mit allen Vorgesetzten neutral analysiert und allfällige Korrekturen vorgenommen werden. Der Aufwand für eine solche Analyse kann selbst für ein kleines Unternehmen rasch eine kritische Grösse erreichen. Ein individueller Datensatz von jedem Arbeitnehmer/in muss erhoben werden; also Interviews usw. Schon bei KMU unter 100 Arbeitnehmern entstehen dadurch Kosten in Höhe eines vierstelligen Frankenbetrags.

[Rz 30] Zusammenfassung der Ergebnisse

Für einen praktischen Einsatz auf Betriebsebene weist die Regressionsanalyse OLS eine Reihe von Problembereichen auf: Beschränkter Anwendungsbereich insbesondere für KMUs, Verständnisprobleme der Rechtsanwender, Notwendigkeit einer manuellen Nachprüfung zur «Lokalisierung» individueller Verstösse gegen Art. 3 GIG sowie zusätzliche Kosten einer derartigen individuellen Nachprüfung bei als «verdächtig» qualifizierten Betrieben.

4. Alternative Lösungsansätze

Angesichts der dargestellten Problembereiche der multiplen Regressionsanalyse lassen sich eine Reihe praktischer Empfehlungen für eine effiziente und rechtssichere Anwendung von Art. 3 GIG formulieren. Das Ziel dieser Empfehlungen besteht ausschliesslich darin, die Rechtssicherheit (Reliabilität) und die Treffsicherheit (Validität) der Rechtsanwendung im Bereich von Art. 3 GIG zu verbessern und die Akzeptanz der Bewertungsverfahren bei den Mitarbeiter/innen und den Rechtsanwendern zu gewährleisten.

4.1. Grundregeln: Konsens und Transparenz

[Rz 31] Eliminieren von Unsicherheit durch Konvention:

Alle methodischen Festlegungen zur Lohnfindung sollten auf *Konsens* beruhen

Um den in Abschnitt 2.4 skizzierten Unsicherheiten (Auswahl und Inhalt der unabhängigen Variablen bzw. Beurteilungskriterien) auf effiziente Weise Rechnung zu tragen, sollten alle Festlegungen zur *Methode* der Lohnfindung zwischen den beteiligten Parteien und gegebenenfalls deren Interessenvertretern im Rahmen der Sozialpartnerschaft *vereinbart* werden. *Und nicht* durch Erlass von zwingendem Gesetzesrecht durch den Gesetzgeber eingeführt werden, wie dies anscheinend die geplante Revision des GIG nunmehr vorsieht. Dies ist der *einzigste effiziente Weg*, mit diesen auf *kognitive Weise* nicht ausreichend begründbaren und eliminierbaren Unsicherheiten umzugehen. Und befreit von unnötigen und unlösbaren Diskussionen über diese methodischen Parameter. Den fraglichen Parametern kommt dadurch (bloss) *konventioneller Charakter* und nicht eine durch kognitive Mechanismen begründete bzw. gewonnene Gültigkeit (Validität) und Reproduzierbarkeit (Reliabilität) zu.⁴²

Dieses Festlegen der Beurteilungskriterien durch Konvention erscheint auch aus dem Blickwinkel des Arbeitsrechts als angezeigt: Auch aus arbeitsrechtlichen Überlegungen sollten alle irgendwie mit *Unsicherheit* behafteten methodischen Elemente der Lohnfindung (Beurteilungskriterien: Auswahl, Inhalt, Skalierung, Gewichte) in generell-abstrakter Weise und *pro futuro* durch *Konsens* zwischen den Beteiligten festgelegt werden. Dabei sollte zur Erzielung optimaler Betriebs- und Realitätsnähe der Inhalt dieser Festlegungen zweckmässigerweise *betriebs-, regionen- oder branchenspezifisch* ausgestaltet erfolgen.

Im einzelnen könnte diese konventionelle Festlegung der methodischen Parameter in folgenden Formen verlaufen: **Zwar** könnten im privatrechtlichen Arbeitsverhältnis *Einzelvereinbarungen* über die Elemente der Lohnfindung getroffen werden, zweckmässiger erscheint es **aber**, diese Parameter einmal *pro futuro* als *Betriebsreglement* festzulegen und gegebenenfalls bei Bedarf anzupassen. Derartige Betriebsreglemente («Bewertungsreglemente») könnten durch kollektive Vereinbarung zwischen Arbeitgeber und Personalkommission geschaffen werden. Sie könnten auch als Teile von GAV oder (soweit inhaltlich) zulässig von «Betriebsordnungen» nach Art. 37/38 des Arbeitsgesetzes (ArG; SR 822.11) ausgestaltet werden. Diesen Reglementen würde der/die zu einem späteren Zeitpunkt rekrutierte Mitarbeiter/in beim Abschluss seines/ihres Individualvertrages jeweils zustimmen und deren Inhalt dadurch zum Vertragsinhalt seines/ihres Arbeitsverhältnisses erheben.⁴³

Die methodischen Parameter der Lohnfindung (Kriterien, Gewichte, Skalierung) könnten zudem von einer neutralen Instanz im Hinblick auf deren Diskriminierungsfreiheit *zertifiziert* werden. Nach der Zertifizierung durch neutrale

Experten müsste indessen wiederum der *verbindliche Konsens* zwischen den beteiligten Sozialpartnern und den direkt Beteiligten erzielt werden. Demgegenüber sieht der Entwurf zur Revision des GIG vor, dass die Methode der Regressionsanalyse als *zwingendes Gesetzesrecht* landesweit auf alle privatrechtlichen Arbeitsverhältnisse anwendbar sein soll und dadurch an die Stelle der konventionellen Festlegung zwischen den Beteiligten treten solle.⁴⁴

[Rz 32] Betriebsinterne Transparenz:

Alle methodischen Parameter zur Lohnfindung werden *betriebsintern* veröffentlicht

Diesem konventionellen Charakter der methodischen Parameter der Lohnfindung entspricht es, dass die methodischen Festlegungen zwischen den Beteiligten nicht nur vereinbart, sondern auch betriebsintern *transparent* gemacht werden. Also: Alle methodischen Einzelheiten der Lohnfindung, insbesondere Auswahl, Inhalt, Skalierung und Gewichte der Beurteilungskriterien (= der unabhängigen Variablen x_i) werden *betriebsintern, für alle Betriebsangehörigen* zugänglich gemacht, so dass alle Beteiligten nachvollziehen können, auf welcher Grundlage und mit welchen Berechnungen seine/ihre persönliche Einstufung und Lohnfindung erfolgt.⁴⁵ Zusätzlich könnte betriebsintern eine geeignete Stelle (Ombudsstelle) organisiert werden, die der/die Arbeitnehmer/in bei einer vermuteten Ungleichbehandlung bei der Einstufung des Arbeitsplatzes (Stufe 1) oder der Positionierung innerhalb des Lohnbandes (Stufe 2) anrufen könnte bzw. vor Einleiten rechtlicher Schritte obligatorisch anrufen müsste.

Zur Ergänzung und Verstärkung der Transparenz könnte zudem ein inhaltlich differenzierter *Auskunftsanspruch* des/der Arbeitnehmers/in nach Vorbild der EU und des Entwurfs der deutschen Bundesregierung (Februar 2017) eingeführt werden. Schwierigkeiten im Hinblick auf den Schutz personenbezogener Daten ergeben sich dabei aus der betriebsinternen Transparenz noch nicht, da die betreffenden methodischen Festlegungen generell-abstrakter Natur sind und zudem erst *pro futuro* personalisierte Wirkungen entfalten. Erst die praktische Durchsetzung eines Auskunftsrechts im Einzelfall könnte Bedenken bezüglich des Datenschutzes entstehen lassen. Diese könnten indessen nach dem Vorbild von §12 des Entwurfs der deutschen Bundesregierung ausgeräumt werden (Einschränkung der Zahl der Vergleichslöhne und Anonymisierung der erteilten Auskünfte).⁴⁶

[Rz 33] Verfahren verläuft in zwei inhaltlich unterschiedlichen Stufen

Das empfohlene Verfahren der Lohnfindung läuft in zwei inhaltlich unterschiedlichen Stufen ab: In einer ersten Stufe werden die Arbeitsplätze *unabhängig von der Person* der Stelleninhaber nach den klassischen Kriterien bewertet und in einer Punktzahl $V(x_i)$ ausgedrückt: *Personenunabhängige Stufe 1*. Diese Punktezahlen $V(x_i)$ werden mit einem Faktor F in einen Lohnbetrag L_i umgerechnet und erzeugen eine Lohnkurve: $L_i = F \cdot V(x_i)$. Diese Lohnkurve ist auf einen längeren Zeithorizont von mehreren Jahren im gleichen Betrieb angelegt und kann nach Bedarf an die Verhältnisse angepasst werden, beispielsweise durch *Indexierung* des Faktors F , Auszahlung eines betriebsweit uniformen Bonus für guten Geschäftsgang und ähnliches.⁴⁷

In der zweiten Stufe wird den einzelnen Lohnbeträgen der Lohnkurve ein prozentmässig definierter betriebsweit konstanter Prozentbetrag c addiert, welcher beispielsweise $c(\max) = 10$ bis 20% der betreffenden Punktzahl $V(x_i)$ beträgt. Dadurch wird für jede Punktzahl $V(x_i)$ ein *Lohnband* mit den Grenzen $F \cdot V(x_i) < L_i < F \cdot V(x_i) + c(\max)$ geschaffen. Mit Hilfe der klassischen Kriterien des sog. *Humankapitals*, welche sich ausschliesslich an der Person des

Stelleninhabers orientieren, wird der Wert der Variablen c im Einzelfall ermittelt und dadurch die Person innerhalb des Lohnbandes positioniert: *Personen-orientierte Stufe 2 bzw. Humankapital-Stufe*. In dieser Stufe 2 werden Parameter wie Dienstalter, Lebensalter, individuelle Leistung und gegebenenfalls auch Parameter der familiären Situation (Familienvater, besondere Unterhaltspflichten, Distanz Wohnort–Arbeitsort usw.) des Stelleninhabers berücksichtigt. Auch für diese Stufe 2 werden die anwendbaren Regeln (Kriterien, Punktzahlen usw.) in generell-abstrakter Weise *pro futuro* festgelegt und gelten für den gesamten Betrieb.

4.2. Stufe 1: Auswahl und Inhalt der Beurteilungskriterien

[Rz 34] *Auswahl*: Es ist bereits oben (Abschnitt 2.4) die Vermutung geäußert worden, dass *niemals ein vollständiger / unwidersprochener Konsens* unter allen landesweit beteiligten Akteuren der Lohnfindung erzielt werden wird über die Auswahl und den Inhalt der Beurteilungskriterien, sei es der OLS-Regressionsanalyse oder sei es der von uns bevorzugten gewichteten Addition SAW. Aufgrund dieser Vermutung erscheint es deshalb aussichtsreicher, den Raster der Beurteilungskategorien der personenunabhängigen Stufe 1 *massgeschneidert* für Branchen-, Regionen-, Einzelbetriebe festzulegen, statt eine landesweit uniforme und flächendeckende Anwendbarkeit eines einzigen Beurteilungsrasters anzustreben.⁴⁸ Dadurch soll eine möglichst *betriebsnahe Beurteilung* erreicht und zugleich vermieden werden, dass eine landesweite (ohnehin unlösbare) Gleichbehandlungsdiskussion zwischen den einzelnen Regionen, Branchen und Betrieben entsteht.⁴⁹ Sogar die vom Bund finanzierte Studie von FELFE et al (2015) anerkennt, dass mit derartigen massgeschneiderten Beurteilungsrastern den betrieblichen Situationen *besser Rechnung getragen* werden kann als mit einer statistischen Beurteilungsmethode wie der multiplen Regressionsanalyse, welche eine landesweit uniforme Anwendung anstrebt.⁵⁰

Ob der massgeschneiderte betriebsnahe Kriterienraster tatsächlich besser ist als ein auf landesweit uniforme Anwendbarkeit zugeschnittener Raster lässt sich problemlos durch Berechnen und Vergleich der Determinationskoeffizienten R^2 bestimmen, wobei die Zahl der eingesetzten Einzelkriterien konstant zu halten ist. Die technische Unabhängigkeit der einzelnen Kriterien voneinander (Multikollinearität, Autokorrelation, Redundanzen) sollte dabei in einem Pretest routinemässig kontrolliert werden durch Ermitteln der bivariaten Korrelationskoeffizienten z.B. nach *Pearson*.

[Rz 35] Inhaltliche Vorgaben: Diskriminierungsfreiheit der Kriterien

Neben der betriebsweit einheitlichen Anwendung der eingesetzten Beurteilungskriterien ist es erforderlich, dass die Kriterien der personenunabhängigen Arbeitsplatzbewertung (Stufe 1 der Lohnfindung) geschlechts-neutral festgelegt werden. Man muss also darauf achten, dass die Kriterien x_i und allenfalls deren Gewichte w_i bei der Arbeitsplatzbewertung nicht eines der beiden Geschlechter benachteiligen (inhaltliche Diskriminierungsfreiheit).⁵¹ Dies wäre dann gegeben, wenn ein Kriterium statistisch *häufiger* bei einem Geschlecht auftritt als beim anderen oder typischerweise von einem Geschlecht deutlich *besser erfüllt* wird als vom anderen Geschlecht.⁵²

Beim generell-abstrakten Festlegen der Beurteilungskriterien *pro futuro* stellt sich die weitere Frage, ob Beurteilungskriterien inhaltlich so formuliert dürfen, dass sie den *Arbeitsmarkt berücksichtigen* (sog. Marktlohn). Und inwieweit damit die Berufung auf den Arbeitsmarkt gegebenenfalls einen Rechtfertigungsgrund für Lohnungleichheit ergeben könnte.⁵³ In der Rechtsprechung wird angenommen, dass der Arbeitsmarkt nur während einer *beschränkten*

Zeitdauer einen Rechtfertigungsgrund für Lohnungleichheit ergebe. Dies spricht eher *gegen* eine Aufnahme derartiger marktorientierter Beurteilungskriterien in einen von vorneherein auf einen längeren Zeithorizont angelegten Kriterienraster der Stufe 1.⁵⁴

4.3. Stufe 1: Verfahren zur Auswertung der Lohndaten

[Rz 36] Verfahren zur Auswertung der erzielten Punkte $V(\mathbf{x})$ im Einzelfall

Angesichts der dargelegten vielfältigen methodischen Probleme der multiplen Regressionsanalyse (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017 Abschnitte 2.1 bis 2.3 und oben Abschnitt 2.4) wird empfohlen, die *Auswertung* der mit dem personenunabhängigen Kriterienraster (Stufe 1) berechneten Punkte $V(\mathbf{x})$ nach dem Vorbild der analytischen Arbeitsbewertung mit der Methode der *gewichteten Addition* (*simple additive weighting*, SAW) auszuwerten, nicht mit dem Modell der statistischen Regressionsanalyse. Diese Auswertung nach SAW beruht *nicht auf statistischen Überlegungen und Zusammenhängen* und ist nicht an mathematische Vorgaben gebunden wie die Regressionsanalyse (Grösse von Stichproben, Normalverteilungen, gleichmässige Verteilung über den gesamten Wertebereich bei F/M, siehe Teil 1, Abschnitt 2.1, Festlegen von Cut-off values Abschnitt 2.4). Sie führt auch nicht zu Diskussionen über den *Informationswert* (siehe Teil 1, Abschnitt 2.2) für die Entdeckung von Verstössen gegen Art. 3 GIG. Selbst die vom Bund finanzierte Studie von FELFE ET AL (2015) anerkennt, dass mit der analytischen Arbeitsbewertung die betrieblichen Besonderheiten auf zweckmässige Weise erfasst werden können.⁵⁵

[Rz 37] *Personenunabhängigkeit*: Im Verfahren der gewichteten Addition (SAW Simple Additive Weighting) wird das Anforderungsniveau der Arbeitsplätze nach einem Beurteilungsraster *unabhängig von den Personen* der Stelleninhaber bewertet. Das Verfahren baut teilweise auf den Ergebnissen der Arbeitspsychologie auf wie das vielfach erwähnte VFA, ABAKABA und zahlreiche andere Verfahren.⁵⁶ Diese gewichtete Addition (SAW) beruht auf dem einfachen Zusammenhang

$V(\mathbf{x}) = \sum w_i v_i(x_i)$, wobei

$V(\mathbf{x})$ den von den verschiedenen Attributen x_i erzeugten Gesamtwert (Punktzahl) des beurteilten Objekts (hier: Arbeitsplatz),

x_i die einzelnen bewerteten Attribute (items, hier notwendige Ausbildung, intellektuelle Anforderungen, Führungsverantwortung usw.),

w_i das Gewicht (weight) der einzelnen Attribute, und

v_i den bei den einzelnen Attributen erzielten Wert (value, Punktzahl) bezeichnet.

Dabei ist zu beachten, dass die Summe aller Gewichte w_i (weight) einen Wert von 1 (oder 100%) nicht übersteigen darf:

$$\sum w_i(x_i) = 1$$

Um Arbeitslöhne von Frauen und Männern vergleichen zu können, sind die $V(\mathbf{x})$ -Werte *interpolierbar* zu gestalten: Fehlt also für einen Frauenlohn mit einer bestimmten Punktzahl $V(\mathbf{x})$ ein korrespondierender Männerlohn mit der exakt gleichen Punktzahl $V(\mathbf{x})$, so kann eine entsprechender hypothetischer Männerlohn (Vergleichslohn) durch lineare Interpolation zwischen zwei benachbarten Punktzahlen $V(\mathbf{x})$ von Männerlöhnen ermittelt und für Vergleiche eingesetzt werden.

[Rz 38] Skalieren / Standardisieren der Kriterien x_i

Damit die Punkte, welche bei den verschiedenen Kriterien x_i erzielt werden, durch einen Vorgang wie die gewichtete Addition miteinander *aggregiert* werden und ein Gesamtergebnis ergeben können, müssen generelle Regeln zur Skalierung und Standardisierung der einzelnen Kriterien x_i festgelegt werden.

Eine unterschiedliche Skalierung einzelner Kriterien kann sich gleich auf das Ergebnis einer Aggregation (hier: der gewichteten Addition) auswirken wie eine sachlich nicht begründete, unterschiedliche Gewichtung einzelner Kriterien, nämlich in einer (beabsichtigten oder ungewollten) Über- oder Unterbewertung bzw. Bevorzugung / Benachteiligen einzelner Kriterien. Dieses Standardisierungsproblem kann beispielsweise derart gelöst werden, dass allen unabhängigen Variablen oder Gruppen von Variablen die *gleiche maximale Punktzahl* zugeteilt wird.⁵⁷

[Rz 39] Unterschiedliche Gewichte w_i für die einzelnen Attribute x_i ?

Schon in der Frühzeit des Aggregierens von einzelnen Indikatoren zu einem Gesamtwert durch *gewichtete Addition* (1970) ist erkannt worden, dass die Zuteilung der *Gewichte* w_i einen Problembereich bildet und Kontroversen über das Festlegen dieser Gewichte im Einzelfall beinahe unausweichlich sein werden. Deshalb wird auch für die Arbeitsplatzbewertung empfohlen, die *Gewichte* w_i der einzelnen Kriterien x_i bzw. Attribute x_i unter Berücksichtigung von betriebs-, branchen- oder regionenspezifischen Gesichtspunkten generell-abstrakt zu formulieren und dadurch die betriebliche Realität möglichst exakt abzubilden. Auch diese Gewichte müssen durch Konsens zwischen den Beteiligten verbindlich festgelegt werden.⁵⁸

Als attraktive Alternative bietet sich an, allen Kriterien *das gleiche Gewicht (equalweighting)* zuzuteilen und das Bewertungsverfahren damit ausschliesslich auf der Tatsache aufzubauen, dass eine bestimmte Eigenschaft/Attribut überhaupt als Kriterium ausgewählt und in die Rechnung einbezogen wird. Wie DANIEL KAHNEMAN (2012) nachgewiesen hat, führt dieses Verfahren des *Equalweighting*, bei dem auf unterschiedliche Gewichte w_i verzichtet wird, mit einiger Wahrscheinlichkeit zu einem ebenso brauchbaren Resultat wie die Anwendung unterschiedlicher Gewichte w_i . Equalweighting bildet daher im Einzelfall eine durchaus prüfenswerte und sinnvolle methodische Alternative. Dies ist darauf zurückzuführen, dass lineare Modelle wie die gewichtete Addition SAW sich gegenüber Abweichungen von einer optimalen Gewichtung w_i ziemlich *robust* verhalten. Oder anders ausgedrückt: Die Kurve zur Bestimmung optimaler Gewichte w_i hat ein «flaches Maximum». ⁵⁹ Gestützt auf diese empirischen Erkenntnisse lässt sich auch für die Arbeitsplatzbewertung der Stufe 1 von Fall zu Fall erwägen, allen Kriterien x_i *das gleiche Gewicht* zuzuteilen.

[Rz 40] Vergleich Regressionsanalyse mit gewichteter Addition SAW

Gegenüber der multiplen Regressionsanalyse weist das Verfahren der SAW eine Reihe von technischen Vorteilen auf:

Besser verständlich: Die Auswertungsmethode der SAW ist für die unmittelbar Betroffenen weit besser verständlich als die multiple Regressionsanalyse unter Verwendung von Dummyvariablen (s. oben Abschnitt 3). Dies ist wohl der *entscheidende und wichtigste Vorteil* der SAW über die multiple Regressionsanalyse. Auch die Berücksichtigung von vielen betriebsspezifischen Variablen x_i dürfte die Betriebsnähe und damit die Akzeptanz der Auswertungsmethode erhöhen.

Keine Untergrenze: Das Verfahren ist nicht an die Untergrenze von 50 Personen gebunden (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.1). Wenn man beispielsweise eine Untergrenze von 10 Personen einführen würde, ab der es eingesetzt werden soll, würde sich der Anwendungsbereich gegenüber Logib und anderen Methoden der statistischen Regressionsanalyse um mindestens 6.5% der Schweizer Unternehmen und 21% der Arbeitnehmer/innen erweitern.⁶⁰

Keine mathematischen Vorgaben: Die vorgeschlagene Auswertung gemäss SAW setzt weder eine Normalverteilung der beobachteten Daten bzw. deren Residuen noch eine gleichmässige Häufigkeitsverteilung bei beiden Geschlechtern über den gesamten Merkmalsraum voraus. SAW ist auch anwendbar, wenn die Häufigkeitsverteilungen der Daten von Frauen und Männern dramatisch unterschiedliche Formen aufweisen würden, also beispielsweise keine Symmetrie, sondern unterschiedliche Schiefe und Kurtosis zeigen würden.

Keine willkürlichen bzw. subjektiven Entscheidungsgrenzen (cut-off values): Da die SAW nicht mit Entscheidungsgrenzen (cut-off values) arbeitet, welche regelmässig mit Unsicherheit und Willkür behaftet sind (s. oben Abschnitt 2.4), werden bei dieser Auswertung der Arbeitsplatzbewertung auch die mit cut-off values regelmässig verbundenen Schwierigkeiten vermieden.

Vordringen bis zur Ebene des Einzelfalles: Ein entscheidender Nachteil der multiplen Regressionsanalyse liegt darin, dass ein beobachteter kollektiver *Gender Wage Gap* (in Form eines hohen Regressionskoeffizienten β_5) weder auf die betroffenen Einzelpersonen (oder einzelne Personengruppen) *lokalisiert* werden kann, noch Hinweise zu den möglichen Ursachen gewonnen werden können. Im Gegensatz zu den statistischen Modellen der Regressionsanalyse würde das hier vorgeschlagene Verfahren der SAW nicht nur Hinweise auf Bereiche liefern, in denen Lohnungleichheit *vermutet* werden könnte, sondern es würde die einzelnen Problemfälle bis auf die Stufe des individuellen Arbeitsplatzes aufzeigen und eine rasche Korrektur ermöglichen. Dadurch ermöglicht es die Auswertung nach SAW, bis zur Ebene der *Lohnfindung im Einzelfall* (bei Einzelpersonen oder einzelnen Personengruppen) vorzudringen und Verstösse gegen Art. 3 GIG rechtlich verbindlich zu entdecken.

4.4. Kollektive Beurteilung des Gender Wage Gap auf Betriebsebene: Trotz allem?

[Rz 41] Falls trotz allem der *Gender Wage Gap* auf Betriebsebene kollektiv beurteilt werden soll, steht eine Reihe statistischer Methoden zur Beurteilung der Ähnlichkeit / Unähnlichkeit von zwei Häufigkeitsverteilungen zur Verfügung, deren

Effizienz derjenigen der multiplen Regressionsanalyse mit einer Dummyvariablen gemäss EBG Standard-Analysemodell mindestens gleichkommt oder diese noch übertrifft. Alle nachfolgend vorgestellten statistischen Methoden sind *nicht-parametrisch*, bauen also *nicht* auf der Annahme auf, die fraglichen Häufigkeitsverteilungen (Punkte oder Löhne) für Männer und Frauen bzw. deren Residuen bildeten je eine Normalverteilung, was sie ziemlich sicher in den meisten Fällen auch nicht tun. Die vorgestellten statistischen Methoden weisen zudem den praktischen Vorteil auf, dass sie keine vorgängige Lohnanalyse mit einer Vielzahl von Kriterien (= unabhängigen Variable x_i) voraussetzen, sondern auf der Grundlage der unverarbeiteten Lohndaten (Löhne oder Lohnsummen von Frauen und Männern) oder der Zahl N der betreffenden Arbeitnehmer/innen ausgeführt werden können und dadurch eine schnelle und kostengünstige *prima facie* Beurteilung auf Betriebsebene ermöglichen.

[Rz 42] Vierfeldertest einzelner Segmente von Arbeitnehmer/innen

Ein *prima facie* Eindruck von der Ähnlichkeit bzw. den Unterschieden zwischen Frauen- und Männerlöhnen in einem Betrieb lässt sich durch einen simplen *Vierfeldertest* gewinnen: Dazu werden die für Frauen und Männer *getrennt ermittelten* Personenzahlen N oder der Lohnsummen durch einen für beide Verteilungen identischen Trenn- oder Grenzwert (cut-off value; Vorschlag: Median) in zwei Segmente (Bereiche, Hälften) aufgeteilt. Beispiel: Trennung durch den Median in niedrige / hohe Löhne). Die Häufigkeiten der derart gebildeten vier Bereiche (hoch – niedrig, F – M) kann in einer Vierfelderanordnung dargestellt und beispielsweise mit Fischer's exaktem z-Test auf Signifikanz gegen Zufall beurteilt werden. Hohe Signifikanz könnte bereits ein Indiz für Lohnungleichheit zwischen den Geschlechtern ergeben.

[Rz 43] Rangkorrelationskoeffizienten r_s nach Spearman

Diese Verfahren können durch die ebenfalls nicht-parametrische statistische Methode einer Beurteilung der *Rangkorrelationen* r_s nach Spearman von Lohnklassen verfeinert werden: Dazu wird die Häufigkeitsverteilung von Punktwerten $V(x_i)$ oder Löhnen L_i von Männern und Frauen wiederum *getrennt* bestimmt und eine Anzahl (wenn möglich > 10) von Punkteklassen bzw. Lohnklassen gebildet. Diesen Klassen kann nach Massgabe der beobachteten Häufigkeit (Zahl der Personen) ein Rang innerhalb der für die Geschlechter getrennten Verteilungen zugeordnet werden. Die von den einzelnen Lohnklassen erzielten Ränge innerhalb der Verteilungen von Frauen und Männern werden miteinander verglichen und der *Rangkorrelationskoeffizient* r_s nach Spearman (1950) berechnet. Die Signifikanz von r_s kann mit der t-Teststatistik beurteilt werden, wobei die Zahl N der verwendeten Lohnklassen die Zahl der Freiheitsgrade bestimmt und lediglich noch ein zweckmässiges Signifikanzniveau ausgewählt bzw. vereinbart werden muss.⁶¹

Auch dieser Test bietet den methodischen Vorteil, dass er *nichtparametrisch* ist; falls dabei die unverarbeiteten Lohnklassen verwendet werden, setzt er auch keine vorgängige Arbeitsplatzbewertung mit einer multi-criteria Methode voraus.

[Rz 44] Kolmogorov-Smirnov *two samples test*

Die folgenden Methoden bauen auf den *kumulierten* Punktwerten (oder Lohnsummen) in den für Frauen und Männer

getrennten Häufigkeitsverteilungen auf. Wie bereits an anderer Stelle dargelegt worden ist, bietet der Kolmogorov-Smirnov *two samples* Test eine nicht-parametrische *Beurteilung des kollektiven Gender Wage Gap* auf Betriebsebene:⁶² Die beobachteten Häufigkeitsverteilungen aus der Arbeitsplatzbewertung (in Lohnbeträgen oder Punkten) von Frauen und Männern innerhalb eines Betriebs werden *getrennt* ermittelt und kumuliert. Die kumulierten Häufigkeitsverteilungen werden in diesem Test anhand der beobachteten maximalen Differenz D_{\max} miteinander verglichen. Die Verteilung der Differenz D ist bekannt (Tabelle) und der beobachtete Wert D_{\max} kann mit dem t-Test auf Signifikanz getestet werden. Dabei muss von den Beteiligten noch ein Signifikanzniveau für die Grösse D als Entscheidungsgrenze (*cut-off*) ausgewählt werden. Dieser Test liefert schon im Ansatz *mehr Erkenntnisse* als eine Berechnung des Regressionskoeffizienten β_5 der Dummyvariablen FM im Kontext der multiplen Regressionsanalyse, da er auch ein Indiz für eine *mittelbare/indirekte* Diskriminierung durch Zuteilung von Arbeit (sog. Beschäftigungsdiskriminierung) bildet.

[Rz 45] Vergleich von Konzentrationsmassen: Gini-Werte

Das Konzentrationsmass nach Gini geht aus von der bekannten doppelt kumulierten *Lorenzkurve*, verwendet wie der Kolmogorov-Smirnov *two-samples* Test ebenfalls eine Akkumulierung (von Arbeitsplatz-Punkten oder Löhnen) und erfasst wie dieser vermutlich nicht nur eine direkte Lohndiskriminierung, sondern auch eine indirekte Beschäftigungsdiskriminierung von Frauen. Auch dieses Gini-Konzentrationsmass ist nicht-parametrisch und setzt keine Normalverteilungen voraus. Deshalb könnte ein Vergleich, beispielsweise der *Quotient Q* oder die *Differenz D*, zwischen den *Gini-Werten* der nach Geschlechtern getrennten Lohnverteilungen in einem Betrieb eine Alternative für die *prima facie* Beurteilung des *wage gender gap* bieten wie der Kolmogorov-Smirnov *two samples test*.⁶³

Es erscheint dabei *prima facie* als plausibel, dass ein niedriger Gini-Wert (relative Lohngleichheit) bei den Frauenlöhnen und ein höherer Gini-Wert (relative Ungleichheit) bei den Löhnen von Männern, also ein von $Q = 1$ stärker abweichender Quotient Q für eine Lohndiskriminierung und Beschäftigungsdiskriminierung zwischen den Geschlechtern sprechen. Der beobachtete Wert des Quotienten Q könnte ebenfalls mit dem t-Test auf Signifikanz getestet werden, wobei man sich wiederum auf ein zweckmässiges Signifikanzniveau verständigen müsste.

4.5. Stufe 2: Lohnbänder & Positionieren innerhalb der Lohnbänder

[Rz 46] Festlegen von generell-abstrakten Regeln für das Umrechnen von Punkten zu Lohnbeträgen:
Festlegen von «Lohnbändern»

Nachdem die Stellen (Arbeitsplätze) in Stufe 1 nach dem einheitlichen und neutralen Kriterienraster unabhängig von der Person der Stelleinhaber bewertet worden sind, müssen die derart erzeugten Punktezahlen in Lohnbeträge transformiert werden. Dadurch werden die erzielten Punktezahlen $V(\mathbf{x})$ (Punktzahlen, Scores) der Arbeitsplatzbewertung unmittelbar den «korrekten» Lohnklassen oder Lohnbändern zugeordnet.

Es muss dazu ein Lohnsystem festgelegt werden, welches für alle Punktezahlen der bewerteten Stellen einen Lohnbetrag L_i (V_i) oder, wie in den meisten Unternehmen üblich, ein Lohnband zuordnet: Diese Zuordnung generiert eine *Lohnlinie*, welche in den unterschiedlichsten geometrischen Formen verlaufen kann. Im vorliegenden

Zusammenhang ist von diesen Lohnlinien einzig zu fordern, dass sie zu *eindeutigen Zuordnungen* führen und dass sie wie alle methodischen Parameter der Lohnfindung durch *Konsens* der Beteiligten gedeckt und *betriebsintern transparent* ausgestaltet sein sollten. Um Diskussionen auf zwei unterschiedlichen Ebenen zu vermeiden, sollten dabei nach Möglichkeit alle kritischen Parameter der Lohnfindung schon bei der Arbeitsplatzbewertung (Stufe 1) ausdiskutiert und gelöst werden.⁶⁴

Zudem sollten für den Vergleich von individuellen Arbeitslöhnen die Lohnlinien $L_i(\mathbf{V}_i)$ *interpolierbar* ausgestaltet werden (zum besseren Verständnis wenn möglich: linear). Also: Steht für einen Vergleich eines individuellen Frauenlohnes kein Männerlohn mit der exakt identischen Punktzahl $V(\mathbf{x})$ des Arbeitsplatzes zur Verfügung, so kann ein hypothetischer Vergleichslohn durch Interpolieren zwischen den beiden benachbarten Männerlöhnen ermittelt werden.

Die benachbarten Lohnbänder eines niedriger und eines höher bewerteten Arbeitsplatzes weisen in der Regel *überlappende Wertebereiche* auf. Jemand, der auf einer höher bewerteten Stelle eingesetzt ist, aber aufgrund der z.B. geringeren Dienstzeit noch im unteren Bereich des entsprechenden Lohnbands eingestuft ist, kann daher den gleichen oder allenfalls sogar einen niedrigeren Lohn beziehen als jemand auf der nächst tiefer bewerteten Stelle, der aber z.B. aufgrund seiner langen Dienstzeit und überdurchschnittlicher Leistungen den oberen Teil seines Lohnbands erreicht hat. Wenn man aber den maximal erreichbaren Lohn in den entsprechenden Lohnbändern betrachtet, sollte dieser Maximallohn für die höher bewertete Stelle durchwegs oberhalb des Maximallohns für eine tiefer bewertete Position liegen.⁶⁵

[Rz 47] Festlegen von generell-abstrakten Regeln für das Positionieren der einzelnen Personen *innerhalb* der Lohnbänder (Stufe 2)

Wenn derartige Lohnbänder zum Einsatz kommen, was in den meisten Unternehmen der Fall ist, müssen auch für die Stufe 2 die Kriterien für die Einstufung der Mitarbeitenden *innerhalb* des Lohnbandes ihres Arbeitsplatzes *in generell-abstrakter Form* definiert werden.⁶⁶ Diese Kriterien bilden im Wesentlichen das sog. *Humankapital* ab, also diejenigen Elemente, welche der Stelleninhaber aufgrund seines Vorlebens (Erwerbsbiographie) mitbringt. Typische Kriterien dafür sind die Dienstjahre (*innerhalb* des Betriebes), allenfalls Berufserfahrung (*ausserhalb* des Betriebes), individuelle Leistungsmerkmale (Beispiel: erzielter Umsatz) allenfalls auch das Lebensalter, auch wenn letzteres aktuell eher an Bedeutung verliert.

Um das Ziel der Lohngleichheit zu erreichen, dürfen auch diese generell-abstrakten Kriterien der personenbezogenen zweiten Stufe für die Einstufung *innerhalb* des Lohnbandes nicht eines der beiden Geschlechter systematisch *benachteiligen*. Bei diesen Kriterien der Stufe 2 dürfte die Gefahr einer Diskriminierung allerdings bedeutend geringer sein, da diese Kriterien im Wesentlichen die bekannten Elemente des Humankapitals abbilden. Um dies sicherzustellen, gibt es wie bei der Stufe 1 mehrere Alternativen:

Die Kriterien werden von einer externen, neutralen Stelle begutachtet und gegebenenfalls korrigiert.

Und/Oder: Die Kriterien werden im Rahmen der Sozialpartnerschaft gemeinsam definiert und zweiseitig durch

Konsens vereinbart (Beispiel: Vertrag zwischen Geschäftsleitung des Arbeitgebers/in und Betriebskommission der Arbeitnehmer/innen, paritätische Bewertungskommission, Betriebsordnung gemäss Arbeitsgesetz Art. 37/38, soweit zulässig, betriebsinterner Zusatzvertrag / Bewertungsreglement zu einem GAV). Auch diese Kriterien / Regeln der Stufe 2 werden *betriebsintern* transparent gemacht, damit alle Beteiligten nachvollziehen können, auf welcher Grundlage ihre Einstufungen erfolgen, und sich bei einer vermuteten Ungleichbehandlung an eine geeignete Stelle (Ombudsstelle, Bewertungskommission o.ä.) wenden können.⁶⁷

Mit diesem zweistufigen Verfahren aus personenunabhängiger *Arbeitsplatzbewertung* (Stufe 1) und personenorientierter *Lohnbandeinstufung* (Stufe 2) kann auf generell-abstrakter Systemebene eine Lohnungleichbehandlung weitgehend ausgeschlossen werden. Es ist indessen erforderlich, regelmässige Stichprobenkontrollen durchzuführen, ob die Mitarbeitenden auch tatsächlich den Lohn erhalten, den sie gemäss Arbeitsplatzbewertung und Einstufungskriterien im Lohnband erhalten müssten.

[Rz 48] Stufe 2: Praktische Gesichtspunkte

Einheitlichkeit und Neutralität der Arbeitsplatzbewertung können in der Praxis nur gesichert werden, wenn es gelingt, die Stellen und die Personen, die sie besetzen, konsequent *auseinanderzuhalten*. Jemand, der die Mitarbeitenden persönlich kennt, ist *de facto* nicht in der Lage, deren Stelle neutral zu bewerten, ohne sich dabei durch seine eigene Einschätzung der Stelleninhaber beeinflussen zu lassen.⁶⁸ Die Arbeitsplatzbewertung Stufe 1 (= Punktzahl der individuellen Arbeitsplätze) sollte deshalb durch ein externes Beratungsunternehmen erfolgen, dessen Ansatz hinsichtlich Diskriminierungsfreiheit zertifiziert ist. Aus diesem Grund setzen die meisten Unternehmen externe Spezialisten für die Aufgabe der Stellenbewertung Stufe 1 ein. Wer den damit verbundenen, finanziellen Aufwand für die externen Beratungshonorare vermeiden möchte, kann aber auch auf andere Lösungen zurückgreifen, indem sich zum Beispiel innerhalb einer Branche oder einer Region die Personalverantwortlichen gegenseitig für die kategorienmässige Bewertung der Arbeitsplätze in den jeweils anderen Unternehmen zur Verfügung stellen.

4.6. Regressionsanalyse und alternative Lösungsvorschläge im Vergleich

Das vorgeschlagene zweistufige Verfahren aus den dargestellten zwei Stufen weist im Vergleich zu der multiplen Regressionsanalyse mehrere gewichtige Vorteile auf:

1. Stufe 1: generell-abstrakten Regeln für die personen-**un**abhängige Arbeitsplatzbewertung, Auswertung der gewonnenen Daten mit dem Verfahren der gewichteten Addition anstelle der Regressionsanalyse,
2. Stufe 2: Generierung von Lohnbändern für die einzelnen Punktwerte der Arbeitsplatzbewertung von Stufe 1 und personen**bezogene** Einstufung innerhalb der Lohnbänder aufgrund von individuellen Personenmerkmalen

Die meisten Unternehmen mit einem professionellen Personalmanagement setzen die beiden Stufen personen-**un**abhängige *Arbeitsplatzbewertung* und personen**bezogene** *Einstufungskriterien* bereits systematisch als integralen Bestandteil ihrer Personalpolitik ein und sie führen ebenfalls bereits routinemässig Stichprobenkontrollen zur Anwendung der Regeln durch. Dadurch ergibt sich für sie kein Mehraufwand, falls das System auch für die Überprüfung der Lohngleichheit eingesetzt und entsprechend für verbindlich erklärt würde.

Das zweistufige Verfahren führt bei richtiger Anwendung (Transparenz der Kriterien und Festlegen aller Beurteilungsparameter durch Konsens / Konvention) zu einer höheren Transparenz auf beiden Stufen, und zwar nicht nur auf der Stufe 1 der personenunabhängigen *Arbeitsplatzbewertung*, sondern auch auf der Stufe 2 der Kriterien für die personenbezogene *Lohnbandeinstufung* bis hin zur Findung des Individuallohns. Dadurch wird es den einzelnen Personen möglich, sich ein eigenes Bild von ihrem Lohn im Vergleich zu anderen Löhnen innerhalb des Betriebes zu machen und damit ungerechtfertigte subjektive Gefühle von Lohnbenachteiligung bis zu einem gewissen Grad zu vermeiden.

[Rz 49] Ergebnis

Es bleibt die Erkenntnis, dass multi-criteria Methoden **zwar** ohne Zweifel unentbehrlich für eine reproduzierbare und damit rechtssichere Anwendung multi-attributiver Gesetzesbegriffe wie «gleichwertige Arbeit» oder «benachteiligen» in Art. 3 GIG sind, falls *Rechtssicherheit* auch nur näherungsweise erreicht werden soll.⁶⁹

Aber zur Vermeidung der dargestellten Problembereiche der multiplen Regressionsanalyse (siehe Abschnitte 2.1 bis 2.4) stehen sowohl für die individuelle Lohnfindung und Lohnvergleiche als auch für die *prima facie* Beurteilung ganzer Betriebe alternative quantitative Methoden zur Verfügung, welche keine vergleichbaren Problembereiche aufweisen.

MATTHIAS MOELLENEY , Center for Human Resources Management & Leadership, Hochschule für Wirtschaft (HWZ), Zürich

FRITZ DOLDER, Prof.Dr.iur. Dr.sc.techn. ETH, (em) Universität Basel

Interessenkonflikte: Keine / unabhängig. Die Autoren sind nicht am gewerblichen Geschäft mit betrieblichen Lohnanalysen beteiligt. Die vorliegende Studie wurde weder durch den Bund noch durch Verbände oder Unternehmungen der Privatwirtschaft angeregt oder finanziell gefördert. M.M ist zurzeit Präsident der Zürcher Gesellschaft für Personalmanagement (ZGP).

Bei dem vorliegenden Beitrag handelt es sich um die Fortsetzung des Beitrages Fritz Dolder / Matthias Moelleney, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017.

1 RONALD OAXACA, Male-female wage differentials in urban labor markets, *International Economic Review* 14(3), 1973, 699: «It is clear that the magnitude of the estimated effects of discrimination *crucially depends upon the choice of the control variables* for the wage regressions. A researcher's choice of control variables *implicitly reveals his or her attitude* toward what constitutes discrimination in the labor market. If it were possible to control for virtually all sources of variation in wages, one could pretty well eliminate labor market discrimination as a significant factor in determining wage differentials by sex (or race)»; ähnlich THOMAS GEISER / MATHIAS MAUCHLE, Rechtsgutachten Marktlohn und konjunkturelle Lage als Rechtfertigungsgrund für geschlechtsspezifische Lohnunterschiede gemäss Art. 3 Abs. 2 GIG, Univ. St. Gallen, 10. Februar 2016, www.ebg.admin.ch/dam/ebg/de/dokumente/2016/06/rechtsgutachten_marketlohnundkonjunkturellegealsrechtfertigungs.pdf (alle Websites zuletzt besucht am 10. November 2016), S. 35: «Bei gleichwertiger Arbeit ist in erster Linie auf wissenschaftliche Arbeitsplatzbewertungsmethoden abzustellen. Welche Kriterien in den Bewertungsmethoden berücksichtigt werden, hängt jedoch wiederum von Wertentscheidungen ab, die sich nicht

- durchwegs wissenschaftlich begründen lassen.»; anders YVES FLÜCKIGER / ROMAN GRAF, Lohndiskriminierung und Entschädigung vor Gericht, in: Human Ressource Management Jahrbuch 2012, Weka Verlag, Zürich, S. 75–96, S. 95: «Die Methode (sc. multiple Regression mit Dummyvariablen FM) ist somit nur schlecht angreifbar und lässt nur *sehr wenig Spielraum für subjektive Betrachtungen*».
- 2 MATTHIAS MOELLENEY, Tages-Anzeiger 6. August 2014: «Eine Saläranalyse mit Logib gibt nur Resultate im Sinne des Erfinders.»; «Logib eignet sich nicht für alle Branchen zur Saläranalyse, denn je nach Zweig werden andere Kriterien für eine Funktion als wichtig erachtet.»; Beispiel von Piloten: Weder das Alter, noch das Dienstalter spielten eine wichtige Rolle, sondern Flugstunden und Landungen. Mit Logib «wäre es aber gut möglich, eine Salärdiskriminierung zu finden, weil viele Frauen in diesem männerlastigen Beruf noch jung sind». [und daher weniger verdienen...]; ähnlich DIETER URBAN / JOCHEN MAYERL, Regressionsanalyse, 4. A. Wiesbaden 2011, S. 61.
 - 3 PETER SCHETTGEN, Arbeit, Leistung, Lohn, Stuttgart 1996, S. 119/120; das von EBERHARD ULICH, Arbeitspsychologie, 7. A. Zürich / Stuttgart 2011, S. 155 bis 167 referierte Tätigkeitsbewertungssystem (TBS und TBS-GA für geistige Arbeit) verwendet auf diese Art ebenfalls fünf «Hauptkategorien» A bis E für die Bewertung von Arbeitstätigkeiten. Wie die Software Logib des Bundes zeigt (5 unabhängige Variable und eine Dummyvariable FM), scheint die Zahl Fünf (bzw. Sechs mit der Dummyvariablen) für die Beurteilungskategorien im Bereich der Arbeitsbewertung beinahe eine Suggestivwirkung auszuüben. Diese Zahlen entsprechen übrigens auch ersteren wissenschaftlichen Erkenntnissen, so der Empfehlung von DANIEL KAHNEMAN (Nobelpreis für Wirtschaft 2002), Thinking, Fast and Slow, London 2012, S. 232: «Don't overdo it – six dimensions is a good number.»; er verweist im übrigen (S. 227) auf den grossen Erfolg des lediglich aus fünf Kriterien bestehenden Test zur medizinischen Beurteilung von Neugeborenen von Dr. Virginia Apgar (1953). Vgl. dazu FRITZ DOLDER, [Rechtsanwendung mit Hilfe von multi-criteria Methoden](#), in: Jusletter 1. Juni 2015, Rz. 57.
 - 4 In diesem Sinne die Antwort des Bundesrats vom 9. November 2016: [16.3657](#) Motion Grüter Franz, Einreichungsdatum: 15. September 2016, Lohngleichheit im Beschaffungswesen. Aber fair und korrekt: «Die Verwendung von Logib ist weder im Bundesbeschaffungsrecht noch im üblicherweise von Bundesbeschaffungsstellen eingesetzten Selbstdeklarationsformular der Beschaffungskonferenz des Bundes vorgegeben. Der Nachweis der Lohngleichheitseinhaltung ist auch mit anderen Instrumenten möglich und zulässig, vorausgesetzt, dass diese *Lohnanalysen nach dem Standard-Analysemodell des Bundes erfolgen*, auf dem auch Logib basiert.» (Hervorhebungen in allen Fussnoten teilweise durch Autoren); Bericht des Bundesrates vom 18. November 2015 in Erfüllung des Postulats [14.3388](#) Noser vom 2. Juni 2014, S. 13: «Im bestehenden Standard-Analysemodell werden die Faktoren *Ausbildung, Dienstalter, potenzielle Erwerbserfahrung, Anforderungsniveau und berufliche Stellung* verwendet (siehe 5.1.). Das Modell beschränkt sich bewusst auf diese fünf Faktoren, damit es *auf alle Unternehmen gleich anwendbar* ist, für die Unternehmen einen möglichst geringen administrativen Aufwand mit sich bringt, und nur Faktoren ohne Diskriminierungspotenzial berücksichtigt.»; CHRISTINA FELFE / JUDITH TRAGESER / ROLF ITEN, Studie zu den statistischen Analysen der Eidgenossenschaft betreffend die Lohngleichheit von Frau und Mann, Auftraggeber: Eidgenössisches Büro für die Gleichstellung von Frau und Mann EBG, Letzte Überarbeitung: 2. November 2015, zitiert als FELFE et al. (2015), S. 37: «Das Modell beschränkt sich bewusst auf die oben genannten fünf Faktoren (plus das Geschlecht), um zu gewährleisten, dass das Modell *(1) auf alle Unternehmen gleich anwendbar ist*, weil diese die Daten im Rahmen der Lohnstrukturhebung des Bundes alle 2 Jahre liefern müssen, (2) möglichst wenig Aufwand bei den Unternehmen generiert und (3) entsprechend der rechtfertigenden Perspektive nur diskriminierungsfreie erklärende Variablen berücksichtigt werden.»; S. 88: «Für ein Standard-Instrument ist es jedoch wichtig, dass das Modell *für alle Unternehmen gleich angewendet* wird und es indirekte Lohndiskriminierungen *gänzlich ausschliessen* kann.»; S. 89: «Wie in Kapitel 3.1 ausgeführt, ist es nicht möglich, das Ausmass der Lohndiskriminierung mit einem standardisierten (d.h. *für alle Unternehmen gleich geltenden*) Analysemodell exakt nachzuweisen: es existieren grundsätzlich *weitere objektive Faktoren*, die die Leistung der Arbeitnehmerin bzw. des Arbeitnehmers beschreiben, die Gleichwertigkeit der Arbeit garantieren und sich in der Entlohnung niederschlagen können (siehe auch Kapitel 5.1 und 5.3.3), jedoch in der standardisierten Analyse *nicht berücksichtigt* werden. Die Toleranzschwelle trägt dieser Tatsache Rechnung.».
 - 5 SCHETTGEN (Fn. 3), S. 151: Gestaltungsempfehlung (2) für eine diskriminierungsfreie Arbeitsbewertung: «Das Bewertungssystem sollte möglichst spezifisch auf die Bedingungen bzw. Gegebenheiten des Unternehmens oder der Branche abgestimmt sein, damit die typischen Merkmale von Frauenberufen erfasst werden können.»; FELFE et al. (2015 [Fn. 4]), S. 106 (Ergebnisse der Umfrage): «Die in den Interviews befragten Unternehmen, die Logib als weniger oder überhaupt nicht geeignet beurteilt haben (15/50 Nennungen), geben hauptsächlich als Grund an, dass die Variablen *nicht auf die Strukturen und Realitäten in ihrem Unternehmen passen*. Als Gründe nennen diese Unternehmen die bereits oben aufgeführten Kritikpunkte (Funktion und Funktionsbewertung für den Lohn relevant und weniger Ausbildung/Dienstjahre; Fehlen der Leistung). Zudem kritisieren diese Unternehmen, dass *betriebs- oder branchenspezifische Eigenschaften sowie die Marktsituation zu wenig berücksichtigt* seien.»; FELFE et al. (2015 [Fn. 4]), S. 112/113 (Ergebnisse der Umfrage): «Gleichwohl sind sich die befragten Unternehmen einig, dass es aufgrund der Einfachheit des Modells die *Realitäten im Unternehmen nicht genau genug abbilde*.».
 - 6 Welches Kriterium soll zwischen diesen beiden gegensätzlichen Gesichtspunkten entscheiden? Die Effizienz der verschiedenen Modelle wird anhand der Qualität der Resultate (*goodness of fit*) und damit anhand des Determinationskoeffizienten R^2 bei *konstanter Zahl* der unabhängigen Variablen x_i beurteilt. Dies würde bedeuten, dass der gleiche betriebliche Personalbestand einmal mit den Variablen des Modells des EBG, einmal mit den Variablen eines branchen- bzw. regionsspezifischen Modells erfasst und ausgewertet wird. Das Modell, welches ein höheres R^2 bei *konstanter Zahl der Variablen* generiert, soll eingesetzt werden.
 - 7 URBAN/MAYERL (Fn. 2), S. 110: «*Giftliste S5*: Suche nach solchen Variablen, welche eine möglichst grosse inhaltliche Nähe

zur abhängigen Variable Y des Regressionsmodells aufweist. [...] Besonders hoch wird R^2 werden, wenn eine oder mehrere X-Variable(n) *das gleiche wie die Y-Variable messen* und dazu nur unterschiedliche Indikatoren benutzen.».

- 8 CHRISTIAN KATZ / CHRISTOF BAITSCH (2006), Arbeit bewerten – Personal beurteilen, Lohnsysteme mit Abakaba, 2. A. Zürich 2006, S. 40–43 haben in einem Abakaba-Projekt mit rund 1'300 einzelnen Funktionsbewertungen maximale Korrelationen von 0.73 (zwischen P1 und P4) und 0.65 (zwischen V1 und I1) gemessen.
Umschreibung der unabhängigen Variable Anforderungsniveau bei Logib:
Das System akzeptiert Zahlen von 1 bis 4 gemäss den folgenden Definitionen:
1 = Arbeitsplatz, der die Verrichtung anspruchsvoller und schwierigster Arbeiten beinhaltet.
2 = Arbeitsplatz, der sehr selbstständiges und qualifiziertes Arbeiten bedingt.
3 = Arbeitsplatz, der Berufs- und Fachkenntnisse voraussetzt.
4 = Arbeitsplatz mit einfachen und/oder repetitiven Tätigkeiten.
Berufliche Stellung: Das System akzeptiert Zahlen von 1 bis 5 gemäss den folgenden Definitionen:
1 = Oberstes und oberes Kader; 2 = Mittleres Kader; 3 = Unteres und unterstes Kader; 4 = Unterstes Kader, 5 = Ohne Kaderfunktion.
- 9 Als abhängige Variable Y verwendet Logib den
– Bruttolohn, inklusive Zulagen, 13. Monatslohn und Sonderzahlungen;
Unabhängige Variablen x_i des Humankapitals:
– Altersjahr (via Geburtsjahr), ergibt theoretisch mögliche (= potenzielle) Erwerbsjahre; Geschlecht (1, 2) = Dummyvariable;
– Dienstjahre (gemäss Eintrittsjahr in den Betrieb);
– Ausbildung, absolvierte, «mitgebrachte», 8 bzw. 9 Stufen.
Unabhängige Variablen x_i des Arbeitsplatzes:
– Anforderungsniveau, 4 Stufen;
– Berufliche Stellung, 5 Stufen, Beschäftigungsgrad.
- 10 FELFE et al. (2015 [Fn. 4]), S. 16: «Das Standard-Analysemodell schliesst insgesamt fünf Erklärungsfaktoren ein, anhand derer die Löhne zwischen Frauen und Männern verglichen werden. Dabei handelt es sich um drei Erklärungsfaktoren, welche die Unterschiede im *Humankapital* (Ausbildung, potenzielle Erwerbserfahrung, Dienstalter) messen und zwei Erklärungsfaktoren, die unterschiedliche *Anforderungen* an die Tätigkeit (Kompetenzniveau, berufliche Stellung) berücksichtigen. Die Analyse der wissenschaftlichen Literatur [?] zeigt, dass dies wesentliche lohnrelevante Merkmale sind, die im Wesentlichen denen entsprechen, die in klassischen Modellen zur Messung von Lohnungleichheiten verwendet werden. *Sie sind insofern geeignet*, Lohnungleichheiten in Unternehmen zu untersuchen.»; wiederholt: S. 126.
- 11 Bericht Bundesrat Noser (Fn. 4), S. 14: «E7 Die bislang im Standard-Analysemodell des Bundes verwendeten Faktoren *Ausbildung, Dienstalter, potenzielle Berufserfahrung, Anforderungsniveau und berufliche Stellung* werden als geeignet beurteilt und sollten beibehalten werden.».
- 12 DOLDER (Fn. 3), Abschnitt 3.2, Rz 26 ff.; dies entspricht spätestens seit dem Genfer Schema (1950) einem allgemein akzeptierten Konsens der Fachwelt für die Arbeitsbewertung: Beispiel: ULLICH (Fn. 3), S. 155 Tätigkeitsbewertungssystem (TBS) mit Tabellen 3.2, S. 157: E1 *Geforderte* berufliche Vorbildung, Tabelle 3.3, S. 159, E1 (identisch), Tab. 3.4 S. 162, E1 (identisch), Tab. 3.5 S. 166, E1 (identisch) auch für TBS-GA (für geistige Arbeit); ZGP Zürcher Gesellschaft für Personalmanagement, Vernehmlassung Änderung GIG, 23. Februar 2016, S. 3: «Ein Beispiel: Logib (und auch das BFS für die LSE) stellen in ihren Analysen auf den *höchsten erworbenen* Ausbildungsabschluss ab, während praktisch alle anderen Funktionsbewertungssysteme sinnvollerweise auf die für eine Funktion *notwendige* Ausbildung abstellen. Unter Fachleuten ist unbestritten, dass für die Lohnfindung in Unternehmen und Verwaltungen die *notwendige* Ausbildung entscheidend ist und *nicht der höchste erworbene* Abschluss. Wenn eine Geomatikerin in einem Kleidergeschäft im Atelier arbeitet und Änderungen vornimmt, muss sie mit den Änderungsschneider/innen verglichen werden, nicht mit Hochschulabsolvent/innen (was in einem konkreten Fall von einem kantonalen Gleichstellungsbüro mit Hinweis auf die korrekte Anwendung von Logib verlangt wurde). Die Erfassung des *höchsten Ausbildungsabschlusses* oder der tatsächlichen für eine Funktion *notwendigen* Ausbildung kann zu beträchtlichen Differenzen im Resultat führen. Die Aussagekraft einer Analyse ist daher beschränkt und bringt den Unternehmen kaum einen Mehrwert.».
- 13 DOLDER (Fn. 3), Abschnitt 3.2, Rz 26 ff.
- 14 Der Bund als Produzent des Standard-Analysemodells EBG und der Software *Logib* scheint nicht zu wissen, was der Bund als Gesetzgeber des öffentlichen Personalrechts getan hat (vgl. Matthaeus 6,3). Beispiele: Bundespersonalverordnung (BPV; [SR 172.220.111.3](#)) vom 3. Juli 2001 (Stand am 1. Januar 2016), Art. 52 Funktionsbewertung (Art. 15 BPG) Abs. 3: «Massgebend für die Bewertung sind die *erforderliche* Vorbildung, der Umfang des Aufgabenkreises sowie das Mass der betrieblichen Anforderungen, Verantwortlichkeiten und Gefährdungen.»; *ähnlich*: Verordnung des VBS über die Bewertung der besonderen Funktionen im VBS (Funktionsbewertungsverordnung VBS; [SR 172.220.111.343.1](#)) vom 21. Juni 2005 (Stand am 1. Januar 2016) Art. 2 Bewertungsgrundlagen; Verordnung des EDI über die Bewertung der besonderen Funktionen im EDI (Funktionsbewertungsverordnung EDI; [SR 172.220.111.343.2](#)) vom 28. März 2002 (Stand am 22. Juli 2003) Art. 2 Bewertungsgrundlagen; Verordnung des EDA zur Bundespersonalverordnung (VBPV–EDA; [SR 172.220.111.343.3](#)) vom 20. September 2002 (Stand am 1. Januar 2016) Art. 34; *ähnlich* der Kanton Zürich: Personalverordnung vom 16. Dezember 1998 ([LS 177.11](#)) § 8. Abs. 2: «Die Richtpositionen werden gemäss dem Verfahren der *«Vereinfachten Funktionsanalyse»* eingereiht. Massgebend sind die *vorausgesetzte* Ausbildung und Erfahrung, die mit der Stelle verbundenen geistigen *Anforderungen*, die Verantwortung, die psychischen und körperlichen *Anforderungen* und Belastungen, die *Beanspruchung* der Sinnesorgane und die besonders äusseren Arbeitsbedingungen, denen die Inhaberin

oder der Inhaber der Stelle ausgesetzt ist.»; Vollzugsverordnung zur Personalverordnung [der evangelisch-reformierten Kirche] vom 6. Juli 2011 (LS 181.401) § 40. Abs. 2: «Massgebend für die *Funktionsbewertung* sind: a. *vorausgesetztes Fachwissen*, b. *erforderliche* Kenntnisse von Strukturen und Abläufen, c. *geforderte* soziale Kompetenzen, [...]».

- 15 Anforderungsniveau (bis 2012 [Fn. 8]), ab 2012: Variable «Betriebliches Kompetenzniveau» mit den vier Kategorien: Tätigkeiten mit 1 [...] höchst komplexen Problemlöse-/ Entscheidungsaufgaben, 2 [...] komplexeren technischen oder praktischen Aufgaben, 3 [...] Aufgaben, die Berufs-/Fachkenntnisse voraussetzen, 4 [...] einfachen manuellen oder routinemässigen Aufgaben; zur Messung der Unabhängigkeit zwischen den Variablen vgl. KATZ/BAITSCH (2006 [Fn. 8]), S. 40 ff.
- 16 Berufliche Stellung: Definitionen der einzelnen Stufen: s. oben Fn. 8.
- 17 ZGP in DENARIS 01/2016, S. 18: «Fragwürdig sind hier vor allem die «(potenziellen) Erwerbsjahre» und die Dienstjahre. [...] Die potenziellen Erwerbsjahre werden so ermittelt: Alter minus Ausbildungsjahre (erneut auf dem höchsten Ausbildungsabschluss basierend) minus sechs Vorschuljahre. Diese Formel wird für Frauen und Männer gleich angewendet. Nach wie vor haben Frauen aber nach der Geburt von Kindern häufig eine Familienpause. Sie haben damit in der Regel weniger Berufserfahrung als gleichaltrige Männer; die Firmen legen den Lohn jedoch auf der Basis der tatsächlichen und *nicht der potenziellen* Berufserfahrung fest.».
- 18 Die Konstruktion der Erziehungsgutschriften stammt aus der Sozialversicherung: Bundesgesetz über die Alters- und Hinterlassenenversicherung (AHVG; SR 831.10) Art. 29^{sexies} (Eingefügt durch Ziff. I des BG vom 7. Oktober 1994 (10. AHV-Revision), in Kraft seit 1. Januar 1997 (AS 1996 2466; BBI 1990 II 1). Abs. 1 «Versicherten wird für diejenigen Jahre eine Erziehungsgutschrift angerechnet, in welchen ihnen die elterliche Sorge für eines oder mehrere Kinder zusteht, die das 16. Altersjahr noch nicht erreicht haben.»; ähnlich Art. 52e ff. der Verordnung über die Alters- und Hinterlassenenversicherung (AHVV; SR 831.101).
- 19 ZGP in DENARIS 01/2016, S. 18: «Ausserdem werden bei Logib die Dienstjahre recht stark gewichtet. Allerdings zählen nur die *innerhalb* des Unternehmens erarbeiteten Dienstjahre, die Erfahrung bei anderen Arbeitgebern zählt an dieser Stelle nicht. [...] Häufig sind aber Unternehmen bei einer Rekrutierung daran interessiert, *externe Erfahrung zu erwerben*. Nach Logib müssten die neuen Mitarbeitenden aber tiefere Löhne erhalten, *da sie nur wenige oder gar keine anrechenbaren Dienstjahre haben*.».
- 20 Genfer Tagung zur Arbeitsbewertung 1950, organisiert vom Comité International de l'Organisation Scientifique (CIOS), Vorschlag von Prof. Dr. Bramesfeld (TH Darmstadt) und Dr. Lorenz, Grundlage bildet eine systematische Gliederung der Arbeitsanforderungen. Gert Zülch Artikel «Arbeitsbewertung» in Handwörterbuch des Personalwesens, Gaugler / Weber (Hrsg.), 2.A. Stuttgart 1992, S. 68–75; KATZ / BAITSCH (2006) [Fn. 8]), S. 38/39 erfassen beispielsweise in ABAKABA im Psycho-sozialen Bereich (PS) psycho-soziale Belastungen als PS d in 6 verschiedenen Varianten und im Physischen Bereich (P) belastende arbeitszeitliche Bedingungen (P c) und belastende Umgebungsbedingungen (P d) in 6 Varianten (P d 1 bis P d 6) von der Unfallgefährdung (P d 1) über Lärm (P d 5) bis zu Kälte, Hitze und Arbeit im Freien (P d 6). Es ist nicht bekannt, warum diese Kriterien als unabhängige Variablen des Standard-Analysemodells EBG ausgeschlossen worden sind.
- 21 FELFE et al. (2015 [Fn. 4]), S. 88 «Das Standard-Analysemodell des Bundes wurde unter Einbezug der wissenschaftlichen Literatur sowie unter Berücksichtigung des Prinzips des gleichen Lohns für gleichwertige Arbeit entwickelt. Die Analyse des Erklärungsgehalts der fünf Erklärungsfaktoren zeigt, dass es sich um *relevante Faktoren* handelt. Ausgehend von der wissenschaftlichen Literatur *gäbe es allerdings durchaus noch weitere Variablen, die zusätzliche Lohnunterschiede erklären könnten*. Das Standard-Analysemodell *soll jedoch nach systematischer Lohndiskriminierung in Unternehmen suchen*. Daher dürfen keine Variablen einbezogen werden, die bereits Diskriminierungen implizieren könnten. Die Analyse des Diskriminierungspotenzials der möglichen zusätzlichen Variablen ergibt, dass lediglich *physische und psychische Belastungen* sowie Sprachkenntnisse prinzipiell diskriminierungsfrei operationalisiert werden könnten. Alle anderen Variablen erfüllen diese Voraussetzungen nicht. Im Einzelfall ist es möglich, dass bei Unternehmen zusätzliche lohnrelevante Merkmale keine indirekten Diskriminierungen beinhalten und das Unternehmen dies nachweisen könnte.»; a.a.O., S. 127: «Eine weitere Möglichkeit wäre, das Modell um Arbeitsbedingungen *zu erweitern*, welche im Rahmen von objektiven und geschlechtsneutralen Berufsbewertungen erhoben werden könnten. Die Machbarkeit und der Aufwand einer solchen Evaluation müssten jedoch vertieft geprüft werden. *Arbeitsbedingungen (z.B. psychische und physische Belastungen)* haben gemäss der wissenschaftlichen Literatur Potenzial, Lohnungleichheiten zwischen Frauen und Männern zusätzlich zu erklären». [...]; a.a.O. S. 126 «*Darüber hinaus wird von verschiedenen Seiten die Berücksichtigung zusätzlicher Erklärungsfaktoren gefordert, welche einen Einfluss auf Lohnunterschiede zwischen Frauen und Männern haben können*.»; a.a.O. S. 129/130: «Empfehlung 8: [...] Geprüft werden könnte jedoch, ob Arbeitsbedingungen wie die *physische und psychische Belastung* anhand einer unabhängigen und geschlechtsneutralen Evaluation der Berufe als zusätzliche Variablen in das Modell aufgenommen werden könnten.».
- 22 Bericht Bundesrates Noser (Fn. 4), S. 14: «E7 Die bislang im Standard-Analysemodell des Bundes verwendeten *Faktoren Ausbildung, Dienstalder, potenzielle Berufserfahrung, Anforderungsniveau und berufliche Stellung* werden als geeignet beurteilt und sollten beibehalten werden.».
- 23 Bericht Bundesrat Noser (Fn. 4), S. 13 «Die Studienautorinnen und der Studienautor haben die folgenden im Postulat genannten Faktoren geprüft: Beschäftigungsgrad in der Berufskarriere (tatsächliche Berufserfahrung), Weiterbildungen, Sprachkenntnisse und Führungserfahrungen. Darüber hinaus haben sie auch die Faktoren Arbeitszeitmodelle (Schicht-, Abend-, Nacht- oder Wochenendarbeit) sowie physische und psychische Belastung am Arbeitsplatz untersucht. Den

Erklärungsgehalt dieser Faktoren für Lohnunterschiede zwischen Frauen und Männern schätzt die Studie folgendermassen ein (siehe dazu auch oben stehend Ziff. 4.2.2.2.): *tatsächliche Berufserfahrung: sehr hoch; Arbeitszeitmodelle und physische und psychische Belastung: hoch; Weiterbildungen: mittel; Sprachkenntnisse und Führungserfahrung: eher tief.* Gemäss der Studie ist bei den Faktoren tatsächliche Berufserfahrung und Weiterbildungen eine aussagekräftige Anwendung schwierig resp. fast unmöglich zu bewerkstelligen. Bei den Faktoren Arbeitszeitmodellen und *physische und psychische Belastung* wird dagegen von einem tiefen resp. eher tiefen Erhebungsaufwand ausgegangen.»; Bericht Bundesrat Noser, a.a.O. S. 14: «E8 Die im Postulat genannten Faktoren *Beschäftigungsgrad in der Berufskarriere (tatsächliche Berufserfahrung), Weiterbildungen, Sprachkenntnisse und Führungserfahrungen* sind *nicht geeignet*, um in das Standard-Analysemodell des Bundes aufgenommen zu werden. Diese Faktoren weisen ein Diskriminierungspotenzial auf und/oder der mit der Erhebung der dazu notwendigen Informationen verbundene administrative Aufwand für die Unternehmen wäre im Vergleich zu deren (zusätzlichem) Erklärungsgehalt unverhältnismässig hoch. *Von den zusätzlich geprüften Faktoren ist nur der Faktor Arbeitszeitmodelle diskriminierungsfrei. Die Faktoren physische und psychische Belastungen bergen zwar die Gefahr der Geschlechterdiskriminierung, letztere kann jedoch durch eine entsprechende Anwendung ausgeschlossen werden. Das Erklärungspotenzial dieser Faktoren soll in empirischen Analysen getestet werden.*».

- 24 Bundespersonalverordnung (BPV; SR 172.220.111.3) vom 3. Juli 2001 (Stand am 1. Januar 2016) Art. 52 Funktionsbewertung (Art. 15 BPG): Abs. 3 «Massgebend für die Bewertung sind die erforderliche Vorbildung, der Umfang des Aufgabenkreises sowie das Mass der betrieblichen Anforderungen, Verantwortlichkeiten und *Gefährdungen*.»; Verordnung des EDI über die Bewertung der besonderen Funktionen im EDI (Funktionenbewertungsverordnung EDI; SR 172.220.111.343.2) vom 28. März 2002 (Stand am 22. Juli 2003) Art. 2: Bewertungsgrundlagen 1 «Massgebend für die Funktionsbewertung sind die erforderliche Vorbildung, der Umfang des Aufgabenkreises sowie das Mass der betrieblichen Anforderungen, *Belastungen* und Verantwortlichkeiten. [...]»; Verordnung des VBS über die Bewertung der besonderen Funktionen im VBS (Funktionenbewertungsverordnung VBS; SR 172.220.111.343.1) vom 21. Juni 2005 (Stand am 1. Januar 2016) Art. 2 Bewertungsgrundlagen Abs. 1: «Massgebend für die Funktionsbewertung sind die erforderliche Vorbildung, der Umfang des Aufgabenkreises sowie das Mass der betrieblichen Anforderungen, Verantwortlichkeiten und *Gefährdungen*. [...]»; Kanton Zürich: Personalverordnung vom 16. Dezember 1998 (LS 177.11) § 8 Abs. 2: «Die Richtpositionen werden gemäss dem Verfahren der «Vereinfachten Funktionsanalyse» eingereiht. Massgebend sind die vorausgesetzte Ausbildung und Erfahrung, die mit der Stelle verbundenen geistigen Anforderungen, die Verantwortung, die *psychischen und körperlichen Anforderungen und Belastungen, die Beanspruchung der Sinnesorgane und die besondern äussern Arbeitsbedingungen, denen die Inhaberin oder der Inhaber der Stelle ausgesetzt ist.*».
- 25 Interessiert man sich für R^2 , so empfiehlt die Praxis der empirischen Sozialforschung, als Entscheidungsgrenze den Wert $R^2 = 0.60$ festzulegen. URBAN / MAYERL (Fn. 2), S. 59 R^2 : «... scheint er auch in idealer Weise dazu geeignet zu sein, mehrere Modell-schätzungen miteinander vergleichbar zu machen. [...] Dementsprechend ist in der Praxis der Regressionsanalyse oftmals auch eine Jagd nach einem hohen R^2 zu beobachten. *Die Stunde des Glücks scheint demjenigen geschlagen*, dessen Regressionsmodell ein R^2 von *über 0.60* aufweist. Bei näherem Hinsehen erweist sich der Wert des Determinationskoeffizienten jedoch als zweischneidige Angelegenheit.»; FLÜCKIGER / GRAF (Fn. 1) haben für betriebliche Regressionsanalysen je nach Rechenbeispiel zwischen $R^2 = .708$ (Tab. 1, S. 90) und $R^2 = .748$ (Tab 3, S. 92) und $R^2 = .887$ (Tab. 2, S. 90, nur Verwaltungspersonal) berechnet. DOLDER (Fn. 3) Rz. 25, Abschnitt 3.5 (Rz. 38 bis 45) und Rz. 128: allgemeine Überlegungen für die Festlegung von Entscheidungsgrenzen bei normativer Verwendung von multi-criteria Methoden.
- 26 Eidgenössisches Büro für die Gleichstellung von Frau und Mann EBG, Standard-Analysemodell zur Überprüfung der Lohngleichheit zwischen Frau und Mann im Beschaffungswesen des Bundes (Methodenbeschrieb), Anleitung zur Durchführung der standardisierten Überprüfung, Bern, Juni 2016, zitiert als Methodenbeschrieb 2016, S. 8/9 gibt folgendes Rechenbeispiel: Test, ob Diskriminierungseffekt statistisch signifikant grösser ist als Toleranzschwelle (Bsp.: Unternehmen X mit $N = 105$ Arbeitnehmerinnen und Arbeitnehmern):
Ergebnisse Standardregression:

Diskriminierungskoeffizient D (β_5)	-0.138
Standardfehler des Diskriminierungseffekts (S.E. von β_5)	0.043
Toleranzschwelle S	-0.050
Teststatistik $t = (D-S) / S.E.$	-2.049
Kritischer t-Wert $t_{krit.}$ (bei Signifikanzniveau 0.05)	-1.660

Entscheidung:

Falls $t \leq t_{krit.}$ folgt:

Signifikanz = Ja, da $-2.049 < -1.66$ [bzw. $+ 2.049$ grösser als $+ 1.66$].

- 27 Der Standardfehler $SE_b = \sqrt{[\sum (y_i - \hat{y}_i)^2 / (N - 2) + \sum (x_i - \bar{x})^2]}$, (formelmässige Herleitung z.B. bei URBAN / MAYERL [Fn. 2], S. 147, Gleichung 3.9) misst die Präzision der Schätzung eines Regressionskoeffizienten (hier β_5 der Dummyvariablen 1/0, Frau/Mann). Je kleiner SE, desto präziser diese Schätzung. Die Auswirkungen von SE auf das Resultat (Betrieb ist «verdächtig» / «nicht verdächtig») können erheblich sein. Auszugehen ist dabei von der «Daumenregel», dass ein Regressionskoeffizient gegenüber seinem SE *mindestens doppelt so gross* sein muss, um bei einem Signifikanzniveau von $\alpha = 5\%$ (0.95) im einseitigen Test signifikant zu werden und die Nullhypothese H_0 zurückweisen zu können (URBAN /

MAYERL [Fn. 2], S. 151). Zu berücksichtigen sind dabei folgende Gesichtspunkte: Ein negativer Testausgang (β_5 : nicht signifikant) kann durch einen *grossen SE* verursacht werden. Ein solcher kann einmal dadurch entstehen, dass der Term $\sum (y_i - \hat{y}_i)^2$ im Zähler von SE sehr gross wird, weil das Regressionsmodell falsch bzw. schlecht spezifiziert ist. Zum anderen dadurch, dass die Varianz von x_i gering wird und der Nenner der zitierten Gleichung (3.9) $\sum (x_i - \bar{x})^2$ daher sehr klein.

Die zitierte «Daumenregel» wird *ceteris paribus* besser erfüllt, wenn die Werte 1 und 0 der Dummyvariablen *ungleich häufig* vorkommen, also im untersuchten Betrieb die Geschlechter *ungleich häufig* verteilt sind. In einem Rechenbeispiel mit $N = 40$ und konstanter Lohnverteilung ergab die Verteilung 50/50 ein $\beta_5 = -.0602$ mit einem $SE = .0532$, also *nicht* signifikant, während die gleichen Löhne bei einer Verteilung von 25/75 ein $\beta_5 = -.2992$ mit einem $SE = .0519$, also signifikant, oder bei einer Verteilung 75/25 ein $\beta_5 = -.3029$ mit einem $SE = .0512$, also ebenfalls signifikant waren. Die zitierte Daumenregel wird *ceteris paribus* auch dann besser erfüllt, wenn die beobachteten Löhne y_i *stärker variieren* und damit der Term $\sum (y_i - \hat{y}_i)^2$ im Zähler der zitierten Gleichung (3.9) gross wird. Rechenbeispiel ($N = 40$): Bei einer Verteilung von 25/75 der Geschlechter ergibt sich bei einem (logarithmierten) Maximallohn von 4.079 ein $\beta_5 = -.2992$ mit einem $SE = .0519$, dagegen bei einem erhöhten (logarithmierten) Maximallohn von 4.301 ein $\beta_5 = -.5891$ mit einer $SE = .1075$; beide Beispiele sind problemlos signifikant.

Wird das Zahlenbeispiel im Methodenbeschrieb 2016 (Fn. 26), S. 9 mit einem kleineren Wert von β_5 und einem kleineren $SE = 0.022$ variiert, so wird die Auswirkung von SE auf die Signifikanz und damit auf das Resultat («verdächtig» – nicht) klar illustriert:

$\beta_5 = -.138$	$SE = .043$	$T = 2.049$	signifikant, da $t(krit) = 1.660$ (=Methodenbeschrieb)
	.022	4.000	sig. da > 1.66

Variiertes Zahlenbeispiel (bei unverändertem $N = 105$ und Toleranzschwelle $S = .050$):

$\beta_5 = -.111$.043	$T = 1.419$	nicht sig. da < 1.66
	.022	2.773	sig. da > 1.66

Die SE können zwar durchaus im Bereich des halben Wertes (0.043) des Regressionskoeffizienten β_5 liegen und dadurch der erwähnten «Daumenregel» entsprechen (so das Zahlenbeispiel aus dem «Methodenbeschrieb 2016 (Fn. 26)»). Aber diese SE von β_5 können auch erheblich von diesem Wert abweichen. Zudem hätte der um $S=0.050$ verminderte Wert von $D (= \beta_5)$ einen *anderen SE*, wenn er «im Original» diesen Wert erreicht hätte. Darf für den verminderten Wert von T einfach derjenige SE verwendet werden, welcher für das unverminderte $D (= \beta_5)$ «nature» gegolten hätte? Damit bleibt offen, wie vor Ort im Betrieb SE prognostiziert und allenfalls beeinflusst werden kann. Wird der Anwender der Regressionsanalyse vor Ort im Betrieb verstehen, dass zwei Betriebe mit dem gleichen oder minimal verschiedenen Regressionskoeffizienten β_5 einmal als «verdächtig», einmal als «nicht verdächtig» qualifiziert werden, weil der Standardfehler SE bei beiden Betrieben unterschiedlich ist? Dass der Betrieb es im Ergebnis also eher dem SE als dem Regressionskoeffizienten β_5 selbst «verdankt», dass er als «verdächtig» qualifiziert wird?

- 28 FELFE et al. (2015 [Fn. 4]), S. 7: «Die derzeit gültige Toleranzschwelle von 5% hat sich, *obgleich nicht wissenschaftlich begründbar*, in der Praxis bewährt. Sofern zusätzliche Variablen hinzugefügt würden, müsste eine Absenkung der Toleranzschwelle geprüft werden.»; a.a.O. S. 89: «6.1 Nutzen der Toleranzschwelle im Rahmen der Kontrollen: *Um der Tatsache Rechnung zu tragen*, dass sich in einem Unternehmen auch weitere objektive Faktoren, die in der standardisierten Analyse nicht berücksichtigt werden, auf die Entlohnung auswirken können, wurde eine Toleranzschwelle von 5% sowie ein statistischer Signifikanztest bei den Kontrollen im *öffentlichen Beschaffungswesen* eingeführt. Wie in Kapitel 3.1 ausgeführt, *ist es nicht möglich*, das Ausmass der Lohndiskriminierung mit einem standardisierten (d.h. für alle Unternehmen gleich geltenden) Analysemodell exakt nachzuweisen: es existieren grundsätzlich *weitere objektive Faktoren*, die die Leistung der Arbeitnehmerin bzw. des Arbeitnehmers beschreiben, die Gleichwertigkeit der Arbeit garantieren und sich in der Entlohnung niederschlagen können (siehe auch Kapitel 5.1 und 5.3.3), jedoch in der standardisierten Analyse *nicht berücksichtigt* werden. Die Toleranzschwelle trägt dieser Tatsache Rechnung. [...] Der *Nutzen der Toleranzschwelle* liegt insbesondere darin, dass sie die Wahrscheinlichkeit gering hält, dass Unternehmen fälschlicherweise der Lohndiskriminierung beschuldigt werden. Damit kommt die Toleranzschwelle sowohl Forderungen nach praktischer Handhabbarkeit als auch nach einer Anwendung mit Augenmass entgegen»; a.a.O. S. 89: «6.2. Einfluss von Modellvariationen auf die Toleranzschwelle: *Die Höhe der Toleranzschwelle von 5% hat weder eine theoretische, empirische, noch juristische Fundierung*. Die «richtige» Höhe der Toleranzschwelle theoretisch und empirisch zu bestimmen, ist nicht möglich. Dazu müssten alle weiteren objektiven Faktoren bekannt, beobachtbar und messbar sein, die einen Einfluss auf die Lohnungleichheit haben. Zusätzlich müsste man das Diskriminierungspotenzial jedes einzelnen Faktors abschätzen können [...]; a.a.O. S. 127/128: «9.2.3. Toleranzschwelle Unter den gegebenen Bedingungen, dass im Einzelfall Lohnunterschiede in Unternehmen mit weiteren – nicht im Analysemodell – berücksichtigten objektiven Faktoren erklärt werden könnten, ist eine *Toleranzschwelle notwendig*. Ohne die Toleranzschwelle würden vermutlich viele Unternehmen *fälschlicherweise in Verdacht geraten*, Lohndiskriminierungen zu praktizieren. Die *Toleranzschwelle* und der zusätzliche Signifikanztest hält die Wahrscheinlichkeit insbesondere für kleine und mittlere Unternehmen für *falsche Verdächtigungen* gering. Die Kehrseite davon ist, dass durch die Toleranzschwelle auch Unternehmen *fälschlicherweise* von einem Verdacht *freigesprochen* werden können. Inwieweit die Höhe der Toleranzschwelle von 5% *angemessen* ist, den richtigen Grad zwischen *falschen Verdächtigungen* und *falschem Freisprechen* zu bilden, lässt sich auf Basis von theoretischen oder empirischen Grundlagen *nicht schlüssig* beantworten. Grund hierfür ist, dass die Diskussion bezüglich aller möglichen Erklärungsfaktoren nicht endgültig abgeschlossen ist und nicht alle Variablen diskriminierungsfrei gemessen werden können. Daher kann die Angemessenheit der Toleranzschwelle nur aufgrund der bisherigen praktischen Erfahrungen beurteilt werden. [...]; Bericht Bundesrat Noser (Fn. 4), S. 14: «Die Toleranzschwelle von 5% hat *keine eigentliche theoretische oder juristische Grundlage*. Es ist deshalb

für die Studienautorinnen und den Studienautoren schwierig, die Angemessenheit der Toleranzschwelle zu beurteilen.».

- 29 Bericht Bundesrat Noser (Fn. 4), S. 14: [Empfehlungen] «E9 Beim bestehenden Standard-Analysemodell sollte die Toleranzschwelle von 5% beibehalten werden. *Sollten zusätzliche diskriminierungsfreie Faktoren in das Standard-Analysemodell aufgenommen werden, müsste die Toleranzschwelle entsprechend gesenkt werden.* E10 Die Zweckmässigkeit der Höhe der Toleranzschwelle sollte nach weiteren gewonnenen Erfahrungen wieder überprüft werden.».
- 30 FELFE et al. (2015 [Fn. 4]), S. 9: «Toleranzschwelle von 5%. Das bedeutet, dass beim Standard-Analysemodell des Bundes eine geschlechtsspezifische Lohnungleichheit von über 5% Anlass zu einer *begründeten Vermutung* systematischer Lohndiskriminierung im Unternehmen gibt. Zusätzlich muss die Bedingung erfüllt sein, dass die geschlechtsspezifische Lohnungleichheit statistisch signifikant über der Toleranzschwelle liegt. Mit der Toleranzschwelle und dem zusätzlichen Signifikanztest sollen mögliche Unsicherheiten aufgefangen werden, welche daraus resultieren können, dass das Standard-Analysemodell keine möglichen weiteren diskriminierungsfreien, unternehmensspezifischen Faktoren berücksichtigt.»; FELFE et al. (2015 [Fn. 4]), S. 38 «Dazu wird das Standard-Analysemodell des Bundes verwendet. Ergibt die Analyse, dass die *unerklärte Lohnungleichheit* unter der Toleranzschwelle von 5% liegt oder diese nicht statistisch signifikant überschritten wird, gelten die Gesetzesvorgaben des BöB *als erfüllt*. [...] Im anderen Fall – wenn das Ergebnis signifikant über der Toleranzschwelle liegt – wird dies als eine *begründete Vermutung* für Lohndiskriminierung gewertet. Das EBG kann dem Unternehmen in diesem Fall eine Frist von sechs bis zwölf Monaten gewähren. Innerhalb dieser Frist kann das Unternehmen Korrekturmaassnahmen ergreifen und einen qualifizierten Nachweis erbringen, dass Lohngleichheit gewährleistet wird.»; SILVIA STRUB, 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, Bern, Juni 2004, S. 43, hat noch vorgeschlagen, die Methode der Zerlegung nach Oaxaca/Blinder (1973), nicht die Technik der Dummyvariablen zu verwenden: «Für eine zukünftige Anwendung des Instrumentariums schlagen wir deshalb vor, wo möglich die etwas aufwändigere, dafür genauere Oaxaca-Methode zu verwenden. Bei kleineren Firmen kann die Einhaltung der Lohngleichheit mittels Dummy-Methode vorgenommen werden.».
- 31 FELFE et al. (2015 [Fn. 4]), S. 7: «Die derzeit gültige Toleranzschwelle von 5% hat sich, obgleich nicht wissenschaftlich begründbar, in der Praxis bewährt.».
- 32 FELFE et al. (2015 [Fn. 4]), S. 34, Tabelle 3: Ergebnisse von Analysen der Lohndifferenzen auf nationaler Ebene; ähnlich FLURINA SCHMID, (FORS, Uni Lausanne), The gender gap in Switzerland over time, Schweizerische Zeitschrift für Soziologie 42 (3), 2016, S. 443–467 (November 2016): 0.065 / 0.125 / 0.265 (S. 454, je nach Kohorte). FELFE et al. (2015 [Fn. 4]), S. 109: «Die Höhe von 5% erachtet etwa ein Drittel der Unternehmen als angemessen, etwas mehr als ein Drittel als zu tief und etwas weniger als ein Drittel als *zu hoch*. Auffällig ist, dass Unternehmen, die bereits eine Analyse mit Logib durchgeführt haben, die Toleranzschwelle *häufiger als eher zu hoch* beurteilen als solche, die noch keine Erfahrungen mit Logib haben.».
- 33 FELFE et al. (2015 [Fn. 4]), S. 7: «Sofern zusätzliche Variablen hinzugefügt würden, müsste eine *Absenkung* der Toleranzschwelle geprüft werden.»; a.a.O. S. 130 «Toleranzschwelle: Empfehlung 9: Sollten zusätzliche Variablen in das Standard-Analysemodell aufgenommen werden, müsste dies mit einer *Absenkung der Toleranzschwelle* einhergehen.».
- 34 FELFE et al. (2015 [Fn. 4]), S. 89/90: «Inwiefern müsste die Toleranzschwelle *angepasst* werden, wenn zusätzliche Variablen in das Modell einbezogen würden, die einen weiteren Teil der Lohnungleichheit erklären können? Klar ist, dass die Toleranzschwelle *nach unten angepasst* werden müsste. In welcher Höhe dies zu erfolgen hat, hängt davon ab, wie stark der zusätzliche Erklärungsfaktor den unerklärten Anteil der Lohnungleichheit im Durchschnitt der Unternehmen senkt. Inwiefern eine Veränderung der statistischen Methode eine *Veränderung der Toleranzschwelle* nach sich ziehen würde, wurde bisher nicht diskutiert. Wichtig ist hervorzuheben, dass eventuelle Mängel in der statistischen Methode nicht Thema bei der Bestimmung der Toleranzschwelle waren. Dementsprechend ist es a priori unklar, wie mit der Toleranzschwelle umgegangen werden soll, wenn mit alternativen statistischen Methoden andere Ergebnisse in Bezug auf den *unerklärten Anteil* der Lohnungleichheit gemessen werden. Bei der Verwendung einer *alternativen Methode*, [??] welche zu einem tieferen *Anteil der unerklärten Lohnungleichheit* führt, *müsste die Toleranzschwelle nach unten angepasst werden*. Ebenso stellt sich die Frage, ob eine *Anpassung* der Toleranzschwelle nötig wäre, wenn – unter dem Wissen, dass eine alternative Methode zu tieferen Ergebnissen führen würde – die bestehende Methode beibehalten würde. Eine *exakte Quantifizierung* einer *Anpassung* der Toleranzschwelle bei einer *Anpassung* des Standard-Analysemodells – sowohl in punkto Modellvariablen als auch statistischer Methode – ist aus oben genannten Gründen *nicht möglich*. Die Festlegung der Toleranzschwelle verlangt daher ein Vorgehen, das sich primär auf Einschätzungen von ExpertInnen abstützt. [...]»; FELFE et al. (2015 [Fn. 4]), S. 127/128: «9.2.3. Toleranzschwelle: Eine Veränderung der statistischen Methode und der einbezogenen Variablen im Standard- Analysemodell müsste eine *Anpassung der Toleranzschwelle* nach sich ziehen, da der Unsicherheitsbereich der Lohndiskriminierung dadurch verändert würde. In welchem Ausmass dies zu erfolgen hätte, ist aber *schwierig zu quantifizieren*, zumal dazu in der Schweiz die notwendigen Daten fehlen. Eine *Anpassung der Toleranzschwelle* unter Erhöhung der Aussagekraft der Methode würde bei den Unternehmen vermutlich aber auf wenig Akzeptanz stossen. [...]»; FELFE et al. (2015 [Fn. 4]), S. 130 «Empfehlung 9: Sofern das Standard-Analysemodell nicht mit zusätzlichen, diskriminierungsfreien Variablen erweitert werden kann (vgl. Empfehlung 3), sollte die *Toleranzschwelle von 5% beibehalten* werden. Sollten zusätzliche Variablen in das Standard-Analysemodell aufgenommen werden, müsste dies mit einer *Absenkung der Toleranzschwelle* einhergehen. Empfehlung 10: Die Zweckmässigkeit der *Höhe der Toleranzschwelle* sollte nach weiteren gewonnenen Erfahrungen wieder überprüft werden, da sich die heutige Beurteilung auf wenige Erfahrungswerte abstützt.».
- 35 Beispiele derartiger Studien: Die zitierten (FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der](#)

- multiplen Regression – Teil 1**, in: Jusletter 22. Mai 2017, Fn. 12) Untersuchungen von ALAN S. BLINDER, Wage discrimination: reduced form and structural estimates, *Journal of Human Resources* 8, 1973, S. 436–455; RONALD OAXACA, Male-female wage differentials in urban labor markets, *International Economic Review* 14(3), 1973, S. 693–709; FLURINA SCHMID (Fn. 32).
- 36 FELFE et al. (2015 [Fn. 4]), S. 30: «Weil die Lohngleichheitsanalyse auf diese Weise *alle möglichen erklärenden Variablen einschliessen kann*, gilt für die individuelle Ebene «Nulltoleranz»; Urteil des Bundesgerichts [4A_585/2015](#) vom 11. April 2016 – Laufenburg: Schon 6% Abweichung sind diskriminierend, ohne die Berechnung im einzelnen zu hinterfragen: «4. Die Feststellungen der Vorinstanz zur Gleichwertigkeit der Arbeit der Beschwerdegegnerin und derjenigen von C. werden von der Beschwerdeführerin nicht gerügt. Sie erhebt auch keine Einwände dagegen, dass die Vorinstanz trotz der *eher geringfügigen Lohndifferenz von durchschnittlich 6%* im streitgegenständlichen Zeitraum von 2008 bis 2013 die Lohndiskriminierung als im Sinn von Art. 6 GIG glaubhaft gemacht beurteilte. *Damit muss es sein Bewenden haben und der angefochtene Entscheid ist insoweit nicht zu überprüfen (vgl. E. 2.1 hiervor).*».
- 37 Auch der vielfach zitierte Modellfall [BGE 130 III 145](#), 149, bildet insofern keine Ausnahme: FRITZ DOLDER / MATTHIAS MOELLENEY, **Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1**, in: Jusletter 22. Mai 2017, Fn. 19; zu dieser Problematik eingehend HEIDI STUTZ et al., Evaluation der Wirksamkeit des Gleichstellungsgesetzes, Synthesebericht, BASS, Bern /Binningen 2005, S. IV: «Der Umgang mit Konfliktfällen auf Betriebsebene zeigt, dass es nach wie vor alles andere als selbstverständlich ist, gegen eine geschlechtsbedingte Diskriminierung am Arbeitsplatz vorzugehen. Die *Angst vor Kündigung* erweist sich hier als ein Haupthindernis.»; a.a.O., S. 99 Abbildung 30 «Einfluss der *Angst vor Kündigung* als Grund für ausbleibende Klagen nach GIG»; a.a.O., S. 101 «Aus der *Befragung der Schlichtungsstellen* geht hervor, dass der Anteil weiter bestehender Arbeitsverhältnisse schon bei Eingang der Schlichtungsverfahren mit 37% nicht hoch ist. Fast zwei Drittel der Arbeitnehmenden wehren sich also erst *nach Beendigung des Arbeitsverhältnisses*. Bis zur Erledigung der Verfahren sinkt dieser Anteil auf 29%. Eine separate Analyse für den öffentlich-rechtlichen und den privatrechtlichen Bereich verweist jedoch auf grosse Unterschiede: Die öffentlich-rechtlichen Arbeitsverhältnisse bestehen nach der Schlichtung zu 60% weiter, bei den *privatrechtlichen* liegt der entsprechende Anteil *bei ganzen 15%*. Bei den Gerichtsentscheiden ist in der *Statistischen Grobauswertung der Gerichtsentscheide* häufiger nicht bekannt, ob das Arbeitsverhältnis weiterläuft. Sonst gleicht sich das Bild: Mindestens die Hälfte der Arbeitnehmenden ziehen erst *nach Beendigung des Arbeitsverhältnisses* vor Gericht. Bis zum Gerichtsentscheid sinkt der Anteil weiter bestehender Arbeitsverhältnisse auf 27% (21% unbekannt). Zwischen der Hälfte und zwei Drittel der Arbeitsverhältnisse existieren zum Erledigungszeitpunkt also nicht mehr. Die öffentlich-rechtlichen Arbeitsverhältnisse bestehen bei 45% der Entscheide weiter. Da bei einem Drittel das Weiterbestehen des Arbeitsverhältnisses nicht bekannt ist, kann davon ausgegangen werden, dass der reale Wert noch deutlich höher liegt. Im privatrechtlichen Bereich liegt der entsprechende Anteil bei ganzen 7% – *bei 84%* der Gerichtsentscheide dagegen besteht das Arbeitsverhältnis mit Sicherheit nicht mehr.».
- 38 FRITZ DOLDER / MATTHIAS MOELLENEY, **Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1**, in: Jusletter 22. Mai 2017, Abschnitt 2.2 Fn. 42 und 43.
- 39 Quelle: BFS Bundesamt für Statistik: Statistik der Unternehmensstruktur STATENT, Stand 10. Oktober 2016; selbst FLÜCKIGER / GRAF (Fn. 1) räumen diese Schwierigkeiten ein, S. 86, empfehlen aber, in solchen Fällen auf Informationen des *Referenzmarktes* zuzugreifen.
- 40 JOHAN GALTUNG, Theory and methods of social research, Oslo 1967, S. 241: «(4) The index should be simple. Obviously, the simpler of two otherwise equal indices will always be preferred.»; a.a.O. S. 242: «The index should be interpretable. The index should have a meaning over and above that given to it by its mathematical definition.»; DOLDER (2015 [Fn. 3]), Rz. 58 (Fn. 41) und Rz. 64 (Fn. 45). Ein Indiz für derartige Kommunikationsschwierigkeiten dürfte auch der Umstand bilden, dass die Software Logib des Bundes, obschon seit 2010 gratis erhältlich, in der Wirtschaft bisher *mit Zurückhaltung* aufgenommen worden ist, so die Ergebnisse der Umfrage von 2015, referiert bei FELFE et al., (2015 [Fn. 4]), S. 102: «Dabei geben insgesamt 20% der mittleren Unternehmen (mit 50–249 Mitarbeitenden), 37% der grossen Unternehmen (mit 250–999 Mitarbeitenden) und 65% der sehr grossen Unternehmen (mit 1'000 und mehr Mitarbeitenden) an, bereits eine Lohngleichheitsanalyse in irgendeiner Form durchgeführt zu haben. [...] *Eine statistische Regressionsanalyse haben erst wenige Unternehmen angewandt, insbesondere wenige mittlere Unternehmen (mittlere: 3%, grosse: 13%, sehr grosse: 39%, [...]).* Die meisten dieser Unternehmen haben dabei Logib verwendet (mittlere: 3%, grosse: 12%, sehr grosse 25%).».
- 41 Aufgabe: Versuchen Sie einmal, dem Personalschef eines KMU mit vielleicht 50 Angestellten zu erklären, warum die Regressionsanalyse mit der Technik der Dummyvariablen F/M und nicht z.B. eine Regressionsanalyse mit dem Determinationskoeffizienten R^2 und geschlechtergetrennten Häufigkeitsverteilungen zur Auswertung verwendet wird! Dagegen wird die Forderung nach unmittelbarer *Verständlichkeit* von der analytischen Arbeitsplatzbewertung nach dem Modell der gewichteten Addition (SAW, z.B. ABAKABA) wohl problemlos erfüllt (s. unten Abschnitt 4.3). Zu den Problemen der Gerichte: ELISABETH FREIVOGEL, Kommentar GIG, 2. A. 2009, N. 119 bis 127 zu Art. 3 GIG, N. 105 zu Art. 3: «Sie (= Gerichte) sind dazu *aufgrund fehlender eigener Sachkenntnisse* in Arbeitsbewertungsfragen häufig nicht selbst in der Lage und müssen sich diese Sachkenntnis mittels Einholen von Expertisen verschaffen, ansonsten sie den Anspruch auf rechtliches Gehör verletzen.»; die Defizite im Verständnis mathematisch-statistischer Zusammenhänge illustriert beispielsweise eine Urteilsbegründung der I. Sozialrechtlichen Abteilung in einem Lohngleichheitsfall nach GIG: Urteil des Bundesgerichts [8C_1006/2012](#) vom 10. April 2013; unbefriedigend ist die Situation insbesondere, wenn zwar Verständnisprobleme bei den Gerichten bestehen und entsprechende Beweisanträge einer Partei vorliegen, aber dennoch *kein Sachverständigengutachten eingeholt* wird. Der bekannte Satz *ludex non calculat* steht nämlich quantitativen Methoden in der Entscheidungsfindung keineswegs entgegen. Dieser Satz ist in der Rechtsgeschichte gründlich missverstanden worden. Er stammt aus dem römischen Zivilprozess und besagt einzig, dass blosse Rechenfehler in einem Urteil keine

Appellation erfordern, sondern bloss zu *berichtigen* sind: Dig. 49.8.0. «Quae sententiae sine appellatione rescindantur. Dig. 49.8.1.1 Macer 2 de appellat: *Item si calculi error in sententia esse dicatur, appellare necesse non est: veluti si iudex ita pronuntiaverit: «cum constat titium seio ex illa specie quinquaginta, item ex illa specie viginti quinque debere, idcirco lucium titium seio centum condemno»: nam quoniam error computationis est, nec appellare necesse est et citra provocationem corrigitur.* (...)»; Die Quote von 44% der Gerichtsfälle, in denen Sachverständigengutachten eingeholt werden, ist nach unserer Ansicht *deutlich zu niedrig*. HEIDI STUTZ et al., (2005 [Fn. 36]) S. 35: «In der andern knappen Hälfte der Lohngleichheitsfälle (44%) wurde mindestens ein gerichtliches Gutachten eingeholt. Davon waren die Hälfte Funktionengutachten zu ganzen Berufsgruppen, die andere Hälfte Individualvergleiche. Bei letzteren handelte es sich immer um den Vergleich inhaltlich gleicher oder sehr ähnlicher Arbeit. In fast allen Fällen folgte das letzturteilende Gericht im Wesentlichen dem Gutachten (die erstinstanzlichen manchmal nicht).».

- 42 Der Vernehmlassungsentwurf zur Revision des GIG sieht für die Methode der Lohnfindung vor:
«Art. 13b Methode der Lohnanalyse:
1 Die Lohnanalyse ist nach einer anerkannten Methode durchzuführen.
2 Die zuständige Behörde veröffentlicht eine Liste der anerkannten Methoden.
3 Der Bundesrat legt die Kriterien fest, die bei der Anerkennung einer Methode zu berücksichtigen sind. Er hört vorgängig die Arbeitgeber- und Arbeitnehmerverbände an.»;
der *Konsens* zwischen den Beteiligten und ihren Interessenvertretungen als Grundlage jeder Arbeitsbewertung wird von der deutschen Bundesregierung im Entwurf eines Lohngleichheitsgesetzes besonders betont (Deutscher Bundestag, Drucksache 18/11133, vom 13. Februar 2017, Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Förderung der Transparenz von Entgeltstrukturen (Beschluss der Bundesregierung vom 11. Januar 2017), S. 3: «Das Gesetz stellt sicher, dass die Verantwortung für eine transparente und faire Lohngestaltung, *insbesondere die Arbeitsbewertung, bei den Tarifvertragsparteien bleibt*, entlässt diese aber gleichzeitig nicht aus der Verantwortung, Entgeltgleichheit zu gewährleisten.».
- 43 Probleme könnten sich hinsichtlich des zulässigen Inhalts von Betriebsordnungen nach Art. 38 Abs. 2 ArG ergeben: «Die vereinbarte Betriebsordnung kann *auch andere Bestimmungen* enthalten, die das Verhältnis zwischen dem Arbeitgeber und den Arbeitnehmern betreffen, jedoch nur soweit, als ihr Gegenstand in dem Bereich, dem der Betrieb angehört, *nicht üblicherweise durch Gesamtarbeitsvertrag oder durch andere kollektive Vereinbarung geregelt wird*.».
- 44 Antwort Bundesrat Grüter (Fn. 4): «Die Verwendung von Logib ist weder im Bundesbeschaffungsrecht noch im üblicherweise von Bundesbeschaffungsstellen eingesetzten Selbstdeklarationsformular der Beschaffungskonferenz des Bundes vorgegeben. Der Nachweis der Lohngleichheitseinhaltung ist auch mit anderen Instrumenten möglich und zulässig, vorausgesetzt, dass diese Lohnanalysen nach dem Standard-Analysemodell des Bundes erfolgen, auf dem auch Logib basiert.» [sic!].
- 45 Empfehlung der EU-Kommission vom 7. März 2014, zur Stärkung des Grundsatzes des gleichen Entgelts für Frauen und Männer durch *Transparenz*, (2014/124/EU), Amtsblatt L69, 112: Empfehlung 2: «2. Die Mitgliedstaaten sollten öffentliche und private Arbeitgeber sowie die Sozialpartner ermutigen, *Transparenzmassnahmen* in Bezug auf die Zusammensetzung und Struktur der Löhne und Gehälter zu beschliessen. Sie sollten gezielte Massnahmen zur Förderung der *Lohntransparenz* ergreifen.»; dazu auch ZGP Vernehmlassung Änderung GIG, 23. Februar 2016, S. 4; «Um Lohngleichheit sicherstellen zu können, sind solche diskriminierungsfreien Einstufungskriterien absolut entscheidend, oder anders ausgedrückt: wenn ein Unternehmen keine *Kriterien für die Einstufung* (möglichst schriftlich) definiert hat, lässt sich überhaupt nicht feststellen, ob gegen den Grundsatz der Lohngleichheit verstossen wurde oder nicht.».
- 46 Dazu HEIDI STUTZ et al., Synthesebericht (2005 [Fn. 36]), S. V: «Fehlender Zugang zu Informationen, Unklarheiten in Bezug auf die Datengrundlagen: Beides führt dazu, dass nicht effizient gegen (vermutete) Diskriminierungen vorgegangen werden kann. Mögliche Massnahmen: *Förderung Lohntransparenz*, Verpflichtung zur *Offenlegung betrieblicher Lohninformationen* zu früherem Zeitpunkte im Konfliktfall, Schaffung von Regeln für offenzulegende Informationen und die Erstellung von Lohngutachten, konsequente Umsetzung der Beweislastleichterung und ihre Ausdehnung auf Nichtanstellung und sexuelle Belästigung.»; Empfehlung der EU-Kommission (2014/124/EU) vom 7. März 2014 (Fn. 45), a.a.O., Empfehlung 3: «3. Die Mitgliedstaaten sollten geeignete und verhältnismässige Massnahmen ergreifen, um zu gewährleisten, dass Arbeitnehmer nach Geschlecht aufgeschlüsselte *Informationen* zur Höhe der Löhne und Gehälter für die Gruppen von Arbeitnehmern *anfordern können*, die eine gleiche oder gleichwertige Arbeit verrichten.»; Deutscher Bundestag, Drucksache 18/11133, vom 13. Februar 2017, Gesetzentwurf der Bundesregierung, Entwurf eines Gesetzes zur Förderung der Transparenz von Entgeltstrukturen (Beschluss der Bundesregierung vom 11. Januar 2017 [Fn. 42]); «§ 10 Individueller Auskunftsanspruch, (1) Zur Überprüfung der Einhaltung des Entgeltgleichheitsgebots im Sinne dieses Gesetzes haben Beschäftigte einen Auskunftsanspruch nach Massgabe der §§ 11 bis 16. [...] Dazu haben die Beschäftigten in zumutbarer Weise eine gleiche oder gleichwertige Tätigkeit (Vergleichstätigkeit) zu benennen. [...] § 12 Reichweite, (3) Bei der Beantwortung eines Auskunftsverlangens ist der *Schutz personenbezogener Daten* der auskunftsbegehrenden Beschäftigten sowie der vom Auskunftsverlangen betroffenen Beschäftigten zu wahren. Insbesondere ist das Vergleichsentgelt *nicht* anzugeben, wenn die Vergleichstätigkeit von *weniger als sechs Beschäftigten* des jeweils anderen Geschlechts ausgeübt wird. Dabei ist sicherzustellen, dass nur die mit der Beantwortung betrauten Personen Kenntnis von den hierfür notwendigen Daten erlangen.».
- 47 Der Faktor F muss nicht zwingend über den gesamten Lohnbereich konstant gestaltet werden; er sollte aber auch nicht allzu stark davon abweichen. Es liegt z.B. nahe, den Faktor F in den höheren Lohnsegmenten etwas zu verkleinern. Dabei muss In Stufe 1 besonders darauf geachtet werden, dass von den Personen der aktuellen Stelleninhaber abstrahiert wird:

SCHETTGEN (Fn. 3), S. 151, Gestaltungsempfehlung «(4) Es muss sichergestellt werden, dass die Arbeitsplätze selbst und nicht die ArbeitsplatzinhaberInnen bewertet werden.».

- 48 Zum Gegensatz zwischen betriebsspezifischen und landesweiten Beurteilungsrastern vgl. die bereits oben (Fn. 5) zitierten Ergebnisse der Umfrage von FELFE et al. (2015 [Fn. 4]): S. 106 (Ergebnisse der Umfrage): «Die in den Interviews befragten Unternehmen, die Logib als weniger oder überhaupt nicht geeignet beurteilt haben (15/50 Nennungen), geben hauptsächlich als Grund an, dass die Variablen *nicht auf die Strukturen und Realitäten in ihrem Unternehmen passen*. [...] Zudem kritisieren diese Unternehmen, dass *betriebs- oder branchenspezifische Eigenschaften sowie die Marktsituation zu wenig berücksichtigt seien*...»; a.a.O. S. 112/113 (Ergebnisse der Umfrage): «Gleichwohl sind sich die befragten Unternehmen einig, dass es aufgrund der Einfachheit des Modells die *Realitäten im Unternehmen nicht genau genug abbilde*»; SCHETTGEN (Fn. 3), S. 151, «Box 3.2: Gestaltungsempfehlungen für eine diskriminierungsfreie Arbeitsbewertung nach «comparable worth», (2) Das Bewertungssystem sollte *möglichst spezifisch* auf die Bedingungen bzw. Gegebenheiten des Unternehmens oder der Branche abgestimmt sein, damit die typischen Merkmale von Frauenberufen erfasst werden können.».
- 49 Dies ist eine Frage der Berücksichtigung von betrieblichen Besonderheiten: Wenn in einem Betrieb (Beispiel: Bauunternehmung) Körperkraft oder (Beispiel: Airline) absolvierte Flugstunden eines Linienpiloten eine zentrale Rolle spielen, warum sollen diese nicht im Gegensatz zu anderen Betrieben (Beispiel: Grossbanken, Versicherungen) *anders* gewichtet werden als andere Kriterien (Beispiel: Intellektueller Bereich oder Gefühlsbereich)? Beispiele für branchenspezifische Kriterien gemäss ABAKABA: Fachliche Anforderungen, I a1 Erforderliche Grundausbildungen, I a3 Erforderliche fachliche Erfahrung: Flugstunden von Piloten, gefahrene Kilometer von Lokomotivführern und LkW-Chauffeuren, PS d4 Konfrontation mit Problemen und Leid anderer Personen.
- 50 FELFE et al. (2015 [Fn. 4]), S. 40, 3.4.2.: Alternative Analyseinstrumente «Zur Messung von Lohndiskriminierung in Unternehmen hat sich neben der statistischen Regressionsanalyse die Methode der *arbeitswissenschaftlichen Arbeitsbewertung / Funktionsbewertung* etabliert. Diese Methode basiert auf psychologischen Arbeitsmarkttheorien. Mit Hilfe dieser Methode kann der Arbeitswert einer Funktion erfasst und die Frage der Gleichwertigkeit (im juristischen Sinn) geklärt werden (SCHÄR MOSER und BAILLOD 2006). Dabei wird jeder Funktion über die Bewertung verschiedener Kompetenzen (z.B. Fähigkeiten, Verantwortung, Erfahrungen), die diese Funktion erfordert, ein bestimmter Wert zugewiesen. Dieser Wert bildet die Grundlage zur Bestimmung des Funktionslohnes. Es existieren viele verschiedene Methoden der Bewertungssystematik. In der Schweiz bieten eine Reihe von privaten Beratungsfirmen Lohngleichheitsanalysen an, welche sich vorwiegend auf die Methode der Arbeitsbewertung/Funktionsbewertung abstützen. Teilweise kombinieren diese Anbieter die Methode mit Regressionsanalysen. Auch die Methode der analytischen Arbeitsbewertung ist durch das Bundesgericht zugelassen (z. B. BGE 125 II 385), da sie es erlaubt, die Löhne bei gleichwertiger Arbeit zu vergleichen.»; nicht beigepflichtet werden kann indessen der weiteren Argumentation von FELFE et al., (2015 [Fn. 4]): «Die statistische Lohnanalyse ist gegenüber der analytischen Arbeitsbewertung dabei etwas flexibler. [??] Sie erlaubt es, gleichzeitig den Einfluss der Anforderungen an die Funktion sowie den Einfluss individueller Merkmale der Arbeitnehmenden (z.B. Erfahrung, Dienstalter) zu berücksichtigen. [??]».
- 51 Das Bundesgericht hat bereits im Entscheid BGE 124 II 409, 436 (1998: Handarbeitslehrerinnen / Primarlehrer Kt. ZH) als zulässige und *diskriminierungsfreie* Kriterien – wenn richtig angewendet – aufgezählt: Ausbildung, Alter, Dienstalter, Qualifikation, Erfahrung, Aufgabenbereich, Leistung; BGE 131 II 393, 399 (Erw. 5.2 und 6, Kanton Solothurn) verwendet die Methode der Vereinfachten Funktionsanalyse (VFA) anhand von 6 Kriterien: Krankenpflegeberufe vs. Polizei; GEISER / MAUCHLE (Fn. 1), S. 7: «Bewertungskriterien sind gemäss dem zur Diskussion stehenden Urteil (= BGE 124 II 409 Erw. 9d) dann geschlechtsspezifisch, wenn sie von den Angehörigen des einen Geschlechts wesentlich leichter oder anteilmässig erheblich häufiger erfüllt werden können als von den Angehörigen des anderen Geschlechts. In Frage kämen demnach Eigenschaften, welche bei Angehörigen des einen Geschlechts aus biologischen Gründen im Durchschnitt weniger häufig erfüllt seien (bspw. Körpergrösse oder -kraft), oder Eigenschaften, welche auf in der gesellschaftlichen Realität vorhandene geschlechtsspezifische faktische Ungleichheiten zurückzuführen seien.»; ähnlich GEISER / MAUCHLE (Fn. 1), S. 25/26 und 35/36.
- 52 Dieses schön klingende Postulat der *Diskriminierungsfreiheit* von Beurteilungskriterien ist zwar schnell formuliert, aber in der betrieblichen Realität weit schwieriger korrekt umzusetzen. Beispiele: Dürfen von einer Fluggesellschaft absolvierte Flugstunden eines Linienpiloten als erforderliche Ausbildung / berufliche Erfahrung honoriert werden? Dürfen von einer internationalen Spedition die gefahrenen Strassenkilometer von Chauffeuren als Berufserfahrung honoriert werden? Dürfen bei Ärzten/Ärztinnen die Anzahl bestimmter ausgeführter chirurgischer Operationen als Berufserfahrung verlangt und honoriert werden? Obschon offensichtlich nur sehr wenige und zwar jüngere Frauen derartige Kriterien erfüllen werden?
- 53 GEISER / MAUCHLE (Fn. 1), S. 35 «In einer dem Grundsatz nach marktwirtschaftlich organisierten Gesellschaft scheint es indessen unumgänglich, dass der Markt bei wirtschaftlichen Entscheidungen mitberücksichtigt wird. In welcher Art und Weise die Berücksichtigung des Marktes erfolgen kann, hängt davon ab, ob es sich bei den zu vergleichenden Arbeitstätigkeiten um gleiche oder um gleichwertige Arbeit handelt.».
- 54 Allerdings übt die Rechtsprechung *Zurückhaltung*; sie verlangt eine *zeitliche Beschränkung* der Berücksichtigung des Arbeitsmarktes (bis maximal ein Kalenderjahr): GEISER / MAUCHLE (Fn. 1), S. 24 Kriterien der I. Zivilabteilung des Bundesgerichts für die Rechtfertigung durch den Arbeitsmarkt: «Die Berücksichtigung der konjunkturellen Lage bzw. der starken individuellen Verhandlungsposition eines Arbeitnehmers bei der Festlegung des Lohnes muss einem wirklichen unternehmerischen Bedürfnis entsprechen.»; weitere Kriterien: Verhältnismässigkeit (geeignet und erforderlich), temporäre Beschränkung, Bereinigung der Lohndifferenz innerhalb von höchstens einem Kalenderjahr; weitere Einzelheiten GEISER /

MAUCHLE (Fn. 1), S. 26 und 35.

- 55 So die bereits zitierte (Fn. 50) Ansicht von FELFE et al. (2015 [Fn. 4]), S. 40 (3.4.2. Alternative Analyseinstrumente).
- 56 Beispiel: CHRISTIAN KATZ / CHRISTOF BAITSCH (1996), Lohngleichheit für die Praxis, 1. A. Zürich 1996, mit Fragebogen ABAKABA S. 70 ff, und Formulareil S. 101–114; CHRISTIAN KATZ / CHRISTOF BAITSCH (2006 [Fn. 8]), Arbeit bewerten – Personal beurteilen, Lohnsysteme mit Abakaba, 2.A. Zürich 2006, S. 38–43 (Kriterienraster und Punktwerte).
- 57 CHRISTIAN LEIPERT (1978), Gesellschaftliche Berichterstattung, Berlin / Heidelberg / New York 1978, S. 83 Standardisierungsproblem allgemein: «Erst nach Durchführung der Standardisierung ist eine Vergleichbarkeit der Indikatoren untereinander, die unabdingbare Voraussetzung für eine Zusammenführung von Indikatoren, gegeben.»; HENNER KLEINWEFERS (2007), Wohlfahrtsökonomie, Stuttgart 2008, S. 251 zum Human Development Index HDI: «Der Leser sollte sich klarmachen, dass die *Normierung* der entscheidende Trick ist, welcher formal die sachliche Aggregation durch gewichtete Addition dimensionsloser Zahlen ermöglicht.»; Skalierung in der Arbeitsplatzbewertung: KATZ / BAITSCH (2006 [Fn. 8]), S. 38: «Die Gesamtpunktzahl pro Merkmalsbereich kann 250 Punkte nicht übersteigen, so dass die Maximalpunktzahl pro Funktion 1000 Punkte beträgt. Diese Konzeption ermöglicht die Vergleichbarkeit zwischen den Merkmalsbereichen...»; für die Skalierung / Standardisierung der Kriterien *in der Rechtsanwendung*: DOLDER (2015 [Fn. 3]) Abschnitt 3.3 (Rz. 33–35) sowie Abschnitt 4.2 (Rz. 77–79).
- 58 Gewichte allgemein: LEIPERT (1978 [Fn. 57]), S. 75 ff.: Gewichtungproblem bei der Formulierung eines Indexes der Lebensqualität; Gewichte für die Arbeitsbewertung: SCHETTGEN (Fn. 3) S. 121–124; ABAKABA von KATZ / BAITSCH, (1996 [Fn. 56]), S. 40 und (2006 [Fn. 8]) S. 34–36 sieht folgende Gewichte vor: Intellektueller Bereich 55–70%, psycho-sozialer Bereich 5–20%, physischer Bereich 5–10%, Führungsverantwortung 15–30%; Gewichte bei SCHETTGEN (Fn. 3), S. 141, Tabelle 3.3: 5 Gruppen von Firmen, total 76 Firmen: Anforderungsgruppen Können 25.1 bis 47.2%, Belastung 19.6 bis 35.0%, Verantwortung 18.4 bis 42.6%, sonstige Einflüsse 1.6 bis 19.9%; vgl. DOLDER (2015 [Fn. 3]), Abschnitt 3.4 Rz. 36, 37 sowie Fn. 26, 27; Beurteilung von *Gewichten* in der Rechtsprechung: BGE 126 II 222: «Beide Funktionen sind dennoch nicht gleichwertig, weil die Kriterien «Ausbildung» und «psychische Beanspruchung» nicht gleich schwer zu *gewichten* sind: Für das erste Kriterium gilt die *Gewichtung* 300 und für das zweite die *Gewichtung* 60.»; BGE 131 II 393, 402 Erw. 6.4. «Indessen sind Werte zwangsläufig subjektiv. Jedes Besoldungssystem beruht notwendigerweise auf bestimmten Werthierarchien, die nicht nach wissenschaftlichen Kriterien richtig oder falsch sein können. Das ist unvermeidlich und daher auch zulässig. Unzulässig ist eine Werthierarchie nicht schon deshalb, weil sie traditionell ist, sondern nur, wenn sie ohne sachlichen Grund ein Geschlecht benachteiligt. Nachdem wie dargelegt sowohl die weiblichen Funktionen der Beschwerdeführerinnen als auch die männlichen der Vergleichsfunktionen von dieser *schwächeren Gewichtung* der Gefühlsarbeit benachteiligt sind, liegt keine Geschlechtsdiskriminierung vor, sondern eine geschlechtsneutrale Wertungsfrage, die legitimerweise so oder anders beantwortet werden kann.»; FREIVOGEL (Fn. 41), N. 114 zu Art. 3 GIG.
- 59 Zum Equalweighting: KAHNEMAN (Fn. 3), S. 226: «A formula that combines these predictors with *equal weights* is likely to be *just as accurate* in predicting new cases as the multiple regression formula that was optimal in the original sample.»; ROBYN M. DAWES (1979), The robust beauty of improper linear models in decision making, *American Psychologist* 34, S. 571–582, Begründung 577: «Why? because linear models are robust over deviations from optimal weighting. [...] The solution to the problem of obtaining optimal weights is one that [...] has a «flat maximum».»; LEIPERT (1978 [Fn. 57]), 86 f: «Wirft man einen Blick in die Literatur zur *Lebensqualitätsmessung*, so fällt auf, dass die Gleichrangigkeit aller Indikatoren und Komponenten als *Gewichtungstyp* vorherrschend ist (...). Es ist mithin zu beachten, dass die Entscheidung für eine Gleichgewichtung aus *rein praktischen Erwägungen* erfolgt.»; zum Equalweighting in der Rechtsanwendung: DOLDER (2015 [Fn. 3]). Rz. 36, 37 und Rz. 86–88.
- 60 Quelle: BFS Bundesamt für Statistik: Statistik der Unternehmensstruktur STATENT, Stand 10. Oktober 2016 (Fn. 39).
- 61 Rangkorrelationen: SIDNEY SIEGEL, Nonparametric Statistics, New York 1956, S. 202–204 (example without ties). Das Verfahren könnte etwa in folgenden Einzelschritten zusammengefasst werden:
- Erstelle eine Liste von N verschiedenen Lohnklassen oder Klassen von Arbeitsplatz-Punkten v_i im Betrieb.
 - Trage für jede dieser Klassen den Häufigkeitsrang (Zahl der betroffenen Personen) in der Verteilung der Frauen (A) und in der Verteilung der Männer (B) ein.
 - Bestimme die Differenz d_i zwischen den beiden Häufigkeitsrängen (A und B) einer Klasse, quadriere diese und berechne die Summe der Quadrate: $\sum d_i^2$
 - Berechne daraus Spearman's Rangkorrelation nach der Formel $r_s = 1 - 6 \sum d_i^2 / (N^3 - N)$
 - Wenn $N > 10$ (grosse Stichproben): Bestimme die Signifikanz von r_s mit Student's t-Test und verwende dazu als Freiheitsgrade die Zahl $N - 2$ der verwendeten Lohnklassen.
- Zwar läge natürlich auch eine Varianzanalyse ANOVA nahe zur Beantwortung der Frage, ob die beiden separaten Verteilungen von Frauen und Männern aus derselben Grundgesamtheit stammen. Aber ANOVA baut (wie bekannt) wiederum darauf auf, dass die beteiligten Zufallsvariablen Normalverteilungen aufweisen und diese Voraussetzung dürfte bei der betrieblichen Lohnanalyse wie dargelegt (FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017, Abschnitt 2.1) häufig bis regelmässig nicht erfüllt sein.
- 62 DOLDER (2015 [Fn. 3]). Fn. 40; zum Kolmogorov-Smirnov «Two-sample Test»: SIEGEL (Fn. 61), S. 127 ff, Summary of procedure S. 135, Tabelle M S. 279; zur Verteilung von D: N. SMIRNOV (1948), Tables for estimating the goodness of fit of empirical distributions, Ann.Math.Statist. 19 (1948), S. 280–281; eine andere Methode wird von FLÜCKIGER / GRAF (Fn. 1), S. 84 vorgeschlagen (2.1 Erster Schritt zur Feststellung von Diskriminierung); sie setzt aber einen Vergleich der beobachteten Verteilung «mit jener des *Referenzmarktes*» voraus und dies erscheint uns aufgrund der Überlegungen in Abschnitt 2.3

(Institutionelle Beschränkungen) als problematisch.

- 63 Gini-Koeffizient: Einzelheiten des von Corrado Gini 1912 anhand der Lorenzkurve formulierten Index als Mass der Ungleichheit z.B. bei HELMUT KROMREY, Empirische Sozialforschung, 9.A. Opladen 2000, S. 436–438 oder GÜNTER BUTTLER / REINHOLD STROH, Einführung in die Statistik, 7. A. Reinbek 2000, S. 174–176, (relative Konzentration); der Gini-Koeffizient kann durch die ebenfalls nicht-parametrische Methode der Quintilsanalyse S80 / S20 aus der EU-Statistik (EURO-STAT) vereinfacht werden, bei der der Quotient zwischen den Lohnsummen der 20% höchsten (S80) und der 20% niedrigsten Löhne (S20) berechnet und ebenfalls als Indikator für Ungleichheit verwendet wird. S80/S20 Einkommensquintilverhältnis nach Geschlecht und nach Altersklassen – EUSILC Erhebung [ilc_di11], Letzte Aktualisierung: 28. März 2017, Quelle der Daten: Eurostat; auf der *nationalen Ebene* sind die Unterschiede des S80/S20 zwischen den Geschlechtern allerdings *bescheiden*: Quotienten S80/S20 Schweiz, alle Altersklassen: Frauen max. 4.9 (2008) bis min. 4.2 (2013); Männer max. 4.8 (2007 und 2008) bis min. 4.3 (2013); Schweiz, unter 65. Altersjahren Frauen max. 4.7 (2008) bis min. 3.9 (2013), Männer max. 4.7 (2008) bis min. 4.1 (2013); auf der Ebene von *Einzelbetrieben* liegen bisher, soweit uns zugänglich, keine derartigen S80/S20-Werte vor. Es erscheint aber *prima facie* zumindest nicht als ausgeschlossen, dass die Werte auf *betrieblicher Ebene* stärker variieren als auf der nationalen Ebene und dass deshalb S80/S20 einen brauchbaren *prima facie* Indikator für den *Gender Wage Gap* auf Betriebsebene ergeben könnte. *Die Autoren danken Herrn Dr.sc.math Andreas Adelman vom Paul-Scherrer-Institut (PSI) für die Hinweise zur S80/S20-Statistik der EU.*
- 64 Es kann im vorliegenden Kontext selbstverständlich keine Diskussion über den Klassiker des zweckmässigen oder optimalen Verlaufs von Lohnlinien geführt werden: Beispiele von Verläufen z.B. bei SCHETTGEN (Fn. 3), S. 293–295.
- 65 Aus den dargestellten Gründen (siehe FRITZ DOLDER / MATTHIAS MOELLENEY, [Betriebliche Lohnanalyse mit Hilfe der multiplen Regression – Teil 1](#), in: Jusletter 22. Mai 2017 Abschnitt 2.3: innerbetrieblicher Charakter) ist es für die praktische Anwendung von Art. 3 GIG von untergeordneter Bedeutung, dass die Lohnbänder auch zum Vergleich der Lohnstruktur eines Betriebes mit dem Arbeitsmarkt und der Lohnstruktur der ausserbetrieblichen «Umgebung» dienen können.
- 66 ZGP Vernehmlassung Änderung GIG, 23. Februar 2016 (Fn. 45), S. 5: «b) Positionierung innerhalb eines Lohnbandes Wenn die Bewertung der Stellen diskriminierungsfrei erfolgt ist, kommt als weitere mögliche Quelle für Lohnungleichheit die Einstufung der Mitarbeiterinnen und Mitarbeiter *innerhalb des Lohnbandes* in Frage, das ihrer Funktion zugeordnet ist. Deswegen muss in einem zweiten Schritt auch überprüft werden, ob diskriminierungsfreie Kriterien für die *Positionierung innerhalb des Lohnbandes* angewandt werden.».
- 67 ZGP Vernehmlassung Änderung GIG, 23. Februar 2016 (Fn. 45), S. 5: «Wenn hingegen beides, die Funktionsbewertung und die Lohnbandeinstufung, diskriminierungsfrei sind, kann es in einer Unternehmung überhaupt keine systematische Lohnungleichheit geben. Allenfalls können *Anwendungsfehler* bei der Lohnfestsetzung gemacht werden, die jedoch einfach zu korrigieren sind.».
- 68 Beispiel: ULICH (Fn. 3), S. 601: «Es müssen Regeln formuliert werden, in denen festgelegt wird,
– wer beurteilt, wer was kann,
– ob und in welchen Abständen eine Überprüfung der Beurteilung vorgenommen werden soll,
– in welchen Abständen Höherstufungen möglich sein sollen.»;
Formulierung übernommen von KARLHEINZ SONNTAG / EKKEHART FRIELING / RALF STEGMAIER, Lehrbuch Arbeitspsychologie, 3.A. Bern 2012, S. 534; vgl. dazu SCHETTGEN (Fn. 3), S. 151, Gestaltungsempfehlung (4 [Fn. 47]).
- 69 Und mit einer Formulierung aus Deutschland (1996) über die Unentbehrlichkeit von multi-criteria Methoden schliesst sich der Kreis zum Ausgangspunkt unserer Studie: SCHETTGEN (Fn. 3), S. 146: «[...] in der BRD sind die Gesetzesnormen viel zu allgemein und vage formuliert, und damit letztlich unbrauchbar (vgl. ... Gleichwertigkeitsgrundsatz im § 611 a BGB; ...). Die Arbeitsgerichte sehen sich angesichts dieser gesetzlichen Situation nicht in der Lage, Entscheidungen zu treffen, und delegieren die Aufgabe der Bestimmung des Arbeitswertes an den Markt zurück, der wiederum genau die sozialen Ungerechtigkeiten schafft, derenthalben die Gerichte angerufen werden.»; *übrigens*: Der kritisierte § 611 a BGB verwendet genau wie der schweizerische Art. 3 Abs. 1 GIG die Formulierung «benachteiligen» und «Benachteiligung».