KATHOLIEKE
UNIVERSITEIT
LEUVEN

Methodologische noot: decompositietechnieken voor loonkloofanalyse

Sem Vandekerckhove, Caroline Vermandere, Tom Vandenbrande en Luc Sels

In opdracht van POD Wetenschapsbeleid, programma 'Samenleving en toekomst'

Januari 2009

Hoger instituut
voor de arbeid

Inhoudstafel

1 Inleiding 3
2 Operationalisering van de loonkloof 5
2.1 Definitie 5
2.2 Welk 'gemiddeld' loon? 6
2.3 Het voordeel van een logaritmische transformatie 9
2.4 De loonverhouding en haar complicaties 11
2.5 Ongelijk of onrechtvaardig? 12
3 Decompositie van de loonkloof 13
3.1 Dummymodel 13
3.2 Oaxaca-Blinder 14
3.3 Algemeen model 16
3.4 Discriminatie of verschillende returns? 17
3.5 Variaties op de decompositie 20
3.5.1 Het bepalen van de genderneutrale returns 20
3.5.2 De methode van Oaxaca-Blinder 20
3.5.3 De methodes van Reimers en Cotton 21
3.5.4 De methode van Neumark 21
3.6 Empirische vergelijking van de methodes 23
3.6.1 Steekproefomschrijving 23
3.6.2 Bepalen van de betrouwbaarheidsintervallen via bootstrapping 24
3.6.3 Selectie van de variabelen 25
3.6.4 Loonmodellen voor mannen en vrouwen 26
3.6.5 Vergelijkingen van de effecten van assets en returns per methode 26
3.6.6 Relatieve bijdrage van de verklarende variabelen 28
4 Conclusie 32
5 Bibliografie 34
6 Appendix 36
6.1 Univariate data 36
6.2 Loonmodellen voor mannen en vrouwen 36
6.3 Decompositie 37

1 INLEIDING

Het WAGE GAP onderzoeksproject neemt een oud zeer opnieuw onder de loep: de loonsongelijkheid tussen mannen en vrouwen. Vrouwen hebben gemiddeld een lager loon, maar verdienen wellicht beter. Dit uitgangspunt leidde tot de Dag van Gelijk Loon (ACV-KAV), de Equal Pay Day (ABVV) en een jaarlijkse monitor door het Instituut voor de gelijkheid van Mannen en Vrouwen. In het maatschappelijk debat is het van het grootste belang dat het niet bij het vaststellen van een loonachterstand van vrouwen blijft. We moeten deze loonkloof ook verklaren en evalueren. Met een genuanceerder beeld van de ruwe cijfers kunnen we de focus van het debat en de beleidsingrepen gerichter maken.

Deze noot licht het methodologische luik van het onderzoek toe, waarbij we zullen focussen op het gebruik van decompositietechnieken in loonkloofonderzoek. Een correcte meting en duidelijke operationalisering van de technieken is essentieel voor het in kaart brengen en verklaren van de loonkloof. Deze blijkt namelijk gevoelig voor misinterpretaties. Daarom gaan we uitgebreid in op de definiëring van de loonkloof en schenken de nodige aandacht aan de discussie omtrent discriminatie. Dit aspect komt daarna aan bod als component in het decompositiemodel ter verklaring van de loonkloof, dat we zullen veralgemenen om zowel positieve als negatieve discriminatie te onderscheiden.

Verschillende methodes om dit te doen worden afgeleid, en we maken een vergelijking tussen de uitkomsten. De technieken worden geïllustreerd aan de hand van een standaard loonmodel en toegepast op de Vacature Salarisenquête 2006.

2 OPERATIONALISERING VAN DE LOONKLOOF

Het operationaliseren van de loonkloof veronderstelt twee belangrijke stappen. Ten eerste moet gedefinieerd worden wat we onder een loonkloof begrijpen, ten tweede moet aangegeven worden hoe de verschillende elementen uit de definitie bepaald kunnen worden.

2.1 Definitie

De loonkloof (gender wage gap) is het verschil tussen de lonen van vrouwen en mannen, uitgedrukt als een fractie van het loon van mannen. In het geval het loon van vrouwen lager is dan dat van mannen zal dit getal tussen 0 en 1 liggen.[1] Volgende formule geeft de loonkloof (G) algebraïsch weer, waarbij W_m het loon (wage) van mannen is en W_f dat van vrouwen:

$$G = \frac{\overline{W}_m - \overline{W}_f}{\overline{W}_m} = 1 - \frac{\overline{W}_f}{\overline{W}_m}$$

[1] Om de tekst niet nodeloos te verzwaren gaan we uit van lagere lonen en negatieve discriminatie van vrouwen, en hogere lonen en positieve discriminatie bij mannen. Dit hoeft echter niet noodzakelijk altijd het geval te zijn.

De uitkomst van deze formule leert ons hoeveel percent *lager* het loon van vrouwen is ten opzichte van dat van mannen. We kiezen voor mannen als referentiegroep in de noemer, volgens de conventie in beleidsgericht onderzoek (Plantenga & Remery, 2006). Een alternatieve specificatie met vrouwen als referentiegroep is echter eveneens mogelijk en geeft een antwoord op de vraag hoeveel percent hoger of lager het loon van mannen is ten opzichte van het loon van vrouwen (Oaxaca, 1973).

2.2 Welk 'gemiddeld' loon?

De loonkloof is een verhouding tussen 'het loon' van vrouwen en mannen. Om tot een operationalisering van dit fenomeen te komen moet nog een tweede beslissing worden genomen, net name die over de uitdrukking van 'het loon' dat in de eerder gegeven functie wordt genomen.

Om 'het loon' van mannen of vrouwen te synthetiseren kan beroep gedaan worden op verschillende statistische maten. In Figuur 1 is de loonverdeling van alle mannen en vrouwen grafisch geïllustreerd, en werd aangegeven hoe deze informatie in een aantal kengetallen kan gereduceerd worden (zie appendix 6.1 voor een vergelijkende tabel).

De meest eenvoudige maat voor centraliteit is het **rekenkundig gemiddelde**,[2] waarbij de totale loonmassa in een groep of populatie verdeeld wordt over het aantal leden van die groep. In de Vacature Salarisenquête van 2006 (zie paragraaf 3.6.1) bedraagt het gemiddelde loon van mannen 3310.57 EUR en dat van vrouwen 2291.93 EUR. De loonkloof is dan 30.77%.

[2] Het rekenkundig gemiddelde van de lonen *W* in een populatie met omvang *n*: $A(W) = \frac{\sum W}{n}$

Voor statistische analyses met het rekenkundig gemiddeld als kernwaarde is het van belang dat de waarden volgens een normaalverdeling (of 'een hoed van Napoleon') gespreid liggen rond deze waarde. In het specifieke voorbeeld van looncurves is dat niet het geval: we vinden wel extreme uitlopers aan de rechterkant van de curve, maar omdat lonen nooit negatieve waarden kunnen aannemen en door een wettelijke ondergrens (minimumloon of bestaansuitkering), ontbreken deze extreme uitlopers aan de linkerkant. Een looncurve zal dus doorgaans een scheve verdeling weerspiegelen.

Figuur 1 laat zien dat de loondistributie in de gehele populatie inderdaad rechtsscheef is en verstoord wordt door enkele uitzonderlijk hoge waarden, die we *outliers* noemen. Dit zijn exuberant veelverdienende respondenten die niet gecompenseerd worden door sterk negatieve lonen aan de onderkant, en ervoor zorgen dat het rekenkundige gemiddelde hoger gaat liggen dan wanneer de looncurve normaal verdeeld zou zijn.

In het geval van scheve verdelingen krijgen we een genuanceerder beeld door te kijken naar de **mediaanwaarde**.[3] Dit is het loon waar de ene helft van de populatie boven zit en de andere helft onder. In tegenstelling tot het rekenkundig gemiddelde is deze maat ongevoelig voor extreme waarden: hoeveel men boven

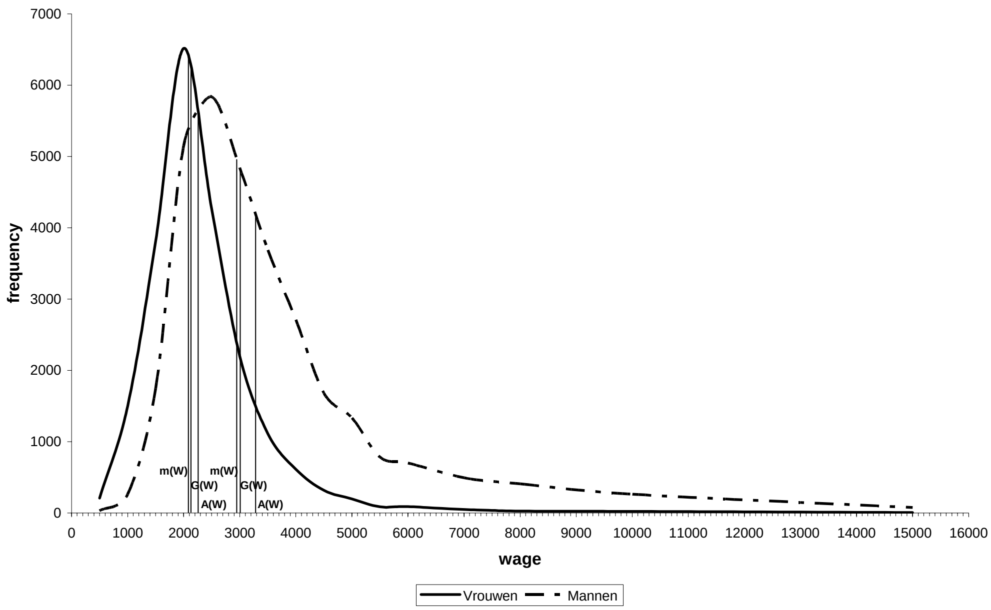
---

[3] De mediaan van de lonen *W* in een populatie met omvang n, af te lezen op de cumulatieve functie K: $\tilde{W} = \left\{ W_i \,\middle|\, K(W_i) = \frac{n+1}{2} \right\}$

Waar de mediaan een centrummaat is, zijn ook andere positiematen mogelijk. Ook op basis van andere quantielen kunnen we een loonkloof bepalen. Het eerste kwartiel zal bijvoorbeeld de loonkloof geven onder de kleine inkomens, terwijl het 95[ste] percentiel de loonkloof bij de hoogste lonen weergeeft.

of onder het mediaan loon zit is van geen belang. In ons geval bedraagt de mediaan bij mannen 2970 EUR en die bij vrouwen 2100 EUR. Baseren we ons op de mediaan, dan verkleint de loonkloof tot 29.29%.

Figuur 1. Brutoloondistributie bij mannen en vrouwen (maandloon bij bedienden in 2006)



Bron: Vacature salarisenquête, release 14 (2006)

Voor de meer complexe statistische analyses gebruiken we het **meetkundig gemiddelde**,[4] dat eveneens *minder* gevoelig is voor outliers dan het rekenkundige en toch enkele nuttige metrische eigenschappen behoudt voor het gebruik in een regressiemodel. In paragraaf 2.3 onderbouwen we verder deze stap. Het meetkundig gemiddelde levert inderdaad resultaten op die tussen de eerdere

---

[4] Het meetkundig gemiddelde van de lonen W in een populatie met omvang n: $G(W) = \sqrt[n]{\prod W} = e^{\frac{\sum \ln(W)}{n}} = e^{A(\ln(W))}$.

berekeningen in liggen. Het gemiddelde loon voor mannen is dan 3020.17 EUR en voor vrouwen 2117.64 EUR. OP basis van het meetkundig gemiddelde bedraagt de loonkloof ten slotte 29.88%.

Samengevat kunnen we stellen dat er – helaas – niet één gemiddelde is dat ons de beste inschatting geeft van het typische loon. Afhankelijk van de onderzoeksvraag gebruiken we de maat die rekening houdt met elk loon (rekenkundig gemiddelde), die aansluit bij het grootste aantal individuen (mediaan) of het midden houdt tussen beide en praktischer is in het gebruik (meetkundig gemiddelde).

2.3 Het voordeel van een logaritmische transformatie

In de analyses van de loonkloof wordt steeds gewerkt met het *natuurlijk logaritme* van de lonen. Door deze transformatie wordt de rechts scheefverdeelde looncurve (zie hoger) omgezet in een curve die beter beantwoordt aan de normaalverdeling.[5] In Figuur 2 zien we wat met de loonverdeling gebeurt door het uitdrukken van de lonen in het linker luik als logaritmes in het rechterluik. Aan de onderkant van het loongebouw, waar heel wat lonen rond het minimumloon hangen, wordt de variatie in lonen wat uitgerekt. Aan de bovenkant van het loongebouw, met zoals aangegeven een aantal extreme waarden van veelverdieners, wordt de variatie in lonen afgevlakt.[6] Het resultaat

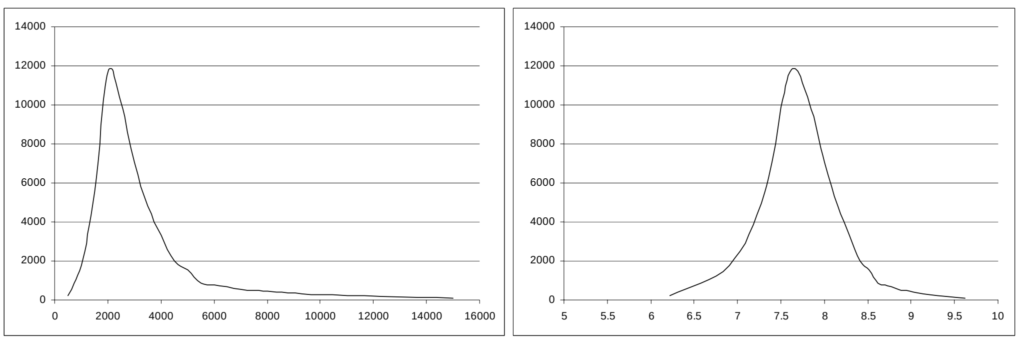
---

[5] Dit is een manier om ervoor te zorgen dat de residuen van de schatting normaal verdeeld zijn, een assumptie bij lineaire regressie.

[6] De reden hiervoor is dat een logaritmische transformatie een absolute schaal omzet in een relatieve schaal. Onderaan de loonverdeling zal een verschil tussen 500 en 1000 EUR een verdubbeling van het loon betekenen, terwijl eenzelfde verschil van 500 EUR tussen lonen van

van deze transformatie is dat het rekenkundig gemiddelde van deze logaritmes niets anders is dan het logaritme van het meetkundig gemiddelde van de oorspronkelijke loonverdeling. In de vorige paragraaf zagen we wat dit impliceert: per definitie onderschatten we het rekenkundige gemiddelde, maar dat is wenselijk omdat we zo een gemiddelde bekomen dat dichter bij de mediaan aansluit.[7] Door de logaritmische transformatie krijgen we een model van de vorm $\ln(W) = X\beta$. Bij een toename met één eenheid in de verklarende variabele, vergroot het loon één maal met factor e, het grondtal van een natuurlijk logaritme (~2.72).

Figuur 2. Logaritmische transformatie van de brutoloondistributie (maandloon bij bedienden in 2006)



Bron: Vacature salarisenquête, release 14 (2006)

6500 en 7000 EUR in relatieve termen een verhoging met 7.69% betreft. Drukken we de lonen uit als logaritmes, dan wordt het verschil tussen de eerste twee lonen, het natuurlijk logaritme van 2, terwijl dit tussen de laatste twee het natuurlijk logaritme van 1.077 wordt. Verschillen tussen logaritmes drukken dus de relatieve grootte uit. Hoe hoger de lonen, des te kleiner de toename in logaritmes zal zijn voor een verhoging van het loon met eenzelfde bedrag.

[7] De mediaan, het meetkundig en het rekenkundig gemiddelde zijn enkel gelijk wanneer alle lonen gelijk zijn.

2.4 De loonverhouding en haar complicaties

We hebben de loonkloof omschreven als het verschil in loon, uitgedrukt als percentage. Men kan deze maat te allen tijde omvormen als een verhouding van lonen. In het geval we de lonen van vrouwen t.o.v. die van mannen plaatsen is de loonverhouding of *gender ratio* eenvoudigweg 1 *min* de loonkloof. Als vrouwen 30.77% minder verdienen dan mannen, dan verdienen ze inderdaad 69.23% $(1-0.3077)$ vàn het loon van mannen, of 0.69 keer zoveel (zie appendix, paragraaf 6.1). Voor de decompositieanalyse zal de loonverhouding echter omgekeerd worden, in navolging van de betreffende literatuur op dat vlak (Oaxaca, 1973; Neumark 1988). We krijgen dus het loon van mannen ten opzichte van dat van vrouwen:

$$GR = \frac{\overline{W}_f}{\overline{W}_m} \Rightarrow \frac{1}{GR} = \frac{\overline{W}_m}{\overline{W}_f}$$
$$G = 1 - GR$$

In decompositieanalyses is de gebruikte maat voor centraliteit het *meetkundige gemiddelde,* dat we aanduiden met G(W) (cf. paragraaf 2.2). Algebraïsch krijgen we dan voor elk geslacht *g*: $G(W_g) = \sqrt[n_g]{\prod(W_g)} = e^{\frac{\sum_{i=1}^{n_g} \ln(W_{gi})}{n_g}} = e^{A(\ln(W_g))}$. Het gevolg is dat we de logloonverhouding kunnen schrijven als het verschil van de rekenkundige gemiddeldes van de gelogaritmeerde lonen. Onderstaande afleiding bewijst dit:

$$\ln\left(\frac{1}{GR}\right) = \ln\left(\frac{G(W_m)}{G(W_f)}\right) = \ln(G(W_m)) - \ln(G(W_f))$$

$$\ln\left(\frac{1}{GR}\right) = \ln\left(e^{\frac{\sum_{i=1}^{n_m}\ln(W_i)}{n_m}}\right) - \ln\left(e^{\frac{\sum_{i=1}^{n_f}\ln(W_i)}{n_f}}\right)$$

$$\ln\left(\frac{1}{GR}\right) = A(\ln(W_m)) - A(\ln(W_f))$$

$$\frac{1}{GR} = e^{A(\ln(W_m)) - A(\ln(W_f))}$$

2.5 Ongelijk of onrechtvaardig?

De loonkloof en genderdiscriminatie worden vaak in één adem genoemd. Beide gaan samen, zonder dat ze hetzelfde betekenen. We moeten namelijk ongelijkheid (inequality) onderscheiden van onrechtvaardigheid (inequity). Een kernvraag in loonkloofonderzoek is inderdaad het bepalen van een rechtvaardige loonsverhouding $(W_m/W_f)_0$ tussen mannen en vrouwen en na te gaan in welke mate de reële situatie hiervan afwijkt. Oaxaca (1973) drukt deze vraag algebraïsch uit als de discriminatiecoëfficiënt *D*. Deze geeft weer hoeveel groter de reële loonverhouding is in vergelijking met de rechtvaardige loonsverhouding. We komen later nog terug op de moeilijkheden bij het bepalen van die "rechtvaardige" loonsverhouding. Volgende formule geeft de discriminatiecoëfficiënt weer:

Comment: Later op terugkomen

$$D = \frac{W_m/W_f - (W_m/W_f)_0}{(W_m/W_f)_0}$$

3 DECOMPOSITIE VAN DE LOONKLOOF

In deze paragraaf wordt ingegaan op de methode om de determinanten van het bestaan van loonverschillen tussen mannen en vrouwen in kaart te brengen. Eerst worden alternatieve methodes gepresenteerd, in een tweede fase illustreren we de kenmerken van deze methodes aan de hand van data over de Belgische loonkloof, in een derde fase spreken we op basis van deze analyses een waardeoordeel uit over de verschillende analysetechnieken.

3.1 Dummymodel

Het eenvoudigste model om de determinanten van het verschil in loon tussen mannen en vrouwen op te meten, is een regressiemodel waarbij één verklarende variabele *geslacht* is.

$$\ln(W_i) = X_0\beta_0 + \sum X_{ik}\beta_k + SEX_i\beta_{sex} + \varepsilon_i$$

De verklaring van het (log)loon wordt gegeven door het intercept ofwel de regressiecoëfficiënt β_0 (als $X_0 = 1$) en door een reeks verklarende variabelen (X_k), die we in het vervolg *assets* zullen noemen.

We zijn uiteraard vooral geïnteresseerd in welke mate het effect van de variabele *geslacht* (β_{sex}) door de andere variabelen verklaard wordt. Daartoe stellen we een volledig model op waaruit we stapsgewijs variabelen wegnemen. De toename in β_{sex} kan uitgedrukt worden als aandeel van het effect van *geslacht* in een model zonder andere variabelen.

Dit model heeft echter een belangrijke methodologische tekortkoming: het is mogelijk dat er een differentiële return bestaat voor bepaalde assets naargelang het geslacht. Zo zou opleidingsniveau in grote mate het loon van mannen kunnen bepalen, maar bij vrouwen nauwelijks effect hebben. Om dit euvel te verhelpen nemen meer geavanceerde methodes aparte parameters voor het bepalen van het loon van mannen en vrouwen op .

3.2 Oaxaca-Blinder

Het meest invloedrijke model voor het in kaart brengen van de loonkloof is het *Oaxaca-Blinder decompositiemodel* (Oaxaca, 1973 & Blinder, 1973). Een decompositiemodel heeft tot doel de effecten van de assets te scheiden van de effecten die te wijten zijn aan directe negatieve of positieve discriminatie, voor zover we er in slagen die te schatten.

Bij aanvang worden de determinanten van het loon voor mannen en vrouwen bepaald via twee eenvoudige regressies: $\ln(W_{gi}) = X_{gi}\beta_g + \varepsilon_{gi}$ met *g* als geslacht. Deze formule betekent dat het populatiemodel en haar schatting voor mannen (*g*=*m*) gegeven worden door $\ln(W_{mi}) = \sum X_{mi}\beta_m + \varepsilon_{gi}$ en $A(\ln(W_m)) = A(X_m)\hat{\beta}_m$, en voor vrouwen (*g*=*f*) door $\ln(W_{fi}) = \sum X_{fi}\beta_f + \varepsilon_{fi}$ en $A(\ln(W_f)) = A(X_f)\hat{\beta}_f$.[8] Voortaan geven we het rekenkundig gemiddelde van de *verklarende* variabelen aan met een platte streep ($A(X) = \overline{X}$).

[8] We schrijven de formules in die zin dat ze ook als matrices kunnen geïnterpreteerd worden.

Vanuit deze specificatie komen we tot het Oaxaca-Blinder model:

$$A(\ln(W_m)) - A(\ln(W_f)) = \overline{X}_m \hat{\beta}_m - \overline{X}_f \hat{\beta}_f$$
$$A(\ln(W_m)) - A(\ln(W_f)) = \overline{X}_m \hat{\beta}_m - \overline{X}_f \hat{\beta}_f + \left(\overline{X}_f \hat{\beta}_m - \overline{X}_f \hat{\beta}_m\right)$$
$$A(\ln(W_m)) - A(\ln(W_f)) = \overline{X}_m \hat{\beta}_m - \overline{X}_f \hat{\beta}_m + \overline{X}_f \hat{\beta}_m - \overline{X}_f \hat{\beta}_f$$
$$A(\ln(W_m)) - A(\ln(W_f)) = \left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_m + \overline{X}_f\left(\hat{\beta}_m - \hat{\beta}_f\right) \equiv E + U$$

Om verwarring over de mogelijkheden van deze analysetechniek te vermijden en gezien de gevoeligheid van het label 'discriminatie', is het belangrijk de juiste interpretatie van de componenten van het model te omschrijven.

Het *verklaarde deel* ($E = \left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_m$) is niets anders dan het verschil in loon tussen mannen en vrouwen, wanneer vrouwen dezelfde return ($\hat{\beta}_m$) zouden hebben voor elke asset als mannen. Deze term (E) zorgt voor een vergroting van de loonkloof als mannen gemiddeld hoger scoren op de verklarende variabelen. Het is de bonus voor mannen ten opzichte van vrouwen op basis van hun kwaliteiten.

Het *onverklaarde deel* ($U = \overline{X}_f\left(\hat{\beta}_m - \hat{\beta}_f\right)$) is het verschil tussen het loon dat vrouwen zouden krijgen gegeven hun profiel bij dezelfde returns ($\hat{\beta}_m$) als mannen en wat ze werkelijk verdienen. Als deze term (U) toeneemt, betekent dit dat mannen om een andere reden dan de verklarende variabelen uit het model (bijvoorbeeld o.w.v. genderdiscriminatie) meer verdienen dan vrouwen. U staat dus voor de extra beloning voor mannen, gegeven een bepaald niveau van assets, en vormt samen met het verschil in kwaliteiten de decompositie van de loonkloof.

3.3 Algemeen model

Het Oaxaca-Blinder model gaat ervan uit dat het verschil tussen de return voor mannen en die voor vrouwen de discriminatie is. Dit betekent dat in een arbeidsmarkt zonder discriminatie, de coëfficiënten voor vrouwen (β_f) gelijk zouden zijn aan die voor mannen (β_m). Het is denkbaar dat daar niet genoeg economische middelen voor bestaan. Met andere woorden: de returns van de mannen gaan ten koste van die voor de vrouwen. We moeten daarom op zoek naar een discriminatieloze coëfficiënt (β_*), voor het effect dat een toename (of afname) van een asset heeft op het loon. We kunnen de geobserveerde situatie hiermee vergelijken en zo een onderscheid maken tussen bevoordeling van mannen (positieve discriminatie) en benadeling van vrouwen (negatieve discriminatie).

Voortbouwend op het Oaxaca-Blinder model kunnen we een algemeen model opstellen (Oaxaca & Ransom, 1994, Silber & Weber, 1999), dat de (log)loonverhouding onderbrengt in een deel assets, een deel positieve returns voor mannen en een deel negatieve returns voor vrouwen. Eerst geven we dit model in abstracto, vervolgens werken we vier meetbare modellen uit om de loonkloof te ontleden.

Onderstaande afleiding voegt enkele termen toe aan het verschil van de gemiddelde (log)lonen van mannen en vrouwen. Deze termen sommeren tot nul, en veranderen dus niets aan de loonkloof. Door te herschikken komen we tot de drie onderscheiden componenten: E, U^+ en U^-.

$$\ln\left(\frac{G(W_m)}{G(W_f)}\right) = \ln(G(W_m)) - \ln(G(W_f)) = A(\ln(W_m)) - A(\ln(W_f))$$

$$A(\ln(W_m)) - A(\ln(W_f)) = \overline{X}_m \hat{\beta}_m - \overline{X}_f \hat{\beta}_f$$

$$A(\ln(W_m)) - A(\ln(W_f)) = \overline{X}_m \hat{\beta}_m - \overline{X}_f \hat{\beta}_f + \left(\overline{X}_m \hat{\beta}_* - \overline{X}_m \hat{\beta}_* + \overline{X}_f \hat{\beta}_* - \overline{X}_f \hat{\beta}_*\right)$$

$$A(\ln(W_m)) - A(\ln(W_f)) = \overline{X}_m \hat{\beta}_* - \overline{X}_f \hat{\beta}_* + \overline{X}_m \hat{\beta}_m - \overline{X}_m \hat{\beta}_* + \overline{X}_f \hat{\beta}_* - \overline{X}_f \hat{\beta}_f$$

$$A(\ln(W_m)) - A(\ln(W_f)) = \left(\overline{X}_m - \overline{X}_f\right)\hat{\beta}_* + \overline{X}_m\left(\hat{\beta}_m - \hat{\beta}_*\right) + \overline{X}_f\left(\hat{\beta}_* - \hat{\beta}_f\right) \equiv E + U^+ + U^-$$

De term E is gelijkaardig aan deze in het Oaxaca-Blinder model, alleen gaat het hier om de verschillen in assets, met de return die geldt in het geval er geen discriminatie is.

We merken dat het onverklaarde deel U een positieve en een negatieve term heeft, wat staat voor de positieve en negatieve returns die dit model capteert. Let wel: het is heel goed mogelijk dat *bepaalde* assets een hogere return zullen kennen bij vrouwen dan bij mannen. In die gevallen zal U^+ *negatief* zijn en dus de loonkloof dichten. We verwachten echter dat het effect vaker omgekeerd zal zijn.

De term U^- kan analoog worden geïnterpreteerd: ze omvat het deel van het loon dat hoger is door een lagere return on assets voor vrouwen dan het geval zou zijn in een arbeidsmarkt zonder differentiële returns volgens geslacht.

Uiteraard stuiten we hier op het probleem dat deze genderneutrale parameter een theoretisch concept is. Om β_* te bepalen zijn dan ook verschillende varianten mogelijk.

3.4 Discriminatie of verschillende returns?

In de literatuur heerst er een zekere onduidelijkheid over verklaarde en onverklaarde delen, waarbij de indruk gewekt wordt dat het onverklaarde deel tot nul herleid kan worden als het model geperfectioneerd wordt. We moeten dit nuanceren.

Kijken we naar de aanvankelijke regressies voor mannen en vrouwen (zie verder, paragraaf 3.6.4), dan is het duidelijk dat deze slechts ten dele de variantie in de lonen kunnen verklaren. De coëfficiënten die we gebruiken in de decompositie zijn echter afgestemd op de gemiddelden in de populatie, zodat de schatting van het gemiddelde loon voor mannen en vrouwen correct is. Bijgevolg kunnen we de logloonkloof aan de hand van deze verklarende variabelen en parameters toch volledig indelen.

Het niet-dekken van de totale variantie geeft aan dat de schattingen die we maken op basis van het loonmodel imperfect zijn. Op basis van de verklarende variabelen voorspellen we dat een individu op een bepaald niveau boven of onder het gemiddelde loon zit. Een goed model kan hier vrij betrouwbare uitspraken over doen. Toch zullen er altijd gevallen waarvoor we het loon niet juist konden inschatten. We spreken dan van *onverklaarde variantie.* Dit is niet te verwarren met de *onverklaarde componenten* U^+ en U^-. Dit zijn constructen die we maken met de geschatte parameters uit de regressies. 'Onverklaard' betekent in dit geval dat er nog verschillen zijn in de coëfficiënten van mannen en vrouwen, die we niet kunnen thuisbrengen: in een situatie zonder discriminatie verwachten we immers dat een bepaalde assets eenzelfde productiviteitswinst oplevert en eenzelfde productiviteit gelijk vergoed wordt, wanneer tenminste met de juiste controlevariabelen rekening gehouden wordt. Om een uitweg te bieden aan de spraakverwarring, is het beter te verwijzen naar U^+ en U^- als de componenten die de effecten weergeven van ongelijke *returns* (soms ook rewards of prices genoemd).

Dezelfde redenering gaat op voor E, of de *verklaarde* component. Met 'verklaard' bedoelen we hier dat we objectieve verschillen hebben kunnen vaststellen op bepaalde variabelen tussen de groepen die we vergeleken. Deze verschillen kunnen op zichzelf de uitdrukking zijn van genderdiscriminatie. Ook hier gebruiken we schattingen van de parameters, die slechts een deel van de variantie kunnen verklaren, en dus nog steeds een deel onverklaard laten. Om opnieuw verwarring te vermijden kan verwezen worden naar E als het effect van ongelijke *assets* (soms ook *characteristics* of *human capital* genoemd).

Helaas is hiermee de kous niet af. De kritiek dat het discriminatie-effect afhangt van de verklaringskracht van de basismodellen, moeten we ter harte nemen. Het is mogelijk dat het verfijnen van het basismodel door het toevoegen van nieuwe variabelen ervoor zorgt dat de verschillen in coëfficiënten afnemen. Deze werden dan veroorzaakt door het feit dat de nieuwe variabele waarvoor gecontroleerd werd, de ene dan wel de andere partij een 'verklaard' voordeel oplevert. Vooral het intercept zal hieraan onderhevig zijn. We kunnen dit een groter effect van assets noemen wanneer er sprake is van productiviteitsverschillen (Weichselbaumer & Winter-Ebmer, 2008), maar het is ook mogelijk dat het gaat om factoren die irrelevant zijn voor de productiviteit maar toch het loon bepalen. In die laatste vorm ondermijnt het verfijnen van het basismodel juist de gebruikelijke interpretatie van de componenten. Al van in den beginne wees Oaxaca (1973) erop dat een oneindig complex model (met uiteindelijk gelijke vectoren $\hat{\beta}_m$ en $\hat{\beta}_f$) volledig in de verklaarde component ondergebracht zou moeten worden. We besluiten daarom dat enig voorbehoud steeds aangewezen is. Het label 'discriminatie' moet eerder als conceptueel dan als feitelijk beschouwd worden, maar is verantwoord wanneer gebruik gemaakt wordt van een model dat rekening houdt met de relevante assets. Kritiek op de geschatte coëfficiënten is altijd mogelijk, maar de beperkingen van inferentiële statistiek zijn niet eigen aan genderstudies alleen.

3.5 Variaties op de decompositie

3.5.1 Het bepalen van de genderneutrale returns

Het algemeen model moet, zoals hoger geargumenteerd, aangevuld worden met een specificatie van β_*. Hiervoor zijn meerdere suggesties gedaan, maar een consensus over een te verkiezen specificatie is er niet (Ma & Chu Ng, 2008). In het algemeen schrijven we β_* als een combinatie van de coëfficiëntenmatrices β_m & β_f. Oaxaca en Ransom (1994) gaven hiervoor de volgende formule:

$$\beta_* = \Omega\beta_m + (I - \Omega)\beta_f$$

Waarbij Ω een wegingsmatrix is met dimensies *k*k* (k = aantal variabelen) en I een identiteitmatrix met dezelfde dimensies.

3.5.2 De methode van Oaxaca-Blinder

Het eerste model dat we in deze algemene vorm kunnen passen is het Oaxaca-Blinder model zélf. Wanneer mannen de referentiecategorie vormen, is $\hat{\beta}_* = \hat{\beta}_m$ en dus $\Omega = I$. U$^+$ valt in dit geval weg ($\hat{\beta}_m - \hat{\beta}_* = 0$), wat betekent dat het om een negatieve discriminatie-model gaat (enkel U$^-$). We kunnen echter net zo goed uitgaan van $\hat{\beta}_* = \hat{\beta}_f$ waarbij $\Omega = 0I$. U$^+$ zal dan extra returns voor mannen weergeven t.o.v. vrouwen, gegeven dat ze dezelfde assets hebben (positieve discriminatie), terwijl E staat voor de loonsvoorsprong van mannen op basis van hun assets in vergelijking met die van vrouwen. Het Oaxaca-Blinder model heeft het grote voordeel dat het reeds zeer vaak gebruikt is, en zich dus gemakkelijker leent voor vergelijkingen met eerder onderzoek.

3.5.3 De methodes van Reimers en Cotton

De volgende modellen onderscheiden zowel U^+ en U^-. Aangezien discriminatie opgedeeld wordt in een positieve en een negatieve component, zal de positieve en de negatieve component verschillen van het verklaarde deel in respectievelijk de positieve en de negatieve variant van het Oaxaca-Blinder model. Reimers (1983) neemt als verwachte waarde voor de genderneutrale returncoëfficiënten het gemiddelde van de geobserveerde voor mannen en voor vrouwen $\hat{\beta}_* = \frac{(\hat{\beta}_m + \hat{\beta}_f)}{2}$. De genderneutrale coëfficiënt β_* is dus de bissectrice van β_m & β_f en de wegingsmatrix wordt gegeven door $\Omega = 0.5I$.

Bij Cotton (1988) is de neutrale coëfficiënt het gewogen gemiddelde volgens de aandelen van mannen en vrouwen in de steekproef of populatie. We kunnen dit schrijven als $\hat{\beta}_* = f_m\hat{\beta}_m + f_f\hat{\beta}_f$, met f_m als de relatieve frequentie voor mannen en $f_f = 1 - f_m$ als deze voor vrouwen. De wegingsmatrix is dan ook $\Omega = f_m I$.

3.5.4 De methode van Neumark

Neumark (1988) en in navolging Oaxaca & Ransom (1994) gebruiken de schatting in de gepoolde steekproef ($\hat{\beta}_* = \hat{\beta}_{m+f}$). De meest praktische methode om deze β_* te bepalen bestaat erin naast de basismodellen voor mannen en vrouwen, ook een regressie uit te voeren voor beide groepen samen en hieruit de genderneutrale coëfficiënten te halen. [9] Ook dit kan geschreven worden als een gewichtenmatrix, namelijk $\Omega = (X'X)^{-1}(X'_m X_m)$.[10] Deze schatting kent een groter gewicht toe aan de variabele die een groter deel van de variantie in de afhankelijke variabele verklaart.

---

[9] Hieronder bewijzen we dat deze oplossing met gewichtenmatrices overeenkomt met de kleinste kwadratenschatting voor de β-vector:

$$
\begin{aligned}
&\Omega\beta_m + (I-\Omega)\beta_f = \beta \\
&\Omega(\beta_m - \beta_f) + \beta_f = X'X^{-1}X'Y \\
&X'X^{-1}X'_m X_m(\beta_m - \beta_f) + \beta_f = X'X^{-1}X'Y \\
&X'_m X_m\left(X'_m X_m{}^{-1}X'_m Y_m - X'_f X_f{}^{-1}X'_f Y_f\right) + X'_f X_f{}^{-1}X'_f Y_f X'X = X'Y \\
&X'_m X_m X'_m X_m{}^{-1}X'_m Y_m - X'_m X_m X'_f X_f{}^{-1}X'_f Y_f + X'_f X_f{}^{-1}X'_f Y_f X'X = X'Y \\
&X'_m Y_m - X'_m X_m X'_f X_f{}^{-1}X'_f Y_f + X'_f X_f{}^{-1}X'_f Y_f X'X = X'Y \\
&X_m Y_m - X'_m X_m X'_f X_f{}^{-1}X'_f Y_f + X'_f X_f{}^{-1}X'_f Y_f X'X - X'Y = 0 \\
&X_m Y_m - X'_m X_m X'_f X_f{}^{-1}X'_f Y_f + X'_f X_f{}^{-1}X'_f Y_f X'_m X_m + X'_f X_f{}^{-1}X'_f Y_f X'_f X_f - X'Y = 0 \\
&X_m Y_m + X'_f Y_f - X'Y = 0
\end{aligned}
$$

[10] We kunnen deze formule opslitsen als $\Omega = (X'_m X_m + X'_f X_f)^{-1}(X'_m X_m)$, zodat duidelijk blijkt dat het om een gewicht gaat. De opsplitsbaarheid van $X'X$ is eenvoudig te begrijpen aangezien de cellen de som geven van alle *n* producten van de waarden op elk paar variabelen. Die som is gelijk aan de som van dezelfde oefening in meerdere subgroepen.

Geheel volgens de definitie is het gewicht voor β_f het verschil van de identiteitsmatrix en de gegeven wegingsmatrix voor β_m. Immers:

$$
\begin{aligned}
\Omega_f &= (X'X)^{-1}(X'_f X_f) = (X'X)^{-1}(X'X - X'_m X_m) \\
\Omega_f &= (X'X)^{-1}(X'X) - (X'X)^{-1}(X'_m X_m) = I - \Omega
\end{aligned}
$$

3.6 Empirische vergelijking van de methodes

Elk van de gepresenteerde methodes is bruikbaar om te bepalen welke determinanten aan de basis van de gender loonkloof liggen. In deze paragraaf wegen we de pro's en contra's van deze methodes af. In meta-analyses werd reeds nagegaan af de gekozen methode een effect heeft op het inschatten van de componenten (Weichselbaumer & Winter-Ebmer, 2008), hier passen we de verschillende methodes toe op één dataset (cf. Silber & Weber, 1999), de Vacature Salarisenquête.

3.6.1 Steekproefomschrijving

De selectie van de data is van belang voor de generaliseerbaarheid van de bevindingen. De loonkloof verschilt volgens land en regio, volgens sector, doorheen de tijd en hangt ook af van het segment van de arbeidsmarkt. Ook voor andere mogelijke afbakeningen van de data geldt dat ze mede bepalen voor welke groep de bevindingen zullen opgaan.

De decompositieanalyse die we hieronder beschrijven is gebaseerd op de data van de Vacature-salarisenquête. Dit is een online vragenlijst die gekoppeld is aan een website van een jobmagazine. De gebruikte release bevat de gegevens die werden verkregen in 2006. Tabel 1 toont de samenstelling van de data. Van belang is de vaststelling dat de steekproefomvang aanzienlijk is (62 284 respondenten), doch het aandeel arbeiders beperkt. Dit is duidelijk een selectiebias omwille van het medium en de doelgroep van het jobmagazine. Om die reden verkiezen we om arbeiders buiten de analyse te laten en gebruik te maken van een weegfactor die de data representatief maakt voor de Belgische beroepsbevolking onder bedienden en ambtenaren.

Tabel 1. Samenstelling steekproef Vacature 2006 salarisenquête

	Vacature 2006	
aantal respondenten	62 284	100%
mannen	36 605	59%
vrouwen	25 679	41%
arbeiders	6 667	11%
bedienden	48 027	77%
ambtenaar	7 585	12%
voltijds	56 059	90%
deeltijds	6 174	10%

Noot: ongewogen cijfers

3.6.2 Bepalen van de betrouwbaarheidsintervallen via bootstrapping

De decompositieanalyse combineert verschillende schattingen van parameters. Als gevolg daarvan zijn de betrouwbaarheidsintervallen rond de componenten en de items niet rechtstreeks te bepalen. Daarom werken we met zogenaamde *exacte* methodes. De techniek die wij gebruiken is de *bootstrap* of Monte Carlo-methode (cf. Silber & Weber, 1999). Om een betrouwbaarheidsinterval rond de geschatte parameter te verkrijgen herhalen we de analyse voor subsamples van de data. Uit het totale bestand met 55 617 respondenten werden 100 subsample gemaakt waarbij aan volgens een binomiale kansenverdeling gewichten werden gegeven aan de respondenten. Op elke subsample werd een decompositieanalyse gedaan. De gemiddelden van de items in deze 100 analyses kunnen gezien worden als het gemiddelde van een steekproevenverdeling, en de standaardafwijking is gelijk aan de standaardfout rond de geschatte parameter. We kunnen elke component en elk item zien als een parameter, en er via de bootstrappingmethode een betrouwbaarheidsinterval rond construeren.

3.6.3 Selectie van de variabelen

3.6.3.1 Afhankelijke variabele

Als afhankelijke variabele gebruiken we het logaritme van het nettomaandloon. We brengen hier in herinnering dat de loonkloof het verschil is in de rekenkundige gemiddeldes van de logaritmes van de lonen van mannen en vrouwen ofwel het logaritme van de omgekeerde loonratio $\ln\left(\frac{G(W_m)}{G(W_f)}\right)$ (zie paragraaf 2.4).

3.6.3.2 Onafhankelijke variabelen

Deze decompositieanalyse dient in de eerste plaats als illustratie bij de techniek. Hoewel deze techniek vrij oud is en internationaal gangbaar, werd deze in de Belgische context nauwelijks toegepast. Door de nadruk op de praktische toepassing was de keuze voor de onafhankelijke variabelen van secundair belang. De variabelen die opgenomen werden in de analyse zijn bekende determinanten of controlefactoren (cf. Vermandere e.a., 2008). Het detailniveau is voor deze illustratie beperkt tot enkele ruime categorieën. Merk tot slot op dat de variabele *geslacht* niet in de lijst voorkomt: we werken immers steeds met aparte vergelijkingen voor mannen en vrouwen.

Volgende variabelen werden gebruikt:

- Leeftijd
- Regio: de regio van tewerkstelling (Vlaanderen, Wallonië of Brussel)
- Sector: primaire, secundaire, tertiaire en quartaire sector
- Arbeidsduur: het de facto aantal gepresteerde uren
- Opleidingsniveau in zes klassen: (1) lager onderwijs; (2) lager; (3) secundair onderwijs; (4) hoger secundair onderwijs; (5) HOKT; (6) HOLT; (7) universiteit

- Functieniveau, of hiërarchische positie in zes klassen: (1) Administratief ondersteunend personeel; (2) Uitvoerend personeel; (3) Professionele medewerker (bv. Stafdienst, expertrol); (4) Middle management; (5) Senior management; (6) Algemene directie/topmanagement
- Bedrijfsgrootte, uitgedrukt in aantal werknemers: (1) 1-9; (2) 10-49; (3) 50-199; (4) 200-499; (5) 500-999; (6) 1000+

3.6.4 Loonmodellen voor mannen en vrouwen

Voor de regressieanalyse hebben we afzonderlijke loonmodellen nodig voor mannen en vrouwen, en een derde voor beide samen (in functie van de methode van Neumark). De resultaten van deze regressies zijn te vinden in appendix 0. Ondanks het beperkte arsenaal aan variabelen kunnen we de loonkloof toch zeer goed inschatten (R^2 = 53% (mannen) en R^2= 52.1% (vrouwen)). Arbeidsduur, leeftijd en opleidingsniveau blijken de sterkste determinanten van het loon, met telkens een positief verband. Wie werkt in Brussel verdient meer dan wie in Vlaanderen en Wallonië werkt. De quartaire sector beloont het minst, de secundaire sector kent de hoogste lonen. Ook hiërarchisch niveau en ondernemingsgrootte hebben een positieve impact op het loon.

3.6.5 Vergelijkingen van de effecten van assets en returns per methode

In navolging van Oaxaca & Ransom (1994) en Silber & Weber (1999) gaan we na of de verschillende methodes voor het bepalen van de componenten significant verschillende resultaten opleveren. Tabel 2 geeft de resultaten van deze oefening, waarbij de positieve en negatieve discriminatie samengeteld worden tot U om de vergelijking te kunnen maken met de methodes van Oaxaca, en ook als fractie van U worden uitgedrukt. De verschillen tussen de methodes situeren zich zowel op het vlak van de indeling van de loonkloof (die uiteraard steeds dezelfde is in de verschillende modellen) en de grootte van het betrouwbaarheidsinterval rond deze schatting. Die betrouwbaarheid hangt af van de gewichtenmatrix die gebruikt werd om een genderneutrale coëfficiënt te schatten.

De belangrijkste component in de verklaring van de loonkloof zijn de assets (E). Bij Neumark staan die in voor zo'n 70% van de loonkloof, bij de andere methodes ligt de schatting in de buurt van 63%. Neumark & de Oaxaca-variant met vrouwen als referentiecategorie vormen respectievelijk de hoogste en de laagste schatting. Voor de 'totale discriminatie' U geldt het omgekeerde: Neumark geeft de laagste schatting, de Oaxaca-decompositie met vrouwen als referentiecategorie geeft de grootste. Deze U komt overeen met de discriminatiecoëfficiënt (zie paragraaf 2.5) wanneer we ze in absolute termen uitdrukken en *e* tot deze macht verheffen. Zo komen we te weten met hoeveel percent de loonverhouding toeneemt t.g.v. ongelijke returns. Bij de techniek van Cotton is dit bijvoorbeeld een toename met 13.84%.

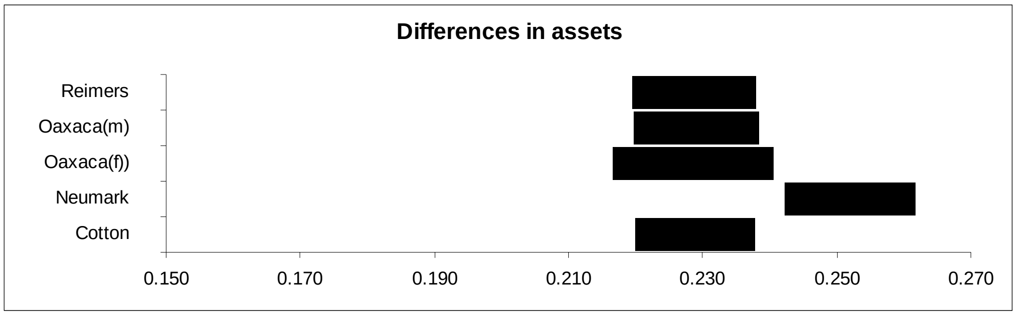
De samenstelling van U varieert eveneens: de Oaxaca-varianten terzijde gelaten is de hoogste inschatting van de negatieve discriminatie (U^-) 58.80% bij Neumark, en de laagste 49.91% bij Reimers. Voor de positieve discriminatie (U^+) geldt het omgekeerde.

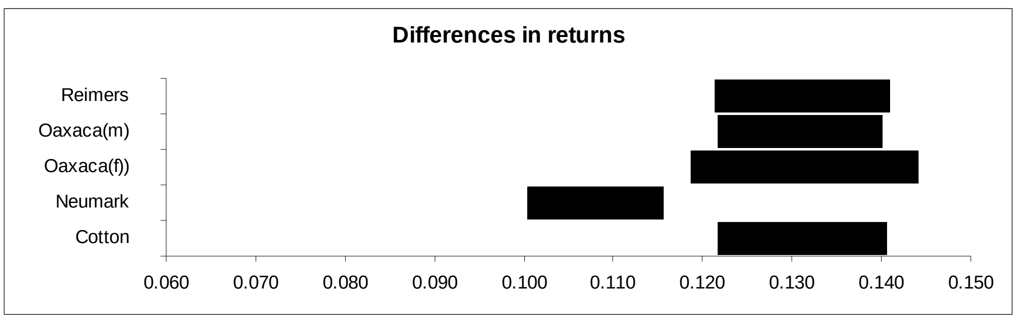
Via bootstrapping construeerden we betrouwbaarheidsintervallen rond de gemiddelde waarden achter deze fracties (zie appendix, paragraaf 6.3). Figuur 3 illustreert de onzekerheidsmarges rond de absolute bijdrage in de loonkloof van de assets en de returns bij een toegelaten fout (α) van 5% met Bonferroni correctie. We merken dat er slechts één techniek significant verschil van de andere: de methode van Neumark. Deze vaststelling komt overeen met de bevindingen van Silber & Weber (1999).

Tabel 2. Vergelijking decompositiemethodes naar verdeling van de loonkloof

	E	U	U^-	U^+
Cotton	63.57%	36.43%	58.00%	42.00%
Neumark	70.00%	30.00%	58.80%	41.20%
Oaxaca(f)	63.49%	36.51%	0.00%	100.00%
Oaxaca(m)	63.63%	36.37%	100.00%	0.00%
Reimers	63.56%	36.44%	49.91%	50.09%

Figuur 3. Betrouwbaarheidsintervallen rond de schattingen van de componenten





3.6.6 Relatieve bijdrage van de verklarende variabelen

In de vorige paragraaf zijn de methodes vergeleken en bleek de som van de items per component niet significant te verschillen. Onderstaande tabel geeft de effecten van elk item weer, uitgedrukt als percentage van de totale loonkloof.[11]

[11] De uitdrukking van de effecten in percentages is mogelijk omdat de logloonverhouding voldoende afwijkt van 0. Zoniet worden de percentages oneindig groot. Merk ook op dat deze percentages sommeren tot 100, ongeacht het aantal verklarende factoren.

We kijken hier in de eerste plaats naar de betekenis van de items voor elke component. Daarbij kan het nuttig zijn de geschatte parameters te bekijken in de appendix, paragraaf 6.2. De items hebben een positief aandeel in de loonkloof als mannen hogere gemiddelde scores vertonen op de assets (E) of wanneer de genderneutrale coëfficiënt op de returns (U^- en U^+) kleiner zijn dan de coëfficiënten bij mannen en groter dan deze bij vrouwen. Dit is het meest waarschijnlijke scenario, maar ook het omgekeerde is mogelijk: negatieve bijdrages tot de loonkloof duiden op lagere scores voor mannen op de assets (E) en genderneutrale regressiecoëfficiënten die groter zijn dan deze bij mannen (U^+) en kleiner dan deze bij vrouwen (U^-). We nemen de analyse van Cotton als vertrekpunt.

De effecten van assets (E) zijn voornamelijk toe te schrijven aan functieniveau (21%), arbeidsduur (18%) en leeftijd (12%). Actieve mannen met een job hebben gemiddeld een hogere hiërarchische positie, presteren meer uren (bv. voltijds) en zijn ouder. Ook bedrijfsgrootte, opleidingsniveau en sector (telkens 4%) hebben een zeker belang. Mannen werken gemiddeld vaker in grotere bedrijven, waar lonen hoger zijn, en hebben een iets hoger opleidingsniveau. Door het belang van deze parameter in het verklaren van het loon, bedraagt het effect toch zo'n 4% van de totale loonkloof. Mocht het opleidingsniveau voor mannen en vrouwen gelijk zijn, dan nam de logloonverhouding m.a.w. met 4% af.[12]

---

[12] We kunnen deze tabel niet lezen als een soort regressiemodel voor de loonkloof om twee redenen. Een afname van de logloonkloof met P% ($P = 1 - \frac{LL'}{LL}$), betekent immers niet dat de loonkloof met eenzelfde percentage krimpt. Stel P = 4% en LL = 0.36, dan *stijgt* de loonverhouding (vrouwen over mannen, $\frac{G(W_f)}{G(W_m)}$ of $\frac{1}{e^{LL}}$) met 1.45% ($1 - e^{P \cdot LL}$). De loonkloof $\frac{G(W_m) - G(W_f)}{G(W_m)}$

De effecten van ongelijke returns (U) zijn opgesplitst in negatieve effecten voor vrouwen en positieve effecten voor mannen. We moeten hierbij meteen opmerken dat een groot deel van de loonkloof te wijten is aan onverklaarde startposities, die we terugvinden in het effect van de constante (44% U^- en 32% U^+). Opmerkelijk hierbij is dat de methode van Neumark dit item bijna volledig onder positieve discriminatie plaatst. Dit wil zeggen dat de genderneutrale coëfficiënt veel dichter bij de waarde voor vrouwen ligt. Bij Reimers en Cotton zijn dit duidelijk de ongewogen en gewogen gemiddelde waarden. In elk geval is het intercept voor mannen hoger dan dat voor vrouwen. Afgezien van de intercepten zijn de significante effecten hier arbeidsduur (-59% U^- en -48% U^+) en functieniveau (20% U^- en 19% U^+). Mannen die opklimmen tot een hoger niveau zullen meer verdienen dan vrouwen die een dergelijke promotie maken. Het effect van arbeidsduur is echter eerder onverwacht: vrouwen zullen per uur dat ze gaan werken, een grotere stijging zien in hun loon dan mannen. Waarschijnlijk schuilt hierachter het effect dat vrouwen met een voltijdse baan een ander type job hebben met een betere verloning. Tot slot is er nog een significant effect van bedrijfsgrootte als we uitgaan van de methode van Neumark. Werken in grotere bedrijven vergroot de loonkloof door het feit dat vrouwen hiervoor een lagere bonus op het loon krijgen, in vergelijking met het genderneutrale effect. Voor de

---

wijzigt ten slotte met -1.012 %-punt ($\frac{1-e^{P \cdot LL}}{e^{LL}}$). Dit is te omslachtig: de indeling van de loonkloof leent zich vooral tot een vergelijking van de relatieve grootte van de effecten.

Verder moet men rekening houden met het effect van gelijke gemiddelde waarden via de returns: de omvang van deze effecten wordt namelijk voor U^- ook bepaald door de gemiddelde score van vrouwen, en voor U^+ door de gemiddelde score van mannen. Worden deze scores gelijk, dan zal de totale component U (U^- + U^+) niet constant blijven. Afhankelijk van de waarde die mannen en vrouwen dan hebben en de coëfficiënten voor deze variabele, kan de loonkloof stijgen of dalen.

interpretatie van de omvang van de effecten verwijzen we naar de bespreking van de relatieve bijdrage van assets, in de vorige alinea.

Tabel 3. Decompositie van de loonkloof

		Constant		Leeftijd		Opl. niv.		Vlaanderen		Wallonië		Primair		Secundair		Ter
Cotton	E	0%	n.s.	12%	*	4%	*	0%	n.s.	0%	n.s.	0%	n.s.	4%	*	0%
	Un	44%	*	9%	*	-5%	*	4%	*	0%	n.s.	0%	n.s.	0%	n.s.	2%
	Up	32%	*	7%	*	-4%	*	3%	*	0%	n.s.	0%	n.s.	1%	n.s.	2%
Neumark	E	0%	n.s.	13%	*	4%	*	0%	n.s.	0%	n.s.	0%	n.s.	4%	*	0%
	Un	5%	n.s.	15%	*	-7%	*	4%	*	0%	n.s.	0%	n.s.	1%	*	3%
	Up	70%	*	0%	n.s.	-2%	n.s.	3%	*	0%	n.s.	0%	n.s.	-1%	*	0%
Oaxaca(f)	E	0%	n.s.	11%	*	4%	*	0%	n.s.	0%	n.s.	0%	n.s.	3%	*	0%
	Un															
	Up	76%	*	17%	*	-9%	*	7%	*	0%	n.s.	0%	n.s.	1%	n.s.	4%
Oaxaca(m)	E	0%	n.s.	13%	*	4%	*	0%	n.s.	0%	n.s.	0%	n.s.	4%	*	1%
	Un	76%	*	16%	*	-9%	*	7%	*	0%	n.s.	0%	n.s.	1%	n.s.	4%
	Up															
Reimers	E	0%	n.s.	12%	*	4%	*	0%	n.s.	0%	n.s.	0%	n.s.	4%	*	0%
	Un	38%	*	8%	*	-4%	*	3%	*	0%	n.s.	0%	n.s.	0%	n.s.	2%
	Up	38%	*	9%	*	-5%	*	3%	*	0%	n.s.	0%	n.s.	1%	n.s.	2%

n.s. = niet significant; * = significant volgens het 95% betrouwbaarheidsinterval

Bron: Vacature salarisenquête, release 14 (2006)

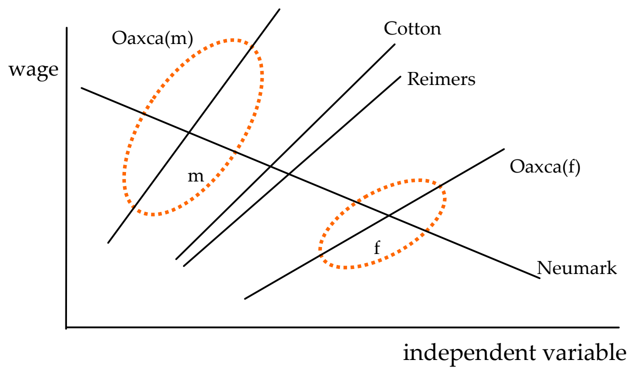
Noot: In de gebootstrapte data is de logloonkloof ln(1/GR) gelijk aan 0.35997. In de totale steekproef was dit 0.35501. Het verschil is te wijten aan het aantal iteraties en aan het feit dat we het gemiddelde logloon bepalen aan de hand van de loonmodellen, waarbij missing het resultaat licht beïnvloeden.

4 CONCLUSIE

De hamvraag blijft welke techniek uiteindelijk de 'beste' is. Hier kunnen we geen exact antwoord op geven. Bij eliminatie komen we toch tot een techniek die we verkiezen te gebruiken. Beschouwen we de varianten op de Oaxaca-decompositie, dan is het duidelijk dat helemaal geen rekening gehouden wordt met herverdelingseffecten in de hypothetische situatie dat een genderneutrale looncurve geldt. Dezelfde returns garanderen voor vrouwen als deze die in het geval van discriminatie golden voor mannen, betekent dat er opeens meer middelen beschikbaar zijn om de loonkost te dekken. Dit is economisch niet realistisch.

Een tweede techniek die we minder waarderen is de methode van Neumark. Het nadeel hier is dat interactie-effecten met gender uitgesloten zijn. Bij een genderneutrale looncurve zal dit per definitie het geval zijn, maar deze analyse kunnen we niet maken. In Figuur 4 is de kwetsbaarheid van de Neumark-decompositie grafisch voorgesteld. De ovalen stellen puntenwolken voor, de lijnen regressiecoëfficiënten voor één verklarende variabele. Door de loonfuncties van mannen en vrouwen te poolen, maar geen interactie-effecten op te nemen, wat de facto gebeurt wanneer we de Neumark gewichtenmatrix gebruiken, *kan* een regressiecoëfficiënt bekomen worden die volledig onwaarschijnlijk is. Ook in minder uitgesproken gevallen zullen interactie-effecten de Neumark-schatting vertekenen.

Figuur 4. Grafische voorstelling van de verschillen tussen de decompositietechnieken



Blijven nog over: de techniek van Reimers en die van Cotton. Finaal geven we de voorkeur aan de techniek van Cotton met de motivatie die Oaxaca & Ransom (1994) al aanhaalden: het is intuïtief logischer dat de genderneutrale coëfficiënten dichter liggen bij de grootste groep op de arbeidsmarkt (mannen). Zeker wanneer grote populaties beschouwd worden lijkt dit de beste optie. In subgroepen, vb. een sector, is het echter mogelijk dat er meer vrouwen actief zijn dan mannen. Voor die subgroepen kan de populatieverdeling naar geslacht vertekende gewichten toekennen aan de groepen. Het hangt af van de inschatting van de selectie van de te onderzoeken populatie om in deze gevallen voor de Reimers, dan wel de Cotton-decomposite te verkiezen. Onze analyses wezen uit dat de impact van die keuze de resultaten niet fundamenteel beïnvloedt.

5 BIBLIOGRAFIE

Cotton J. (1988) 'On the decomposition of wage differentials', The Review of Economics and Statistics, 70 (2), p. 236-243

Ma Y & Chu Ng Y (2008) 'Bootstrapping statistical inferences of decomposition methods for gender earnings differentials', Applied Economics, 40, p. 1583-1593

Neumark D. (1988) 'Employers' disciminatory behavior and the estimation of wage discrimination', Journal of Human Resources, 23 (3), p. 279-295

Oaxaca R.L. & Ransom M.R. (1994) 'On discrimination and the decompostition of wage diferentials', Journal of Econometrics, 61, p. 5-21

Oaxaca R.L. (1973) 'Male-Female Wage Differentials in Urban Labor Markets', International Economic Review, 14 (3), p. 693-709

Plantenga J. & Remery C. (2006) 'The gender pay gap. Origins and policy responses. A comparative review of thirty European countries.', Draft report. Utrecht School of Economics

Reimers C.W. (1983) 'Labor market discrimination against hispanic and Black Men', Review of Economics and Statistics, 65 (4), p. 570-579

Silber J. & Weber M. (1999) 'Labor market discrimination: Are there significant differences between the various decomposition procedures?', Applied Economics, 31 (), p. 359-365

Vermandere C., Vandekerckhove S., Vandenbrande T. & Sels L. (2008) 'Naar een verklaring van de loonkloof', working paper HIVA

Weichselbaumer D. & Winter-Ebmer R. (2005) 'A meta-analysis of the international gender wage gap', Journal of Economic Surveys, 19 (3), p. 479-511

6 APPENDIX

6.1 Univariate data

Tabel 4. Het loon in de Belgische beroepsbevolking onder bedienden en ambtenaren: vergelijking van de maten van centraliteit

	Mannen	Vrouwen	Loonkloof	Loonverhouding
Rekenkundig gemiddelde	3310.57	2291.93	30.77%	69.23%
Mediaan	2970.00	2100.00	29.29%	70.71%
Meetkundig gemiddelde	3020.17	2117.64	29.88%	70.12%

Bron: Vacature salarisenquête, release 14 (2006)

6.2 Loonmodellen voor mannen en vrouwen

Tabel 5. Loonmodellen voor mannen, vrouwen en gepoolde data, met gemiddelde waarden en standaardafwijking voor mannen en vrouwen.

	Totaal		Mannen		Vrouwen		Mannen		Vrouwen	
	b	p	b	p	b	p	$\overline{X}_m$	$\sqrt{s^2}$	$\overline{X}_f$	$\sqrt{s^2}$
Intercept	**5.933**	0.000	**6.186**	0.000	**5.912**	0.000	**1.00**	0.00	**1.00**	0.00
Leeftijd	**0.014**	0.000	**0.015**	0.000	**0.013**	0.000	**39.90**	10.24	**36.76**	9.86
Opl. Niv.	**0.090**	0.000	**0.088**	0.000	**0.097**	0.000	**3.83**	1.28	**3.68**	1.17
Regio*										
Vlaanderen	**-0.052**	0.000	**-0.034**	0.000	**-0.072**	0.000	**0.62**	0.48	**0.63**	0.48
Wallonië	**-0.105**	0.000	**-0.102**	0.000	**-0.113**	0.000	**0.15**	0.36	**0.15**	0.35
Sector**										
Primair	**0.073**	0.014	**0.049**	0.225	**0.103**	0.011	**0.00**	0.04	**0.00**	0.05
Secundair	**0.155**	0.000	**0.147**	0.000	**0.130**	0.000	**0.28**	0.45	**0.19**	0.39
Tertiair	**0.074**	0.000	**0.078**	0.000	**0.049**	0.000	**0.49**	0.50	**0.47**	0.50
Arbeidsduur	**0.014**	0.000	**0.008**	0.000	**0.018**	0.000	**43.84**	8.95	**38.31**	9.13
Functieniv	**0.101**	0.000	**0.112**	0.000	**0.060**	0.000	**3.13**	1.22	**2.30**	1.14

Ond. Grootte	**0.042**	0.000	**0.041**	0.000	**0.031**	0.000	**3.92**	1.75	**3.50**	1.78
R^2	0.572		0.530		0.521					

Bron: Vacature salarisenquête, release 14 (2006)
* Referentiecategorie: Brussel
** Referentiecategorie: quartaire sector

6.3 Decompositie

Comment: Updaten: bonferroni!

Tabel 6. Betrouwbaarheidsintervallen rond de schattingen van de componenten

	Comp.	Effect	Relative size	95% C.I.
Cotton	E	0.22866	63.52%	(0.1646-0.2927)
	U	0.13131	36.48%	(0.0634-0.1991)
	Un	0.07665	58.37%	(0.0347-0.1184)
	Up	0.05466	41.63%	(0.0175-0.0917)
Neumark	E	0.25202	70.01%	(0.1834-0.3206)
	U	0.10795	29.99%	(0.0519-0.1639)
	Un	0.06353	58.85%	(0.0297-0.0973)
	Up	0.04443	41.15%	(0.0209-0.0678)
Oaxaca(f))	E	0.22946	63.75%	(0.1441-0.3148)
	U	0.13050	36.25%	(0.0429-0.218)
	Un	-	-	-
	Up	0.13050	100.00%	(0.0429-0.218)
Oaxaca(m)	E	0.22809	63.36%	(0.1613-0.2948)
	U	0.13188	36.64%	(0.061-0.2027)
	Un	0.13188	100.00%	(0.061-0.2027)
	Up	-	-	-
Reimers	E	0.22878	63.55%	(0.1632-0.2942)
	U	0.13119	36.45%	(0.0621-0.2002)
	Un	0.06594	50.26%	(0.0305-0.1013)
	Up	0.06525	49.74%	(0.0214-0.109)