



## Bruger ikke-vestlige efterkommerkvinder deres uddannelser? Baggrundsnotat 2: Løngevinst ved uddannelse

### Indledning

I denne analyse undersøges, om ikke-vestlige efterkommerkvinder får lige så meget ud af at tage en erhvervskompetencegivende uddannelse, som kvinder med dansk baggrund gør.

Først beregnes den såkaldte løngevinst ved uddannelse for de to grupper. Altså hvor stor økonomisk gevinst, der opnås ved at tage en erhvervskompetencegivende uddannelse i forhold til, hvis man ingen uddannelse har. Dernæst sammenholdes de to gruppers løngevinst. Til sidst undersøges gevinsten for de kvinder, som er i beskæftigelse.

Der indledes med data og metode samt beskrivelse af grupperne.

### Data og metode

Analysen er baseret på en række registre fra Danmarks Statistik med oplysninger på individniveau.

Datasættet består af ikke-vestlige efterkommerkvinder og kvinder med dansk oprindelse mellem 24-40-år, som boede i Danmark den 1. januar 2017 og 2018 og ikke er under uddannelse. Den nedre aldersgrænse er valgt, fordi der fokuseres på personer med fuldført erhvervskompetencegivende uddannelse. Den øvre aldersgrænse er valgt for at gøre grupperne mere sammenlignelige, da der kun findes ganske få efterkommerkvinder over 40 år med en erhvervskompetencegivende uddannelse.

Løngevinsten ved uddannelse beregnes i analysen ud fra kvindernes erhvervsindkomst i 2017. Personer som ikke er i beskæftigelse vil dermed indgå i analysen med en erhvervsindkomst på 0 kr. På den måde inkluderes beskæftigelsens betydning i beregningen af gruppernes løngevinst (se også Arendt Nielsen 2018 eller Kraka 2015). Der benyttes en såkaldt winsorizing, så personer med en negativ erhvervsindkomst på mere end 2 mio. kr. samt en positiv erhvervsindkomst på mere end 2 mio. kr. sættes til hhv.  $\pm 2$  mio. kr. i analysen.

Udlændinge- og  
Integrationsministeriet

Analysekontoret  
Slotsholmsgade 10  
1216 København K

Tel. 6198 4000  
Mail [uim@uim.dk](mailto:uim@uim.dk)  
Web [www.uim.dk](http://www.uim.dk)

CVR-nr. 36977191

For at beregne løngevinsten og komme så tæt på den egentlige effekt af at tage en erhvervskompetencegivende uddannelse som muligt, er valgt et analysedesign, hvor løngevinsten modelleres ved udelukkende at se på forskelle i erhvervsindkomst mellem søskende – i vores tilfælde forskellene mellem søstre (se Produktivitetskommissionen 2013 eller Kraka 2015).

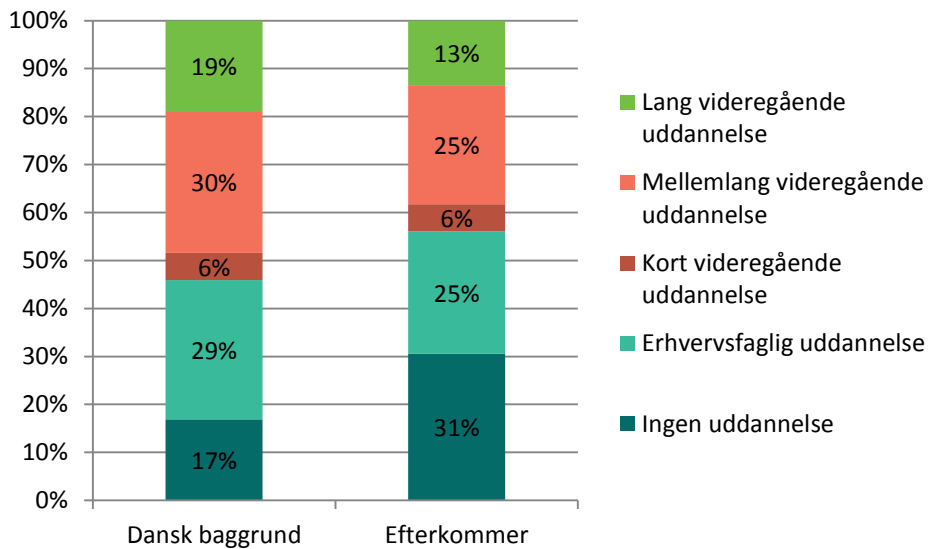
I denne analyse kompliceres beregningen ved, at ikke-vestlige efterkommerkvinder og kvinder med dansk baggrund har en meget forskellig alderssammensætning, da alder har stor betydning for lønnens størrelse. For at tage højde for disse forskelle afgrænses datasættet som nævnt til de 24-40-årige, ligesom der benyttes forskellige statistiske greb til at tage højde for alderssammensætningen (se nærmere beskrivelse af grupperne i bilag 1 og den statistiske model i bilag 2).

### **Beskrivelse af grupperne**

Der indgår i alt 370.527 kvinder med dansk baggrund i datagrundlaget og 10.212 efterkommerkvinder. Figur 1 viser gruppernes fordeling på uddannelse – for yderligere beskrivende statistik se bilag 1.

Figuren viser, at kvinderne med dansk baggrund er bedre uddannede end efterkommerkvinderne. Blandt kvinderne med dansk baggrund har cirka hver sjette ingen erhvervskompetencegivende uddannelse, mens det for efterkommerkvinderne er knap hver tredje og altså næsten dobbelt så mange. Forskellen i uddannelsesniveau har stor betydning for kvindernes indkomst, men når løngevinsten ved uddannelse estimeres, ser man på, hvor stor indkomsten er for de to grupper i forhold til, hvis de ingen uddannelse havde. Dermed kan *løngevinsten* for en efterkommer med en universitetsuddannelse eksempelvis godt være højere end gevinsten for en kvinde med dansk baggrund med en universitetsuddannelse, selvom kvinden med dansk baggrunds *indkomst* er højere.

**Figur 1: Højest fuldførte erhvervskompetencegivende uddannelse fordelt på kvinder med dansk baggrund og ikke-vestlige efterkommerkvinder**

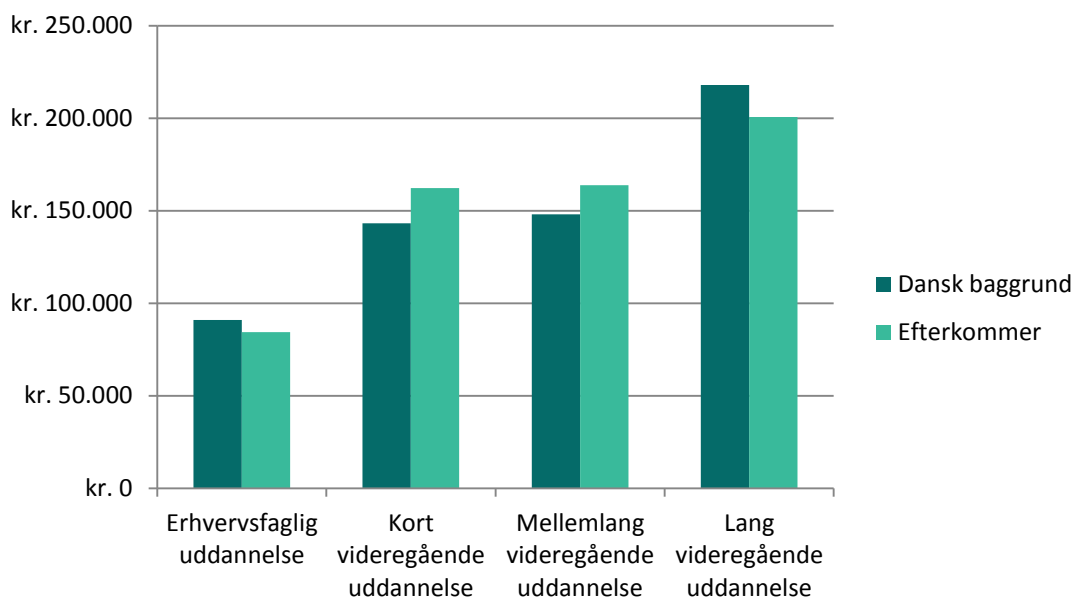


Anm.: Ingen uddannelse inkluderer i analysen personer, som ikke har nogen registreret værdi for deres højest fuldførte uddannelse. Mellemlang videregående inkluderer universitetsbacheloror.

### Gruppernes løngevinst ved uddannelse

Figur 2 viser, at løngevinsten er ganske stor for både kvinder med dansk baggrund og efterkommerkvinder, uanset hvilket uddannelsesniveau det drejer sig om. Der er en tendens til, at gevinsten stiger med uddannelseslængde, og gevinsten er omtrent lige stort for begge grupper.

**Figur 2: Løngevinsten ved forskellige uddannelsesniveauer fordelt på kvinder med dansk baggrund og ikke-vestlige efterkommerkvinder**

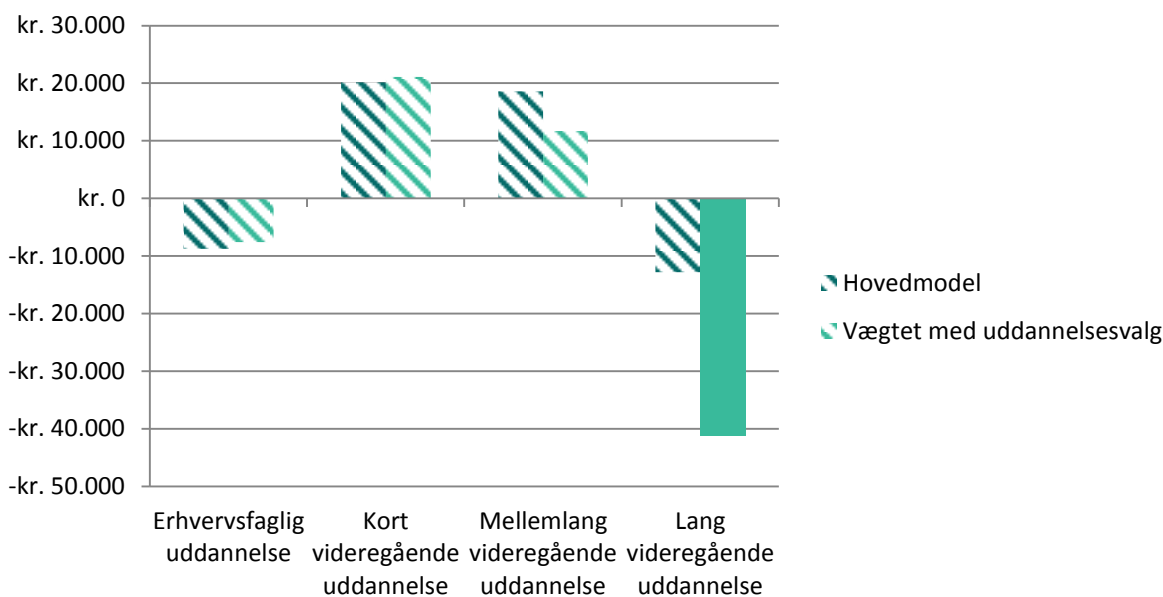


Anm.: Figuren viser den estimerede løngevinst ved forskellige uddannelsesniveauer. Estimaterne udtrykker indkomstforskellen mellem personer med forskellige uddannelsesniveauer og sammenlignelige personer uden erhvervskompetencegivende uddannelse. Personer uden erhvervskompetencegivende uddannelse inkluderer i analysen personer, som ikke har nogen registreret værdi for deres højst fuldførte uddannelse. Mellemlang videregående inkluderer universitetsbachelor. Løngevinsten i denne figur er estimeret på baggrund af en regressionsmodel for kvinder med dansk baggrund og en model for efterkommerkvinder. En beskrivelse af metoden findes i bilag 2. Antallet af personer som indgår i beregningen af gevinsten reduceres til cirka en tredjedel af datasættet, idet der udelukkende ses på forskelle inden for søskendepar – se også bilag 1.

Ud fra figur 2 ses det, at der er ganske små forskelle mellem grupperne. For at teste, om forskellene er reelle eller kan være udtryk for tilfældighed, estimeres også én samlet regressionsmodel, som kaldes hovedmodel i figur 3 (se beskrivelse af metoden i bilag 2).

Resultaterne fra hovedmodellen i figur 3 viser, at forskellene på efterkommerkvinders og kvinder med dansk baggrunds løngevinst på de forskellige uddannelsesniveauer ikke er statistisk sikre. Der er derfor generelt ikke indikationer på forskelle mellem efterkommerkvinders og kvinder med dansk baggrunds løngevinst af at tage en uddannelse.

**Figur 3: Løngevinst ved uddannelse for ikke-vestlige efterkommerkvinder i forhold til kvinder med dansk baggrund**



Anm.: Figuren viser ikke-vestlige efterkommerkvinders estimerede løngevinst ved uddannelse relativt til kvinderne med dansk baggrund. Søjlernes sribede udfyldning angiver, at forskellene i efterkommer og kvinder med dansk baggrunds løngevinst *ikke* med statistisk sikkerhed er forskellig fra 0 ved et 5%-signifikansniveau. Personer uden erhvervskompetencegivende uddannelse inkluderer i analysen personer, som ikke har nogen registreret værdi for deres højst fuldførte uddannelse. Mellemlang videregående inkluderer universitetsbacheloror. Løngevinsterne i figuren er estimeret på baggrund af regressionsmodeller med både kvinder med dansk baggrund og efterkommerkvinder. En beskrivelse af metoden findes i bilag 2.

Uddannelsesniveaue dækker over forskellige uddannelsesvalg. Der kan fx være forskel på, hvilke erhvervsfaglige uddannelser kvinderne med dansk baggrund og efterkommerkvinderne vælger, eller hvilke videregående uddannelser, der vælges. Disse uddannelsesvalg har betydning for løngevinsten, og de estimerede gevinster kan således dække over forskellige uddannelsesvalg inden for niveauerne. For at undersøge om de relativt lige løngevinst, som fremgår af figur 2, skyldes forskellige uddannelsesvalg, estimeres også en statistisk model, hvor personer vægtes ud fra deres uddannelsesvalg (metoden beskrives i bilag 2). Resultaterne af modellen, hvor der vægtes på uddannelsesvalg fremgår også af figur 3.

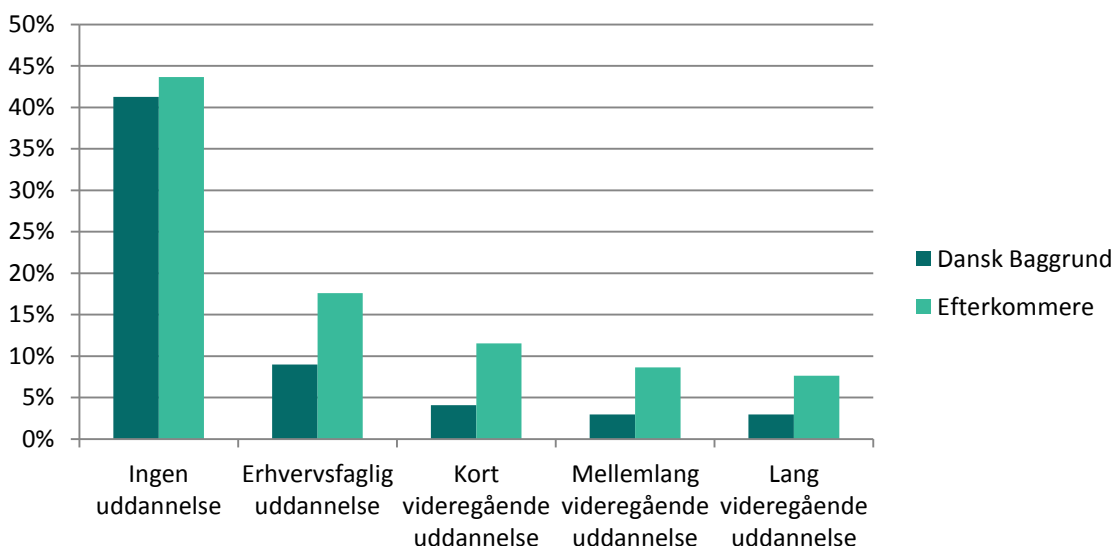
Resultaterne fra den vægtede model viser, at selv når der tages højde for forskellige uddannelsesvalg, er der ikke statistisk sikre forskelle på gruppernes løngevinst ved uddannelse på nær på de lange videregående uddannelser. Når der tages højde for uddannelsessammensætningen bliver efterkommerkvinders gevinster af lange videregående uddannelser væsentligt mindre ift. kvinderne med dansk baggrunds. Dette kunne indikere, at de tilnærmelsesvist lige store gevinster af lange videregående uddannelser i figur 2 skyldes, at efterkommerkvinderne vælger uddannelser med højere indkomst. De ti største lange videregående uddannelser for hhv. kvinder med dansk baggrund og efterkommerkvinder fremgår af bilag 1.

## Løngevinsten ved uddannelse for kvinder i beskæftigelse

Generelt fordeler indkomsten sig jævnt mellem de to grupper. Der er dog én vigtig undtagelse; der er langt flere efterkommerv kvinder med en erhvervskompetencegivende uddannelse, som har en erhvervsindkomst på 0 kr., og som dermed kan betragtes som uden beskæftigelse.

Det fremgår af figur 4, at mens andelen af efterkommere uden uddannelse, som er uden beskæftigelse, kun er marginalt større end kvinderne med dansk baggrund, så er andelen uden beskæftigelse mere end dobbelt så stor blandt kvinder med en uddannelse. Samlet set er 12 pct. af efterkommerv kvinderne med en kompetencegivende uddannelse uden beskæftigelse, mens det gælder 5 pct. af kvinderne med dansk baggrund.

**Figur 4: Andel uden beskæftigelse fordelt på kvinder med dansk baggrund og ikke-vestlige efterkommerv kvinder og erhvervskompetencegivende uddannelse**



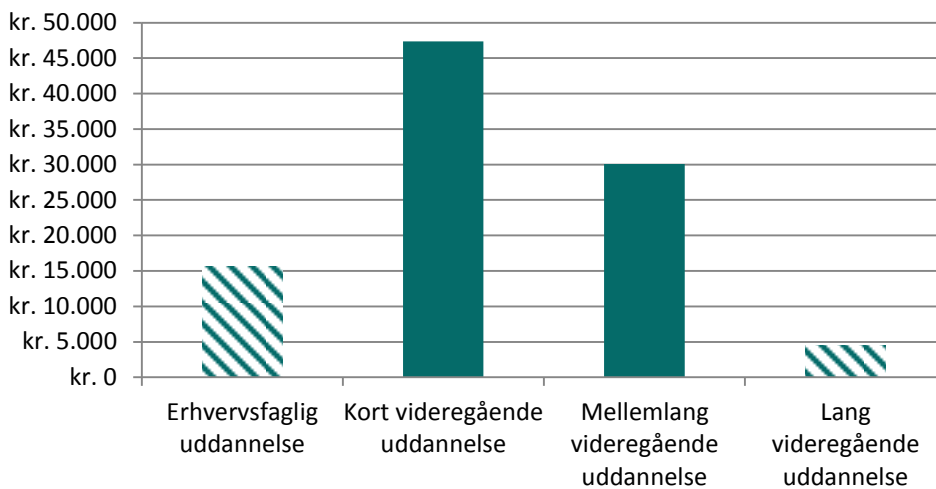
Anm.: Uden beskæftigelse defineres her som kvinder med en erhvervsindkomst på 0 kr. Gruppernes samlede indkomstfordeling på tværs af uddannelsesnivea fremgår af bilag 1. Ingen uddannelse inkluderer i analysen personer, som ikke har nogen registreret værdi for deres højst fuldførte uddannelse. Mellemlang videregående inkluderer universitetsbachelor.

Når den relative andel af efterkommere med en uddannelse, som er uden beskæftigelse, er større end kvinderne med dansk baggrund, vil det trække efterkommernes samlede løngevinst af uddannelse ned i forhold til kvinderne med dansk baggrund, idet personer uden beskæftigelse inkluderes i gevinstberegningen. Det er derfor interessant at undersøge, hvorvidt der er forskelle på de to gruppers løngevinst, hvis man alene ser på dem, som er i beskæftigelse, dvs. som ikke har en erhvervsindkomst på 0 kr.

Figur 5 viser resultatet af en statistisk model, hvor vi ser på forskellen i løngevinst for beskæftigede kvinder med hhv. dansk baggrund og efterkommere. Figuren viser, at når de beskæftigede sammenlignes, er efterkommerv kvindernes løn-

vinst større end kvinderne med dansk baggrund. Forskellen er dog kun statistisk sikker på de korte- og mellemlange videregående uddannelser.

**Figur 5: Beskæftigede ikke-vestlige efterkommerkvinders løngevinst ift. beskæftigede kvinder med dansk baggrunds løngevinst, fordelt på uddannelsesniveau**



Anm.: Figuren viser beskæftigede ikke-vestlige efterkommerkvinders estimerede løngevinst relativt til kvinderne med dansk baggrund. Søjlernes sribede udfyldning angiver, at forskellene i efterkommere og kvinder med dansk baggrunds løngevinst *ikke* med statistisk sikkerhed er forskellig fra 0 ved et 5%-signifikansniveau. Personer uden erhvervskompetencegivende uddannelse inkluderer i analysen personer, som ikke har nogen registreret værdi for deres højst fuldførte uddannelse. Mellemlang videregående inkluderer universitetsbachelor. Løngevinsterne i figuren er estimeret på baggrund af regressionsmodeller med både kvinder med dansk baggrund og efterkommerkvinder. En beskrivelse af metoden findes i bilag 2.

Der kan være forhold, som har betydning for, om kvinderne er i beskæftigelse eller ej, og som samtidig påvirker løngevinsten ved uddannelse, men som ikke kan forklares med de variable, som indgår i modellen. Hvis disse forhold er forskellige for efterkommerkvinder og kvinder med dansk baggrund, vil beregningen af forskelle i løngevinst ved uddannelse for de beskæftigede skævvrides. Den benyttede model antager, at disse eventuelle forhold er ens på tværs af søskende.

## Bilag 1 – Beskrivelse af grupperne

**Tabel B.1: Deskriptiv statistik for de to grupper, først kvinder med dansk baggrund, dernæst for efterkommerkvinderne, i parentes fremgår hhv. antal personer og gennemsnit for den subgruppe, som har søskende i datasættet**

<b>DANSK BAGGRUND</b>						
Variabel	Antal personer		Gennemsnit		Min	Max
Erhvervsindkomst	370.527	(109.221)	kr. 290.038	(kr. 296.563)	kr. -2.000.000	kr.2.000.000
<b>Højest fuldført erhvervskompetencegivende uddannelse</b>						
Ingen	370.527	(109.221)	17%	(15%)	0	1
EUD	370.527	(109.221)	29%	(29%)	0	1
KVU	370.527	(109.221)	6%	(6%)	0	1
MVU	370.527	(109.221)	30%	(31%)	0	1
LVU	370.527	(109.221)	19%	(20%)	0	1
<b>Alder</b>						
24	370.527	(109.221)	3%	(2%)	0	1
25	370.527	(109.221)	4%	(3%)	0	1
26	370.527	(109.221)	5%	(4%)	0	1
27	370.527	(109.221)	5%	(5%)	0	1
28	370.527	(109.221)	6%	(6%)	0	1
29	370.527	(109.221)	6%	(6%)	0	1
30	370.527	(109.221)	6%	(7%)	0	1
31	370.527	(109.221)	6%	(7%)	0	1
32	370.527	(109.221)	6%	(7%)	0	1
33	370.527	(109.221)	6%	(7%)	0	1
34	370.527	(109.221)	6%	(7%)	0	1
35	370.527	(109.221)	6%	(7%)	0	1
36	370.527	(109.221)	7%	(7%)	0	1
37	370.527	(109.221)	7%	(7%)	0	1
38	370.527	(109.221)	7%	(6%)	0	1
39	370.527	(109.221)	7%	(6%)	0	1
40	370.527	(109.221)	8%	(6%)	0	1
<b>Hjemmeboende</b>	370.527	(109.221)	0%	(0%)	0	1
<b>Enlig</b>	370.527	(109.221)	30%	(28%)	0	1
<b>Antal børn</b>						
0	370.527	(109.221)	35%	(34%)	0	1
1	370.527	(109.221)	21%	(21%)	0	1
2	370.527	(109.221)	32%	(33%)	0	1
3	370.527	(109.221)	10%	(10%)	0	1
4	370.527	(109.221)	1%	(2%)	0	1
5-	370.527	(109.221)	0%	(0%)	0	1
<b>Barsel</b>	370.527	(109.221)	20%	(22%)	0	1



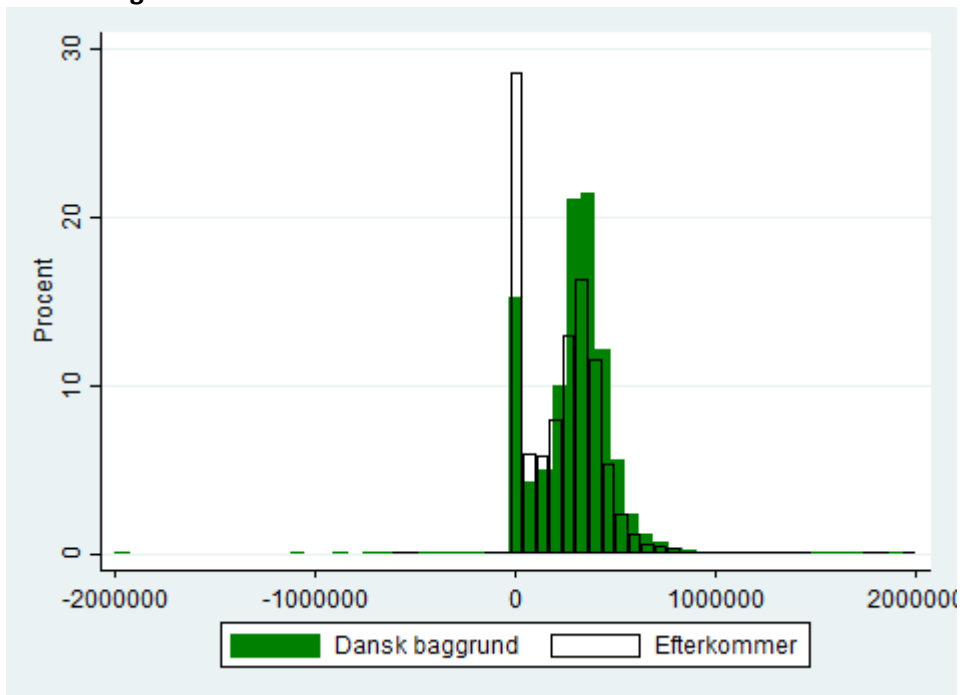
<b>Indlagt ifm. fødsel</b>	370.527	(109.221)	10%	(11%)	0	1
<b>Antal indlæggelser ifm. sygdom</b>	370.527	(109.221)	0,1	(0,1)	0	49
<b>Antal kontakter med almen læge</b>	370.527	(109.221)	8,9	(8,8)	0	371
<b>Kontakt til speciallæge</b>	370.527	(109.221)	32%	(31%)	0	1
<b>Antal kontakter med psykolog/psykiater</b>	370.527	(109.221)	0,4	(0,3)	0	69
<b>Fars højeste fuldførte erhvervskompetencegivende uddannelse</b>						
Ingen	370.527	(109.221)	28%	(27%)	0	1
Erhvervsfaglig	370.527	(109.221)	45%	(47%)	0	1
Videregående Uddannelse	370.527	(109.221)	23%	(24%)	0	1
Uoplyst	370.527	(109.221)	3%	(2%)	0	1
<b>Mors højeste fuldførte erhvervskompetencegivende uddannelse</b>						
Ingen	370.527	(109.221)	31%	(31%)	0	1
Erhvervsfaglig	370.527	(109.221)	38%	(37%)	0	1
Videregående uddannelse	370.527	(109.221)	29%	(31%)	0	1
Uoplyst	370.527	(109.221)	1%	(1%)	0	1
<b>Mors beskæftigelse som 14-årig</b>						
Beskæftigelse	370.527	(109.221)	82%	(84%)	0	1
Ledig/uden for arbejdsstyrken	370.527	(109.221)	16%	(14%)	0	1
Ingen oplysninger	370.527	(109.221)	2%	(1%)	0	1
<b>Boede med forældre som 14-årig</b>						
Begge forældre	370.527	(109.221)	68%	(77%)	0	1
Hos mor, der er i nyt par	370.527	(109.221)	11%	(7%)	0	1
Hos enlig mor	370.527	(109.221)	15%	(11%)	0	1
Hos far, der er i nyt par	370.527	(109.221)	1%	(1%)	0	1
Hos enlig far	370.527	(109.221)	2%	(2%)	0	1
Ikke hos forældrene	370.527	(109.221)	2%	(1%)	0	1
Ingen oplysninger	370.527	(109.221)	1%	(1%)	0	1
<b>Far ubetinget eller betinget dom som 0-14-årig</b>	370.527	(109.221)	8%	(6%)	0	1
<b>Antal i søskendeflok som 14-årig</b>	370.527	(109.221)	2,0	(2,3)	0	13
<b>Har søskende</b>	370.527	(109.221)	29%	(100%)	0	1
<b>IKKE-VESTLIG EFTERKOMMER</b>						
<b>Erhvervsindkomst</b>	10.212	(4.248)	kr. 224.312	(kr. 231.423)	kr. -613.472	kr.2.000.000
<b>Højest fuldført erhvervskompetencegivende uddannelse</b>						
Ingen	10.212	(4.248)	31%	(29%)	0	1
EUD	10.212	(4.248)	25%	(27%)	0	1
KVU	10.212	(4.248)	6%	(6%)	0	1

MVU	10.212	(4.248)	25%	(25%)	0	1
LVU	10.212	(4.248)	13%	(13%)	0	1
<b>Alder</b>						
24	10.212	(4.248)	8%	(5%)	0	1
25	10.212	(4.248)	8%	(5%)	0	1
26	10.212	(4.248)	8%	(7%)	0	1
27	10.212	(4.248)	8%	(7%)	0	1
28	10.212	(4.248)	8%	(8%)	0	1
29	10.212	(4.248)	8%	(8%)	0	1
30	10.212	(4.248)	7%	(7%)	0	1
31	10.212	(4.248)	6%	(7%)	0	1
32	10.212	(4.248)	5%	(6%)	0	1
33	10.212	(4.248)	4%	(5%)	0	1
34	10.212	(4.248)	5%	(6%)	0	1
35	10.212	(4.248)	5%	(6%)	0	1
36	10.212	(4.248)	5%	(6%)	0	1
37	10.212	(4.248)	4%	(5%)	0	1
38	10.212	(4.248)	4%	(5%)	0	1
39	10.212	(4.248)	3%	(4%)	0	1
40	10.212	(4.248)	3%	(3%)	0	1
<b>Hjemmeboende</b>	10.212	(4.248)	2%	(1%)	0	1
<b>Enlig</b>	10.212	(4.248)	41%	(39%)	0	1
<b>Antal børn</b>						
0	10.212	(4.248)	42%	(37%)	0	1
1	10.212	(4.248)	19%	(19%)	0	1
2	10.212	(4.248)	25%	(27%)	0	1
3	10.212	(4.248)	11%	(14%)	0	1
4	10.212	(4.248)	2%	(3%)	0	1
5-	10.212	(4.248)	1%	(1%)	0	1
<b>Barsel</b>	10.212	(4.248)	20%	(21%)	0	1
<b>Indlagt ifm. fødsel</b>	10.212	(4.248)	13%	(13%)	0	1
<b>Antal indlæggelser ifm. sygdom</b>	10.212	(4.248)	0,1	(0,1)	0	16
<b>Antal kontakter med almen læge</b>	10.212	(4.248)	9,9	(10,1)	0	82
<b>Kontakt til speciallæge</b>	10.212	(4.248)	40%	(41%)	0	1
<b>Antal kontakter med psykolog/psykiater</b>	10.212	(4.248)	0,2	(0,2)	0	33
<b>Fars højeste fuldførte erhvervskompetencegivende uddannelse</b>						
Ingen	10.212	(4.248)	52%	(57%)	0	1
Erhvervsfaglig	10.212	(4.248)	23%	(22%)	0	1
Videregående	10.212	(4.248)	13%	(12%)	0	1

Uddannelse						
Uoplyst	10.212	(4.248)	12%	(9%)	0	1
<b>Mors højeste fuldførte erhvervskompetencegivende uddannelse</b>						
Ingen	10.212	(4.248)	56%	(59%)	0	1
Erhvervsfaglig	10.212	(4.248)	18%	(17%)	0	1
Videregående Uddannelse	10.212	(4.248)	11%	(10%)	0	1
Uoplyst	10.212	(4.248)	14%	(14%)	0	1
<b>Mors beskæftigelse som 14-årig</b>						
Beskæftigelse	10.212	(4.248)	38%	(36%)	0	1
Ledig/uden for arbejdsstyrken	10.212	(4.248)	57%	(61%)	0	1
Ingen oplysninger	10.212	(4.248)	5%	(2%)	0	1
<b>Boede med forældre som 14-årig</b>						
Begge forældre	10.212	(4.248)	77%	(83%)	0	1
Hos mor, der er i nyt par	10.212	(4.248)	2%	(1%)	0	1
Hos enlig mor	10.212	(4.248)	14%	(11%)	0	1
Hos far, der er i nyt par	10.212	(4.248)	1%	(1%)	0	1
Hos enlig far	10.212	(4.248)	1%	(1%)	0	1
Ikke hos forældrene	10.212	(4.248)	1%	(1%)	0	1
Ingen oplysninger	10.212	(4.248)	4%	(2%)	0	1
<b>Far ubetinget eller betinget dom som 0-14-årig</b>	10.212	(4.248)	14%	(13%)	0	1
<b>Antal i søskendeflok som 14-årig</b>	10.212	(4.248)	3,0	(3,5)	0	12
<b>Har søskende</b>	10.212	(4.248)	42%	(100%)	0	1

Anm.: I parentes angives hhv. antal personer og gennemsnit for den del af datasættet, som har en søskende og derfor anvendes i den statistiske hovedmodel.

**Figur B.1: Fordeling af erhvervsindkomst for kvinder med dansk baggrund og for ikke-vestlige efterkommerkvinder**



Tabel B.2 viser, at efterkommerkvinderne og kvinderne med dansk baggrund har stort set samme uddannelser i top 5. Til gengæld udgør denne top 5 hhv. 27 pct. hos kvinderne med dansk baggrund og 41 pct. hos efterkommerne. Desuden er tre humanistiske uddannelser i top 10 hos kvinderne med dansk baggrund, mens der kun er en blandt efterkommerne.

**Tabel B.2: De ti største lange videregående uddannelser, fordelt på kvinder med dansk baggrund og efterkommerv kvinder**

Placering	Uddannelse	Procent	Akkumuleret
<b>DANSK BAGGRUND</b>			
1	Erhvervsøkonomi, cand.merc.	8%	8%
2	Jura, cand.jur.	8%	15%
3	Medicin, c.med.	5%	21%
4	Psykologi, c.psych.	4%	25%
5	Revisorkandidat, cand.merc.aud.	2%	27%
6	Kommunikation, cand.comm	2%	29%
7	Dansk, c.mag.	2%	31%
8	Statskundskab, cand.scient.pol.	2%	33%
9	Engelsk, c.mag.	1%	34%
10	Sundhedsvidenskab, ph.d.	1%	36%
<b>EFTERKOMMERE</b>			
1	Jura, cand.jur.	12%	12%
2	Erhvervsøkonomi, cand.merc.	9%	21%
3	Medicin, c.med.	9%	30%
4	Farmaci, c.pharm.	8%	37%
5	Revisorkandidat, cand.merc.aud.	4%	41%
6	Tandlæge, cand.odont.	3%	44%
7	Erhvervsøkonomi og jura, cand.merc.(jur.)	3%	47%
8	Psykologi, c.psych.	2%	50%
9	Socialt arbejde, c.soc.	2%	51%
10	Engelsk, c.mag.	2%	53%

## Bilag 2 – Den statistiske model

Analysens hovedmodel til at estimere uddannelsesniveaets betydning for indkomst, i denne analyse omtalt løngevinst, er en såkaldt *sibling fixed effects* model (se Produktivitetskommissionen 2013 eller Kraka 2015). En sådan model ser udelukkende på variation i indkomst inden for søskendepar. Tanken bag modellen er, at der kan være forhold, som er vanskelige at måle, der kan have væsentlig betydning for både uddannelsesvalg og indkomst. Det kunne fx være evner eller opvækstvilkår. Disse ikke-observerbare såvel som observerbare forhold skulle være nogenlunde konstante inden for søskendeparrene, og dermed kan der tages højde for dem ved kun at se på variation inden for søskendepar.

Modellen bygger på en central antagelse om, at der ikke er udeladt væsentlige individuelle faktorer, som varierer inden for søskendeparrene og korrelerer med både uddannelsesniveau og erhvervsindkomst. Det antages således eksempelvis, at aldersdummyerne opfanger forskelle i søstrenes erhvervs erfaring. Som det fremgår af bilag 4, er der som robusthedstjek kørt en model, som tillader heterogene alderseffekter, hvilket ikke ændrer ved de grundlæggende resultater.

Analysens hovedmodel kan beskrives ved følgende formel:

$$erh.ind_{is} = \beta_{0s} + \sum_{j=2}^5 \beta_{1j} udd_{jis} + \beta_2 eft_{is} + \sum_{j=2}^5 \beta_{3j} (udd_{jis} eft_{is}) + \mathbf{X}'_{is} \gamma + \varepsilon_{is}$$

Hvor:

- $erh.ind_{is}$  er erhvervsindkomsten for person  $i$  i søskendepar  $s$
- $\beta_{0s}$  er det søskendespecifikke konstantled i søskendepar  $s$
- $udd_{jis}$  er dummyvariable som indikerer uddannelsesniveau  $j$  for person  $i$  og søskendepar  $s$ , hvor  $j=1$  er referencekategorien og indikerer ingen uddannelse.
- $eft_{is}$  er en dummy som indikerer om person  $i$  og søskendepar  $s$  er efterkommer
- $\beta_{3j}$  er det centrale parameterestimat, som angiver forskellen mellem efterkommeres og kvinder med dansk baggrunds uddannelseseffekt for uddannelsesniveau  $j$ , beregnet som interaktionsledet  $eft_{is} * udd_{jis}$
- $\mathbf{X}'_{is}$  er en vektor af kontrolvariable, som opfanger individuelle baggrundskarakteristika, som varierer mellem personer inden for søskendeparrene. Vektoren inkluderer aldersdummyer samt forhold vedrørende civilstatus, herunder antal børn, om man er på barsel mv. samt psykiske og fysiske sundhedsforhold.

Man kunne have tilføjet eksempelvis grundskolekarakterer til rækken af kontrolvariable for at tage højde for eventuelle forskelle i evner inden for søskendeparrene, som ikke opfanges med den nuværende række variable. Men i datasættet findes kun oplysninger om grundskolekarakter på knap en tredjedel af personerne, da Danmarks Statistik først registrerer grundskolekarakterer fra skoleåret 2001/2002, og derfor er karakterer ikke inkluderet.

I den statistiske model, hvor der tages højde for uddannelsessammensætningen, benyttes en nyere matchingprocedure kaldet Coarsened Exact Matching (se Blackwell et al. 2009; Iacus, King & Porro 2012) til at skabe et datasæt, som er balanceret ud fra personernes uddannelser. På den måde fjernes personer, der ikke kan sammenlignes med andre personer i datasættet, mens andre vægtes op eller ned. Proceduren foregår ved, at uddannelser udvælges, som mindst én efterkommer og mindst én kvinde med dansk baggrund har fuldført. Dernæst tilskrives personerne en vægt ud fra, hvilken uddannelse de har. Matchede efterkommerkvinde får vægten 1, alle ikke-matchede personer får vægten 0, mens matchede kvinder med dansk baggrund får en vægt, hvormed personerne tilpasses efter, hvor mange fra begge grupper, der er med hver uddannelse. Vægtene beregnes med følgende formel:

$$w_i = \begin{cases} 1, & i \in T^u \\ \frac{m_C m_T^u}{m_T m_C^u}, & i \in C^u \end{cases}$$

hvor;

$w_i$  = vægten for individ  $i$

$m_C$  = antal kvinder med dansk baggrund

$m_T$  = antal efterkommerkvinder

$m_C^u$  = antal kvinder med dansk baggrund med uddannelsen  $u$

$m_T^u$  = antal efterkommere med uddannelsen  $u$

$T^u$  = efterkommere med uddannelse  $u$

$C^u$  = kvinder med dansk baggrund med uddannelse  $u$

**Tabel B.3: Resultatet af den valgte matching**

Antal uddannelser	1.429
Antal matchede uddannelser	362
Antal matchede efterkommerkvinder	4.222
Antal matchede kvinder med dansk baggrund	92.742

### Bilag 3 – Robusthedsanalyse

Foruden en lang række forskellige modelsøgninger og tests, er der gennemført tre robusthedsanalyser. Den første ser på betydningen af at inkludere aldersheterogene effekter, den anden ser på betydningen af at skifte indkomstbegreb og den tredje ser på betydningen af at skifte statistisk modeltype.

#### Aldersheterogene effekter

Som nævnt er den måske største udfordring ved at estimere forskellen på kvinder med dansk baggrunds- og efterkommerkvinders løngevinst, at grupperne har en meget forskellig alderssammensætning, da alder har stor betydning for indkomst. Ikke-vestlige efterkommerkvinder er væsentlig yngre end kvinder med dansk baggrund, som det fremgår af bilag 1. I den statistiske hovedmodel, som benyttes i denne analyse, inkluderes således aldersdummyer for at tage højde for disse aldersforskelle. Men man kan forestille sig, at der findes samspilseffekter, som gør at betydningen af alder for indkomst ikke er den samme for alle grupper. I denne robusthedsanalyse inkluderes derfor interaktioner mellem aldersdummyer og herkomst samt aldersdummyer og uddannelsesniveau som kontrolvariable for at tage højde for disse eventuelle aldersheterogene effekter. Løngevinsten estimeres fortsat på tværs af alder.

#### Indkomstbegreb

I den statistiske hovedmodel bruges erhvervsindkomst som afhængig variabel for at inkludere overskud fra selvstændig virksomhed. Som robusthedsanalyse er modellen gennemført med skattepligtig lønindkomst i stedet.

## Statistisk modeltype

Den statistiske hovedmodel ser alene på forskelle i løngevinst ved uddannelse mellem søskende. Dette for at tage højde for ikke-observerbare forskelle i fx evner. Men dermed reduceres det samlede datasæt til de kvinder, som har en eller flere søskende i datasættet. Som alternativ til denne metode er løngevinsten også forsøgt estimeret ved de mindste kvadraters metode (OLS) og ved en model, som antager en latent indkomst-variabel, der ikke begrænses ved 0 kr. og estimeres ved maximum likelihood (Tobit). Begge disse alternative modeller antager, at alle væsentlige individuelle karakteristika, som både påvirker uddannelsesvalg og erhvervsindkomst, er observerbare og kan kontrolleres for i modellen. Tobit-modellen antager endvidere, at fejledet er normalfordelt, hvilket OLS ikke gør. Der inkluderes således en række kontrolvariable om kvindernes familiebaggrund mv. i disse to modeller. Der kontrolleres for forældres uddannelsesniveau, mødrenes beskæftigelsessituation, da kvinderne var 14 år, fædrenes kriminalitetshistorik, da kvinderne var 0-14 år, hvorvidt de boede med forældrene som 14-årig og antal i søskendeflok som 14-årig. For at tage højde for, at hovedmodellen kun estimerer vha. subgruppen som har søskende i datasættet føjes også en dummy-variabel til rækken af kontrolvariable i OLS- og Tobit-modellerne.

## Resultat af robusthedsanalyse

Resultaterne af robusthedsanalysen fremgår af tabel B.4, hvor estimater fra hovedmodellen (1) kan sammenlignes med alternative modeller (4-7).

### Overordnede resultater

Overordnet set resulterer alle regressionsmodeller i ganske små forskelle i løngevinsterne ved uddannelse mellem danske kvinder og ikke-vestlige efterkommerkvinder, som kun i få tilfælde er statistisk sikre. Ved et almindeligt 5 % signifikansniveau er de eneste signifikante forskelle mellem de to grupper i OLS- og Tobit-modellerne, hvor efterkommerkvinder med en erhvervsfaglig uddannelse har en mindre løngevinst ved deres uddannelse end erhvervsfagligt uddannede danske kvinder samt dem med kort videregående uddannelse i OLS-modellen. Alle andre forskelle i løngevinst mellem de to grupper er på tværs af modellerne insignifikante (bemærk at model 2 og 3 ikke er en del af robusthedsanalysen).

Overordnet set kan det også nævnes, at model 1, 4 og 5, som alle er estimeret som sibling fixed-effects modeller, har stort set samme rangordning af uddannelserne, mens rangordningen er lidt anderledes i model 6 og 7. Da næsten alle estimater er statistisk insignifikante skal man dog være varsom med at fortolke på rangering og estimatstørrelser.

### Specifikke resultater

Aldersheterogene effekter:

Estimaterne ændrer sig kun ganske lidt og er heller ikke i denne model statistisk signifikante.



### Indkomstbegreb:

Det ændrer ikke nævneværdigt ved resultaterne at estimere løngevinsten ved lønindkomst i stedet for erhvervsindkomst.

### Alternative statistiske modeltyper:

Robusthedsanalysen viser for det første, at Tobit- og OLS-estimerne ligger ganske tæt på hinanden. For det andet viser disse alternative statistiske modeltyper et væsentligt andet billede end vores hovedmodel. I disse modeller klarer efterkommerne sig relativt dårligere ift. de danske kvinder, når vi ser på EUD-, KVVU- og MVU-uddannede. Dette er ikke overraskende, da man må forvente en hvis bias, idet disse modeller ikke nødvendigvis opfanger lige så mange forhold, eksempelvis forhold, der ikke kan observeres, som hovedmodellen. Med andre ord kan der være forhold, som trækker efterkommerkvinders løngevinst ned relativt til de danske kvinders, som bedre opfanges, når vi kun ser på forskelle i mellem søskende. Hvad angår de LVU-uddannede kvinder, er resultatet dog lidt mere overraskende, da efterkommerkvinder i hovedmodellen klarer sig en smule bedre end de danske kvinder i de alternative statistiske modeltyper. Forskellene er dog små og ikke statistisk signifikante.

**Tabel B.4: Efterkommerkvinders estimerede løngevinst ved uddannelse ift. danske kvinders ved forskellige statistiske modeller**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hovedmodel	Vægtet på uddannelse	Personer i job	inkl. aldersinteraktioner	Lønindkomst	OLS	Tobit
Ingen udd.	reference	reference	reference	reference	reference	reference	reference
EUD	-8.756 (9.495)	-7.608 (9.694)	15.644 (12.206)	-10.967 (9.444)	-10.109 (9.386)	-14.916*** (4.181)	-16.805*** (5.498)
KVVU	20.137 (15.585)	21.098 (15.985)	47.352*** (16.016)	16.212 (15.604)	15.309 (15.710)	-15.422** (7.777)	-14.585 (9.020)
MVU	18.567* (10.252)	11.708 (10.546)	30.052** (12.013)	15.012 (10.274)	13.574 (10.167)	2.691 (4.297)	5.424 (5.433)
LVU	-12.815 (15.606)	-41.196** (16.044)	4.544 (15.797)	-2.931 (15.507)	-17.632 (14.877)	-6.176 (6.592)	-3.160 (7.501)
Modeltype	Fixed effect	Fixed effect	Fixed effect	Fixed effect	Fixed effect	OLS	Tobit
Observationer	113.469	85.039	93.483	113.469	113.469	379.941	379.941

Klyngerobuste standardfejl i parentes \*\*\* p<0,01, \*\* p<0,05, \* p<0,1

## Litteratur

Arendt Nielsen, Jacob 2018: "Sammenhængen mellem beskæftigelse og uddannelse for ikke-vestlige flygtninge og familiesammenførte". *Arbejdsrapport 54*. Rockwoolfondens Forskningsenhed. Juni 2018.

Kraka 2015: "Løngevinst ved faglige 3F-uddannelser – Rapport til 3F til konferencen 'Vision Danmark'". Kraka analysenotat af Andreas Mølgaard og Nicolai Kaarsen. 21. december 2015.

Produktivitetskommissionen 2013: "Søskendes uddannelsesvalg og indkomst". Baggrundsnotat. 17. december 2013.

Blackwell, Matthew, Stefano Iacus, Gary King & Guiseppe Porro 2009: "cem: Coarsened exact matching in Stata" in *Stata Journal*, 2009, vol. 9, issue 4, 524-546

Iacus, Stefano M., Gary King & Guiseppe Porro 2012: "Causal Inference without Balance Checking: Coarsened Exact Matching". In *Political Analysis* 20, 1, Pp. 1--24.