



BØRNE- OG
UNDERVISNINGSMINISTERIET



Effekter af uddannelsesløft

Fra ufaglært til faglært

Effekter af uddannelsesløft
Fra ufaglært til faglært

2022

Design: Center for Kommunikation og Presse
Denne publikation kan ikke bestilles.
Der henvises til webudgaven.

Publikationen kan hentes på:
www.uvm.dk
Børne- og Undervisningsministeriet
Departementet
Frederiksholms Kanal 21
1220 København K

Indhold

1	Indledning og sammenfatning	4
1.1	Effekter af uddannelsesløft fra ufaglært til faglært	4
1.1.1	Rapportens hovedkonklusioner:	5
2	Eksisterende empiriske analyser	7
2.1	Fire danske analyser	7
2.2	Stort spænd i effektestimaterne	7
3	Genestimering af eksisterende analyser og estimering af effekter på beskæftigelsesgrad	9
3.1	Muligt at genskabe tre ud af fire studier	9
3.2	Genestimering af DØR (2018) og AE (2019) på nye data	9
3.3	Genestimering af Søgaard (2011) på nye data	10
3.4	Udvidelse af DØR (2018) og AE (2019)	12
3.5	Udvidelse af Søgaard (2011)	13
4	Estimeret beskæftigelseseffekt af opkvalificering af ufaglærte til faglært niveau, som funktion af antagelse om gennemslag	14
5	Litteratur	16
6	Appendiks	17
6.1	Arbejderbevægelsens Erhvervsråd (AE) 2019	17
6.1.1	Konstruktion af data	17
6.1.2	Empirisk model	18
6.2	De Økonomiske Råd (DØR) 2018	18
6.2.1	Konstruktion af data	18
6.2.2	Empirisk model	19
6.3	Søgaard (2011)	19
6.3.1	Konstruktion af data	20
6.3.2	Empirisk model	21

1 Indledning og sammenfatning

1.1 Effekter af uddannelsesløft fra ufaglært til faglært

Der er gode grunde til at forvente, at individers muligheder på arbejdsmarkedet påvirkes af deres uddannelsesniveau. Gennem uddannelse opnås kompetencer, som kan kvalificere til flere forskellige arbejdsfunktioner eller mere generelt kan øge produktiviteten. Derigennem kan uddannelse forventes at have effekter på såvel tilknytning til arbejdsmarkedet som på løn.

Hvor meget beskæftigelsen i samfundet kan påvirkes ved yderligere opkvalificering af befolkningen, er imidlertid vanskeligt at fastslå. Den observerede forskel i beskæftigelse mellem personer i forskellige uddannelseskategorier kan ikke alene danne grundlag for en sådan vurdering. Dels kan forskellen efter al sandsynlighed ikke alene tilskrives uddannelse – forskelle i personlige forudsætninger må forventes også at spille en rolle. Og dels kan effekten af uddannelse for personer, der ikke i dag er tilbøjelige til at opnå en uddannelse være forskellig fra effekten for de, der allerede er tilbøjelige til at blive uddannet. Det er eksempelvis muligt, at uddannelseseffekten for personer, der ikke allerede opnår uddannelse, er mindre end effekten for de, der gør – fordi de eksempelvis har ringere forudsætninger i form af helbred, motivation eller intelligens for at klare sig godt på arbejdsmarkedet, selv hvis de opnår højere uddannelse (Søgaard 2011). Det vil sige, det er muligt, at marginaleffekten af yderligere uddannelsesløft er lavere end den gennemsnitlige effekt af uddannelse historisk.

Denne rapport har til formål at belyse og udvide det empiriske grundlag for at vurdere beskæftigelses-effekter af at løfte ufaglærte til faglært niveau, eksempelvis i forbindelse med konsekvensvurderinger. DØR (2021) anbefaler, at der løbende sker udvikling og forbedring af vidensgrundlaget for konsekvensvurderinger, herunder blandt andet, at beregningsforudsætninger med regelmæssige mellemrum opdateres i forhold til datagrundlaget, erfaringer fra Danmark og udlandet, og den videnskabelige litteratur¹. Rapporten er udarbejdet af Børne- og Undervisningsministeriet (BUVM) og indeholder en genestimering på nye data, af tre eksisterende danske studier, som har søgt at belyse beskæftigelses-effekterne ved uddannelsesløft. Det drejer sig om DØR (2018), AE (2019) og Søgaard (2011). Desuden præsenteres resultaterne af Bækgaard og Helsø (2018). Da det ikke har været muligt at genskabe den oprindelige analyses resultater, genestimeres sidstnævnte dog ikke på nye data. Der tages i indeværende rapport ikke særskilt stilling til kvaliteten af de metodiske tilgange, idet en gennemgang og vurdering af studierne metodiske tilgange, herunder metodiske udfordringer og usikkerheder, som er væsentlige for fortolkningen af estimationsresultaterne, fremgår af Andersen m.fl. (2021), som publiceres sammen med denne rapport. I tillæg til genestimering på nye data, udvides analyserne i rapporten til også at se på effekter af uddannelse på beskæftigelsesgrad, dvs. hvor meget den enkelte arbejder.

De genestimerede analyser adskiller sig indbyrdes bl.a. ved, at nogle ser på effekten af uddannelse på beskæftigelsen, mens andre fokuserer på erhvervsdeltagelsen, dvs. summen af beskæftigede og ledige. Det diskuteres i det følgende ikke, hvilket mål der er mest relevant i forhold til at evaluere effekterne af

¹ DØR anbefaler at "[d]er bør ske en løbende udvikling og forbedring af vidensgrundlaget for konsekvensvurderingerne. Med regelmæssige mellemrum bør anvendte modeltilgange og beregningsforudsætninger opdateres i forhold til datagrundlaget, erfaringer fra Danmark og udlandet og den videnskabelige litteratur" DØR (2021) "Dansk Økonomi, efterår 2021, kapitel V, s. 275.

uddannelse. Det bemærkes dog, at såfremt der antages et konstant niveau for ledigheden ved en opkvalificering af befolkningen, er effekter på beskæftigelse og erhvervsdeltagelse ens.

1.1.1 Rapportens hovedkonklusioner:

- Punktestimater fra eksisterende analyser med et specifikt fokus på uddannelsesløft fra ufaglært til faglært (DØR 2018, AE 2019 og Bækgaard og Helsø 2018) giver anledning til et gennemslag på beskæftigelsen på mellem 25 pct. og 67 pct.² Det vil sige, at punktestimaterne indebærer, at et uddannelsesløft til ufaglærte forventeligt vil forøge beskæftigelsesfrekvensen for disse med mellem 25 pct. og 67 pct. af forskellen i gennemsnitlig beskæftigelse mellem ufaglærte og faglærte. For en faglig vurdering af de anvendte metoder, henvises der til Andersen m.fl. (2021).
- BUVM har kunnet genskabe de oprindelige resultater fra to af disse analyser på tilfredsstillende vis, og punktestimater fra genestimering af disse analyser på nyere data giver anledning til gennemslag på mellem 41 pct. og 66 pct. (DØR 2018 og AE 2019).
- Baseret på forskellen i beskæftigelsesfrekvens mellem ufaglærte og faglærte 30-årige i 2019, kan et gennemslag på hhv. 41 pct. og 66 pct. oversættes til en beskæftigelseseffekt som 30-årig på ca. 135 personer og ca. 220 personer ved et hypotetisk løft af 1.000 personer fra ufaglært til faglært niveau.
- Betragtes arbejdsmarkedstilknytning målt ved beskæftigelsesgraden (det vil sige, hvor stor en andel af fuldtid, svarende til 37 timer ugentligt, den enkelte arbejder), giver analyseresultaterne anledning til gennemslag mellem 35 pct. og 65 pct.
- Baseret på forskellen i beskæftigelsesgraden mellem ufaglærte og faglærte 30-årige i 2019, kan et gennemslag på hhv. 35 pct. og 65 pct. oversættes til en beskæftigelseseffekt som 30-årig på ca. 125 og ca. 230 fuldtidspersoner ved et hypotetisk løft af 1.000 personer fra ufaglært til faglært niveau. Dette kan betragtes som summen af effekten på beskæftigelsesfrekvensen og arbejdstiden.
- Estimaterne er alle behæftede med betydelige usikkerheder. Andersen m.fl. (2021) konkluderer, at ingen af analyserne giver et statistisk velidentificeret bud på den marginale effekt af at gennemføre en faglært uddannelse (Andersen m.fl. 2021, s. 27). Dette bør understreges, når der foretages evalueringer af beskæftigelseseffekter af uddannelsespolitiske tiltag mv. Det kan eksempelvis være relevant at præsentere en kvalitetsdeklaration sammen med beregningerne, jf. også DØR (2021). Det vil ligeledes være relevant at inddrage følsomhedsberegninger ved konkrete anvendelser, jf. både Andersen m.fl. (2021) og DØR (2021).

² Det fjerde studie, Søgaard (2011) differentierer ikke mellem uddannelsesløft på forskellige niveauer, og har altså ikke specifikt fokus på uddannelsesløft fra ufaglært til faglært, hvorfor det ikke omfattes af opsummeringen her.

Tabel 1: Opsummering af centrale empiriske estimater fra analysen

	Gennemslag på beskæftigelses-/erhvervsfrekvens i oprindelig analyse (pct.)	Muligt at genskabe oprindelige resultater	Gennemslag beskæftigelses-/erhvervsfrekvens på nye data (pct.)	Gennemslag på arbejdstid målt ved beskæftigelsesgrad (pct.)
Uddannelsesløft fra ufaglært til faglært				
AE (2018) ^a	67	Ja	66	65
DØR (2019) ^b	39 eller 25	Ja	51 eller 41	46 eller 35
Bækgaard og Helsø (2018) ^c	38	Nej	-	-
Generelt uddannelsesløft				
Søgaard (2011) ^d				
Direkte estimering	47	Ja	38	-
Indirekte estimering	43	Omtrentligt	-3	-

Anm.: (a) Baseret på resultater for de "bedst egnede" undtagen førtidspensionister. AE (2019) præsenterer flere resultater – i indeværende rapport følges gengivelsen af det centrale estimat i Andersen m.fl. (2021); (b) To resultater præsenteres som udtryk for den marginale effekt af en ufaglært uddannelse i DØR (2018) – et baseret på ufaglærte med over 90 pct. gennemførelsessandsynlighed (første tal) og et baseret på et estimeret tredje-gradspolynomium (andet tal), jf. afsnit 2. Her præsenteres begge estimater. Resultaterne præsenteres ikke som gennemslag i den oprindelige analyse. I BUVMs genestimering har det ikke været muligt at inddrage sessionsdata, som ellers indgår i den oprindelige analyse. Dette vurderes ikke at påvirke resultaterne nævneværdigt, da det har været muligt at genskabe punkttestimatet for uddannelseseffekten i den oprindelige analyse, og da sessionsdata i den oprindelige analyse betyder ganske lidt for de individualspecifikke effekter, særligt ved høje estimerede gennemførelsessandsynligheder. (c) Baseret på et uddannelsesløft af 1 pct. af analysens yngste kohorter fra ufaglært til faglært niveau. Her er angivet gennemsnittet for mænd og kvinder. I den oprindelige analyse præsenteres gennemslagene separat. (d) Søgaard (2011) præsenterer resultater for en lang række alders- og årsgropper. Det oprindelige resultat præsenteret her refererer til 30-59-årige i perioden 1980-2006. Der differentieres i studiet ikke mellem uddannelsesløft på forskellige niveauer, hvorfor resultatet ikke er et estimat for et løft fra ufaglært til faglært specifikt, og heller ikke er sammenligneligt med resultaterne fra de øvrige analyser. Analysen præsenterer resultater for to forskellige tilgange (en "direkte" og "indirekte", beskrevet i appendiks). I genestimeringen på nye data findes der, på baggrund af den indirekte estimering, et negativt gennemslag, dog meget tæt på nul. Et negativt gennemslag vil, givet at beskæftigelsesniveauet er højere på højere uddannelsesniveauer, indikere, at et uddannelsesløft fører til en reduktion af beskæftigelsen. At tilgangen fører til et sådant kontraintuitivt resultat, og at resultaterne baseret på den "direkte" og "indirekte" tilgang afviger meget, indikerer udfordringer med metodens pålidelighed.

Notatet er struktureret således: I afsnit 2 præsenteres resultaterne af fire danske studier, der har forsøgt at identificere effekten af uddannelse på beskæftigelse eller erhvervsfrekvens. I afsnit 3 præsenteres resultaterne af BUVMs genestimering af disse studier. Endelig illustreres, i afsnit 4, implicerede beskæftigelseseffekter (målt i personer eller fuldtidspersoner), som funktion af antagelse om gennemslaget, ved et hypotetisk uddannelsesløft af 1.000 30-årige fra ufaglært til faglært. Estimationsligninger og information om data findes i appendiks.

2 Eksisterende empiriske analyser

2.1 Fire danske analyser

Der er identificeret fire empiriske analyser på danske data, der har søgt at fastlægge beskæftigelses- eller erhvervsdeltagelseeffekter af uddannelse, jf. Andersen m.fl. (2021). Det drejer sig om:

- Arbejderbevægelsens Erhvervsråd (AE), 2019, Fra ufaglært til faglært forbedrer de offentlige finanser.
- De Økonomiske Råds Formandskab (DØR), 2018, Dansk Økonomi, Efterår 2018, kapitel 3: "Ufaglærtes tilknytning til arbejdsmarkedet".
- Bækgaard, H. og A.-L. Koch Helsø, 2018, The Danish Labour Force 1980-2015 – and the importance of education, DREAM Working Paper 2018:1.
- Søgaard, J.E., 2011, Sammenhængen mellem uddannelse og erhvervsdeltagelse, Arbejdspapir 24/2011, Finansministeriet³.

Metoderne, der anvendes i de fire danske analyser, afviger indbyrdes betragteligt i deres forsøg på at håndtere de generelle udfordringer ved at vurdere effekten af uddannelsesløft til ufaglærte (se Andersen m.fl. (2021) for en diskussion af disse udfordringer). Disse forskelle kan bidrage til at forklare forskellene i de rapporterede resultater. I Andersen m.fl. (2021) konkluderes det, at ingen af analyserne kan siges at give et statistisk velidentificeret bud på den marginale effekt af en faglært uddannelse (Andersen m.fl. 2021, s. 27).

2.2 Stort spænd i effektestimaterne

Tre af analyserne har specifikt fokus på afkastet af en faglært uddannelse. Det drejer sig om AE (2019), DØR (2018) og Bækgaard og Helsø (2019)⁴. I alle tre studier findes en umiddelbar forskel i beskæftigelses- eller erhvervsfrekvensen mellem ufaglærte og faglærte på ca. 35 pct. point. Analyserne forsøger på forskellig vis at korrigere for forskelle mellem ufaglærte og faglærte, som påvirker beskæftigelsen eller erhvervsfrekvensen ud over uddannelsen i sig selv. Ligeledes forsøger de tre analyser på forskellig vis at adressere, at den marginale effekt af et yderligere uddannelsesløft efter al sandsynlighed er forskellig fra den gennemsnitlige effekt af uddannelse, jf. Andersen m.fl. (2021). Analyserne når hver især frem til flere forskellige estimater for effekten af uddannelse, for forskellige grupper af ufaglærte. I Andersen m.fl. (2021) fremhæves følgende resultater, som vedrører forventede marginale effekter af et uddannelsesløft til ufaglærte (se også tabel 2):

- AE (2019) finder, at for de 25 pct. "bedst egnede" ufaglærte, forventes beskæftigelsesfrekvensen at stige med 25 pct. point ved opkvalificering til en faglært uddannelse.

³ Søgaard (2011) udgjorde baggrunden for fastlæggelsen af størrelsesordenen af de gennemslag, som finansministeriet benytter i de mellemfristede fremskrivninger, jf. Finansredegørelse 2014.

⁴ I Bækgaard og Helsø (2019) er fokus ikke begrænset til et løft fra ufaglært til faglært. Der estimeres separate effekter for forskellige typer af uddannelsesløft, hvor uddannelsesløftet fra ufaglært til faglært er et.

- DØR (2018) finder, at for de 100 ufaglærte, der i analysen vurderes at have over 90 pct. gennemførelssandsynlighed ved påbegyndelse af en erhvervsuddannelse⁵, er den forventede effekt en forøgelse af beskæftigelsesfrekvensen på 14 pct. point. Den forventede beskæftigelseseseffekt for ufaglærte med 100 pct. gennemførelssandsynlighed bestemmes på baggrund af et estimeret tredjegradspolynomium for sammenhængen mellem beskæftigelseseffekt og gennemførelssandsynlighed til at være 9 pct. point. At gennemslaget er faldende i gennemførelssandsynligheden, afspejler, at de personer, der har højst estimeret sandsynlighed for at gennemføre en erhvervsuddannelse, samtidig er dem, der har højst beskæftigelsesfrekvens som ufaglærte, hvorfor den potentielle beskæftigelsesfremgang må være mindre.
- Bækgaard og Helsø (2018) finder, at hvis 1 pct. af analysens yngste mandlige fødselsårgang løftes fra ufaglært til faglært niveau, forventes hele den mandlige fødselskohortes erhvervsfrekvens at stige med 0,08 pct. point. For kvinder er det tilsvarende tal 0,18 pct. point.
- Disse resultater kan omsættes til såkaldte gennemslagsgrader, som udtrykker de estimerede effekter ved uddannelsesløft, som andelen af den umiddelbare forskel i beskæftigelses- eller erhvervsfrekvens - mellem faglærte og ufaglærte. Det refererede resultat fra AE (2019) giver anledning til et gennemslag på 67 pct.; resultaterne fra DØR (2018) giver anledning til et gennemslag på hhv. 39 pct. og 25 pct. (selvom effektestimaterne ikke oversættes til gennemslag i DØRs analyse) og resultaterne fra Bækgaard og Helsø (2018) giver anledning til et gennemslag på 24 pct. for mænd og 51 pct. for kvinder. Resultaterne er opsummeret i tabel 2.

Det fjerde studie, Søgaard (2011), fokuserer ikke specifikt på et uddannelsesløft fra ufaglært til faglært, men konkluderer, at et gennemsnitligt fremadrettet skøn for gennemslaget af et øget uddannelsesniveau i befolkningen ligger i intervallet 0-40 pct. Det er væsentligt at understrege, at dette ikke er et skøn for gennemslaget af et uddannelsesløft fra ufaglært til faglært niveau specifikt, men et skøn for et generelt uddannelsesløft i befolkningen.

Tabel 2: Estimerede effekter og gennemslag på beskæftigelses- eller erhvervsfrekvens af et uddannelsesløft fra ufaglært til faglært i danske analyser.

	Estimeret effekt (pct. point)	Gennemslag (pct.)
AE (2019) ^a	25	67
DØR (2018) ^b	14 eller 9	39 eller 25
Bækgaard og Helsø (2018) ^c	13	38

Anm.: Resultaterne fra Søgaard (2011) er udeladt her, da de ikke refererer til uddannelsesløft fra ufaglært til faglært specifikt. Alle analyserne præsenterer flere forskellige estimater for effekten af uddannelse. Her gengives alene hovedresultaterne som præsenteret i Andersen m.fl. (2021). (a) Baseret på resultater for de "bedst egnede" undtagen førtidspensionister; (b) To resultater præsenteres som udtryk for den marginale effekt af en ufaglært uddannelse i DØR (2018) - et baseret på ufaglærte med over 90 pct. gennemførelssandsynlighed (første tal) og et baseret på et estimeret tredjegradspolynomium (andet tal), jf. hovedtekst. Resultaterne præsenteres ikke som gennemslag i den oprindelige analyse. (c) Baseret på et uddannelsesløft af 1 pct. af analysens yngste kohorter fra ufaglært til faglært niveau. Her er angivet gennemsnittet for mænd og kvinder, og det er skaleret med 100 for sammenlignelighed med de andre analyser. I den oprindelige analyse præsenteres effekter og gennemslag separat for mænd og kvinder.

⁵ DØR (2018) prædikerer, på baggrund af en estimeret model, hver af de ufaglærtes sandsynlighed for at gennemføre en erhvervsuddannelse, såfremt de påbegynder en, og finder at blandt de ca. 5.400 ufaglærte i analysepopulationen, har kun ca. 100 af dem over 90 pct. sandsynlighed for at gennemføre.

3 Genestimering af eksisterende analyser og estimering af effekter på beskæftigelsesgrad

3.1 Muligt at genskabe tre ud af fire studier

Børne- og Undervisningsministeriet har søgt at replikere og med nye data genestimere og udvide de fire eksisterende danske analyser, som har forsøgt at kvantificere beskæftigelses- eller arbejdsudbudseffekter ved uddannelsesløft.

I forbindelse med replikationen er det lykkedes at genskabe resultaterne fra DØR (2018), AE (2019), og i nogen grad for Søgaard (2011). Dog bemærkes det, at resultaterne fra Søgaard (2011) i replikationen viser sig meget følsomme overfor, hvordan observationer med manglende uddannelsesoplysninger håndteres. Desuden bemærkes det, at Søgaard (2011) anvender to estimeringsmetoder: En direkte, og en indirekte (se Andersen m.fl. 2021). Resultaterne fra replikationen af den indirekte estimeringstilgang afviger mere fra den oprindelige analyse, end det er tilfældet for den direkte estimering. Dette kunne indikere, at den indirekte estimering er mere følsom overfor mindre variationer i data og/eller variabelkonstruktion, end den direkte estimering.

At det ikke har været muligt at genskabe resultaterne for Bækgaard og Helsø (2018), skyldes formentlig en kombination af, at resultaterne kan være følsomme overfor mindre forskelle i data og variabelkonstruktion, og at beskrivelsen af tilgangen i den oprindelige artikel ikke er fyldestgørende til formålet. Da Bækgaard og Helsø (2018) ikke har kunnet replikeres, er analysen heller ikke genestimeret på nye data eller udvidet til at omfatte arbejdstid.

3.2 Genestimering af DØR (2018) og AE (2019) på nye data

Analysen i DØR (2018) er baseret på 29-årige i 2016, mens analysen i AE (2019) er baseret på 30-årige i 2017. De nyeste tilgængelige data for beskæftigelsen, som den er opgjort i analyserne, er fra 2019. De nye resultater er for begge analyser baseret på sammenlægning af data fra 2018 og 2019, og der kontrolleres i analysen for år.

Tabel 3: Sammenligning af resultater fra oprindelige analyser og genestimering på nye data baseret på DØR (2018) og AE (2019).

	Baseret på alle ufaglærte		Baseret på ufaglærte på marginen af uddannelse ^a	
	Oprindelig analyse	Nye data	Oprindelig analyse	Nye data
Baseret på DØR (2018)				
Beskæftigelseseffekt (pct. point)	29	38	14 eller 9	22 eller 18
Gennemslag (pct.)	81	88	39 eller 25	51 eller 41
Baseret på AE (2019)				
Beskæftigelseseffekt (pct. point)	32	33	25	24
Gennemslag (pct.)	86	89	67	66

Anm.: Opdateringen er baseret på data fra 2018 og 2019. (a) For analyser baseret på DØR (2018) er dette defineret som enten ufaglærte med over 90 pct. estimeret gennemførelssandsynlighed ved påbegyndelse af en erhvervsuddannelse (første tal) eller ufaglærte med teoretisk 100 pct. gennemførelssandsynlighed (andet tal); for analyser baseret på AE (2019) er det defineret som de 25 pct. "bedst egnede" undtagen førtidspensionister. Bemærk at det i BUVMs genestimering af (DØR) ikke har været muligt at inddrage sessionsdata, som ellers indgår i den oprindelige analyse. Dette vurderes ikke at påvirke resultaterne nævneværdigt, da det har været muligt at genskabe punktestimatet for uddannelseseffekten i den oprindelige analyse, og da sessionsdata i den oprindelige analyse betyder ganske lidt for de individualspecifikke effekter, særligt ved høje estimerede gennemførelssandsynligheder.

Det fremgår af tabel 3, at genestimering af analyserne på nye data giver anledning til estimater for gennemslaget på beskæftigelsesfrekvensen af et uddannelsesløft fra ufaglært til faglært, som er meget lig resultaterne fra den oprindelige analyse hvad angår AE (2018), men som er højere hvad angår DØR (2018). I genestimeringen af AE (2019) findes et gennemslag baseret på den fulde analysepopulation af ufaglærte på 89 pct., mod 86 pct. i den oprindelige analyse. På baggrund af den gruppe, som i analysen vurderes mest relevant i forhold til at evaluere den marginale effekt af uddannelsesløft, dvs. de 25 pct. "bedst egnede" eksklusiv førtidspensionister (jf. ovenfor), findes et gennemslag på 66 pct., mod 67 pct. i den oprindelige analyse.

I genestimeringen af DØR (2018) på nye data findes et gennemslag, baseret på den fulde analysepopulation af ufaglærte, på 88 pct., mod 81 pct. i den oprindelige analyse. Hvad angår ufaglærte på marginen af uddannelse, fås der ved estimering på nye data også højere punktestimater for beskæftigelseseffekten, end i den oprindelige analyse. Baseret på de personer, der på baggrund af analysen vurderes at have mere end 90 pct. sandsynlighed for at gennemføre en erhvervsuddannelse, såfremt de påbegynder den, findes i genestimeringen et gennemslag på 51 pct., mod 39 pct. på baggrund af den oprindelige analyse. For ufaglærte med en gennemførelssandsynlighed på 100 pct. bestemmes gennemslaget til at være 41 pct. mod 25 pct. på baggrund af den oprindelige analyse.

For en faglig vurdering af metoderne henvises der til Andersen m.fl. 2021, s. 9-14 (omhandlende AE 2019) og s. 14-18 (omhandlende DØR 2018).

3.3 Genestimering af Søgaard (2011) på nye data

Analysen i Søgaard (2011) giver, som tidligere nævnt, ikke mulighed for at konkludere specifikt på effekten af et uddannelsesløft fra ufaglært til faglært niveau (for en faglig vurdering henvises til Andersen m.fl. 2021 s. 19-25). Ikke desto mindre er analysen medtaget for fuldstændighedens skyld.

Tabel 4: Sammenligning af resultater fra oprindelig analyse og genestimering på nye data baseret på Søgaard (2011)

	Oprindelig analyse (1980-2006)	Oprindelig analyse (1994-2006)	Nye data (1980-2019)	Nye data (2008-2019)
Gennemslag (pct.) baseret på "direkte estimering"				
30-49-årige	25	-	35	55
30-54-årige	39	-	46	54
30-59-årige	47	39	52	38
Udtydningseffekt (pct. point) baseret på "indirekte estimering"				
30-49-årige	6,1	-	6,3	9,9
30-54-årige	5,3	-	6,3	9,8
30-59-årige	4,6	8,2	6,6	10,0
Gennemslag (pct.) baseret på "indirekte estimering".				
30-49-årige	25	-	38	-3
30-54-årige	35	-	38	-1
30-59-årige	43	1	35	-3

Anm.: Søgaard (2011) præsenterer resultater for en lang række alders- og årsgrupperinger, her gengives et udvalg. Der differentieres i studiet ikke mellem uddannelsesløft på forskellige niveauer, hvorfor resultatet ikke er et estimat for et løft fra ufaglært til faglært specifikt, og heller ikke er sammenligneligt med resultaterne fra de øvrige analyser, som er genestimeret i denne rapport. Analysen præsenterer resultater for to forskellige tilgange (en "direkte" og "indirekte", beskrevet i appendiks). I genestimeringen på nye data findes - på baggrund af den indirekte estimering - et negativt gennemslag, dog meget tæt på nul. Et negativt gennemslag vil, givet at beskæftigelsesniveauet er højere på højere uddannelsesniveauer, indikere, at et uddannelsesløft fører til en reduktion af beskæftigelsen. At tilgangen fører til et sådant kontraintuitivt resultat, og at resultaterne baseret på den "direkte" og "indirekte" tilgang afviger meget, indikerer udfordringer med metodens pålidelighed.

Den oprindelige analyse er baseret på arbejdsstyrkestatistik fra perioden 1980 til 2006. Det har været muligt at genestimere analysen med 13 års ekstra data, dækkende perioden frem til 2019. Der er estimeret to nye versioner: Én, der anvender data for hele perioden 1980 til 2019, og én der alene baseres på data for perioden 2008 til 2019. Resultaterne fra både den direkte og indirekte estimering er præsenteret i tabel 4 ovenfor (se appendiks for en beskrivelse af fremgangsmåden). Bemærk, at Søgaard (2011) præsenterer en lang række estimater for forskellige grupperinger af aldre og analyseår. Her gengives et udvalg.

Det fremgår af tabellen, at mens der ved den oprindelige analyse overvejende er god overensstemmelse mellem resultaterne fra den direkte og indirekte estimering, i hvert fald på baggrund af hele perioden 1980 til 2006⁶, er dette ikke tilfældet ved genestimeringen på data for perioden 2008 til 2019.

⁶ For perioden 1994-2006 giver den direkte og indirekte estimering anledning til vidt forskellige estimater i den oprindelige analyse, og det er disse resultater (hhv. 39 pct. og 1 pct.), der afgrænser intervallet for analysens skøn for det fremadrettede gennemslag af uddannelsesløft. I BUVMS forsøg på at genskabe den oprindelige analyses resul-

Her spænder de præsenterede estimater for gennemslag på erhvervsfrekvens baseret på den direkte tilgang fra 38 til 55 pct., mens det ved den indirekte estimering ligger omkring 0⁷.

Allerede ved replikationen fremstod den indirekte estimering følsom overfor potentielt mindre variationer i data og variabelkonstruktioner. I den sammenhæng, kan det være relevant, at der er et større data-brud i den registerbaserede arbejdsstyrkestatistik i 2008. Ligeledes kan det være relevant, at kvaliteten af uddannelsesdata forøgedes markant fra begyndelsen af 1970'erne, og de fødselsårangange, som blev uddannet før dette tidspunkt fylder mindre i analysepopulationen i de senere år. Endelig er det muligt, at de markante afvigelser mellem den direkte og indirekte estimering i senere år, hænger sammen med, at den generelle uddannelsesfremgang, der forekommer i senere perioder, er anderledes sammensat, end i de tidligere perioder. Tilgangene i Søgaard (2011) tillader, som nævnt, ikke at differentiere mellem forskellige former for uddannelsesløft.

I Søgaard (2011) vurderes resultaterne at være usikre i de tilfælde, hvor der er store afvigelser mellem resultater fra den direkte og indirekte estimering.

3.4 Udvidelse af DØR (2018) og AE (2019)

Analyserne fra DØR (2018) og AE (2019) er i forbindelse med genestimering udvidet til at omfatte effekter på beskæftigelsesgraden i november. En beskæftigelsesgrad på 100 pct. svarer til fuldtidsarbejde (37 timer ugentligt), og en person med en beskæftigelsesgrad på 100 pct. kan altså regnes for én fuldtidsperson. Det bemærkes, at der er stor usikkerhed forbundet med beskæftigelsesgraden, idet denne fx altid er beregnet for selvstændige og medarbejdende ægtefæller, mens den for lønmodtagere er baseret på betalte arbejdstimer ifølge e-Indkomstregistret.

Tabel 5: Udvidelse af analyser baseret på DØR (2018) og AE (2019), effekt på beskæftigelsesgrad.

	Beskæftigelsesgrad	
	Alle ufaglærte	Ufaglærte på marginen af uddannelse ^a
Baseret på DØR (2018)		
<i>Beskæftigelseseffekt (pct. point)</i>	34	19 eller 14
<i>Gennemslag (pct.)</i>	83	46 eller 35
Baseret på AE (2019)		
<i>Beskæftigelseseffekt (pct. point)</i>	34	25
<i>Gennemslag (pct.)</i>	88	65

Anm.: Analysen er baseret på data fra 2018 og 2019. (a) For analyser baseret på DØR (2018) er dette defineret som enten ufaglærte med over 90 pct. estimeret gennemførelsessandsynlighed ved påbegyndelse af en erhvervsuddannelse (første tal) eller ufaglærte med teoretisk 100 pct. gennemførelsessandsynlighed (andet tal); for analyser baseret på AE (2019) er det defineret som de 25 pct. "bedst egnede" undtagen førtidspensionister. Alle resultater er statistisk signifikante.

tater, findes for perioden 1995-2006 væsentligt højere gennemslag på hhv. 47 og 27 pct. ved den direkte og indirekte estimering. Det understreger, at analysen er meget følsom overfor forventeligt mindre variationer i datakonstruktion, og at der er stor usikkerhed forbundet med estimaterne.

⁷ Estimaterne er faktisk negative, men ganske små. Et negativt gennemslag ville, givet at beskæftigelsesniveauet er højere på højere uddannelsesniveauer, indikere, at et uddannelsesløft fører til en reduktion af beskæftigelsen. At tilgangen fører til et sådant kontraintuitivt resultat, indikerer udfordringer med metodens pålidelighed.

Resultaterne er gengivet i tabel 5 overfor. Bemærk, at der i analyserne ikke er betinget på, at man faktisk er i beskæftigelse (dvs. der indgår ikke-beskæftigede i analysen). Det betyder, at effekterne på beskæftigelsesgraden skal forstås som en kombination af effekter på den ekstensive og den intensive margin (beskæftigelsesfrekvens og beskæftigelsesomfang, givet beskæftigelse). Det fremgår af tabellen, at beregnede gennemslag på beskæftigelsesgraden ligger tæt på de beregnede gennemslag for beskæftigelsesfrekvensen. Baseret på de metodiske tilgange benyttet i AE (2019) og DØR (2018) findes således et gennemslag på beskæftigelsesgraden på hhv. 88 pct. og 83 pct., baseret på de fulde analysepopulationer. For gruppen af bedst egnede, som ikke omfatter førtidspensionister, findes på baggrund af den metodiske tilgang i AE (2019) et gennemslag på beskæftigelsesgraden på 65 pct. På baggrund af den metodiske tilgang i DØR (2018) findes for gruppen med over 90 pct. sandsynlighed for at gennemføre en faglært uddannelse et gennemslag på beskæftigelsesgraden på 46 pct. For ufaglærte med en gennemførelssandsynlighed på 100 pct. bestemmes gennemslaget til at være 35 pct.

3.5 Udvidelse af Søgaard (2011)

Også analysen af Søgaard (2011) er udvidet til at se på beskæftigelsesgrad. Da data for arbejdstid først er tilgængelige fra 2008, er denne udvidelse foretaget for perioden 2008 til 2019. Resultaterne viser sig, ligesom for erhvervsfrekvensen, at afvige meget på tværs af den direkte og indirekte estimering. På den baggrund vurderes resultaterne meget usikre, og gengives ikke her.

4 Estimeret beskæftigelseseffekt af opkvalificering af ufaglærte til faglært niveau, som funktion af antagelse om gennemslag

Der er store usikkerheder forbundet med alle gennemgåede analyser, jf. Andersen m.fl. (2021). I Andersen m.fl. (2021) konkluderes, at alle analyserne befinder sig et stykke fra toppen af evidenshierarkiet, og at ingen af analyserne kan siges at give et statistisk velidentificeret bud på den marginale effekt af at gennemføre en faglært uddannelse (Andersen m.fl. 2021, s. 27). Samtidig understreges, at selektive tiltag rettet mod personer, hvor forskellige sociale omstændigheder har været afgørende for manglende uddannelse, kan have en anden effekt end et løft rettet mod et repræsentativt udsnit af ufaglærte (Andersen m.fl. 2021, s. 29).

Da der er store empiriske udfordringer forbundet med at fastlægge størrelsesordner for effekter af uddannelse, og der derfor er faglig usikkerhed omkring størrelsesordenen, vurderer Andersen m.fl. (2021) det relevant eksplicit at inddrage følsomhedsanalyser i forbindelse med kvantificering af forventede effekter ved uddannelsesløft (Andersen m.fl. s. 30). Dette harmonerer med DØRs anbefalinger om, at der ved konsekvensvurderinger mv. bør præsenteres beregninger baseret på alternative forudsætninger (DØR 2021, s. 277). DØR (2021) anbefaler desuden, at der indføres kvalitetsdeklarationer både af konkrete konsekvensvurderinger og af de underliggende regneregler og metoder (DØR 2021, s. 272-275).

Figur 1 nedenfor illustrerer implicerede effekter på beskæftigelsen i hhv. antal personer og fuldtidspersoner, som funktion af antagelsen om størrelsen på gennemslaget. Der tages udgangspunkt i den observerede forskel i beskæftigelse på tværs af ufaglærte (her defineret som personer med ukendt uddannelse eller grundskole som højst fuldførte uddannelse, uanset øvrige karakteristika og uddannelseshistorik) og faglærte (personer med en erhvervsuddannelse som højst fuldførte uddannelse, uanset øvrige karakteristika og uddannelseshistorik) i alderen 30 år i november 2019. Forskellen i beskæftigelsesfrekvens for grupperne er 33 pct. point, og forskellen i beskæftigelsesgraden er 36 pct. point. Der regnes med et hypotetisk uddannelsesløft af 1.000 ufaglærte til faglært niveau. Et uddannelsesløft af 1.000 ufaglærte på én årgang vurderes at være en relativ ambitiøs målsætning, ligesom der heller ikke er anvist et specifikt tiltag til at generere det givne uddannelsesløft⁸. Den illustrative beregning tjener alene det formål at belyse størrelsesordenen på beskæftigelseseffekten ved forskellige antagelser, hvad angår gennemslaget. Beregningen er skalerbar.

Til illustration af den implicerede effekt ved konkrete antagelser om gennemslagets størrelse for uddannelsesløft anvendes resultaterne fra BUVMs genestimering af DØR (2018) og AE (2019), idet det samtidig skal bemærkes, at det ligger uden for indeværende rapport at fastlægge det relevante usikkerhedsinterval for specifikke beregningskontekster⁹. De nye punktestimater for effekten af et uddannelsesløft fra ufaglært til faglært på baggrund af tilgangene fra disse analyser kan omsættes til gennemslag på beskæftigelsesfrekvensen på hhv. ca. 41 pct. eller 51 pct. og ca. 66 pct., når der tages udgangspunkt i de delanalyser, der i Andersen m.fl. (2021) fremhæves som centrale resultater.

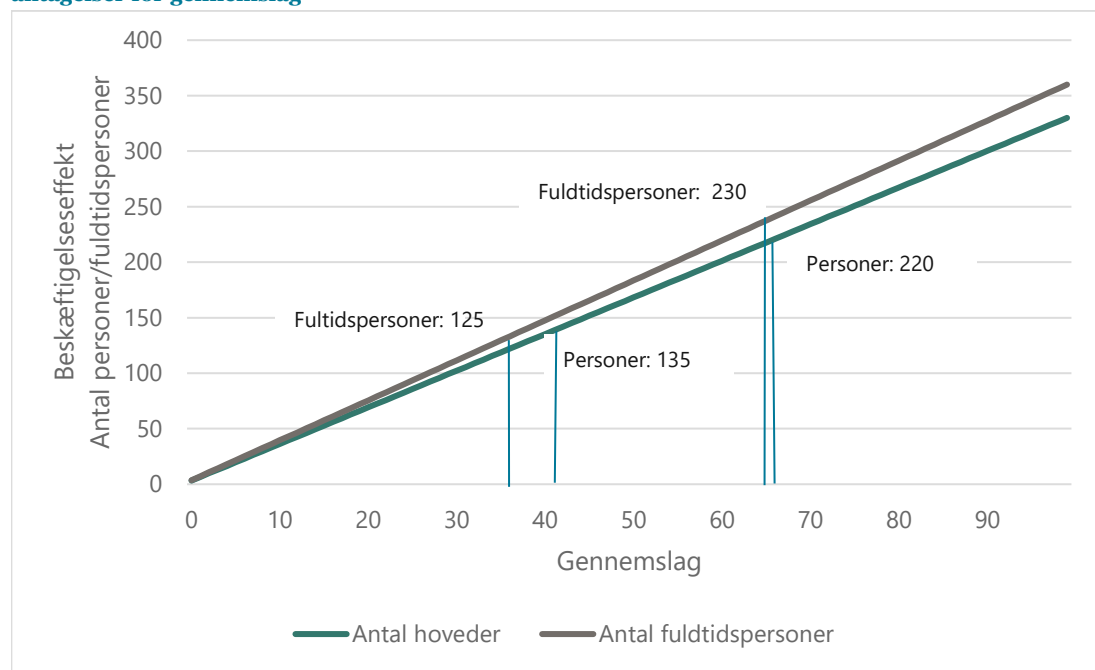
⁸ Det svarer til at løfte ca. 8 pct. af personer med ukendt uddannelse eller grundskole som højst fuldført uddannelse fra årgangen til faglært niveau.

⁹ Foruden usikkerheder forbundet med middelret identifikation af marginale effekter og heterogenitet, vil estimaterne desuden være forbundet med statistisk usikkerhed. Det skal i den forbindelse nævnes, at BUVM ikke er bekendt med en metode til at bestemme den statistiske usikkerhed for det laveste præsenterede estimat, ligesom dette heller ikke præsenteres i den oprindelige analyse.

Ved et løft af 1.000 ufaglærte giver et gennemslag på 41 pct. (det laveste estimat baseret på genestimering af DØR (2018)), med udgangspunkt i en forskel i beskæftigelsesfrekvensen på 33 pct. point, anledning til en forventet beskæftigelseseffekt på ca. 135 personer ($0,41 * 0,33 * 1000 = 135$), og et gennemslag på 66 pct. (baseret på genestimering af AE (2019)) giver anledning til en beskæftigelseseffekt på ca. 220 personer.

Et gennemslag på beskæftigelsesgraden på 35 pct. (det laveste estimat baseret på udvidelse af DØR (2018)) giver, med udgangspunkt i en forskel i beskæftigelsesgrad på 36 pct. point, anledning til en forventet effekt på ca. 125 *fuldtids*personer ($0,35 * 0,36 * 1000 = 126$), og et gennemslag på 65 pct. (baseret på udvidelse af AE (2019)) giver anledning til en effekt på ca. 230 fuldtidspersoner.

Figur 1: Beskæftigelseseffekt ved et løft af 1.000 ufaglærte til faglært niveau, ved forskellige antagelser for gennemslag



Anm.: Baseret på forskellen i beskæftigelse mellem faglærte og ufaglærte 30-årige i 2019, som er 33 pct. point. fsva. beskæftigelsesfrekvens og 36 pct. point fsva. beskæftigelsesgrad.

Kilde: Egne beregninger pba. den registerbaserede arbejdsstyrkestatistik (RAS).

5 Litteratur

Andersen, T. M., M. Svarer, R. M. Vejlin (2020) *Potentielle beskæftigelses- og løneffekter ved at løfte ufaglærte til faglært niveau – empirisk viden og mulighederne for metodisk udvikling*. Rapport udarbejdet for Børne- og Undervisningsministeriet.

Arbejderbevægelsens Erhvervsråd (2019) *Fra ufaglært til faglært forbedrer de offentlige finanser*.

Bækgaard, H. og A.-L. Koch Helsø (2018) The Danish Labour Force 1980-2015 – and the importance of education. *DREAM Working Paper*, 2018:1.

Cragg, J. G. (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39: 829–844.

De Økonomiske Råds Formandskab (2018) *Dansk Økonomi, Efterår 2018*.

De Økonomiske Råds Formandskab (2021) *Dansk Økonomi, Efterår 2021*.

Finansministeriet (2014) *Finansredegørelsen, Kap. 6: Uddannelse, vækst og offentlige finanser*.

Leuven, E. og B. Sianesi (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. <http://ideas.repec.org/c/boc/bocode/s432001.html>. Version 2.0.10

Søgaard, J.E. (2011) Sammenhængen mellem uddannelse og erhvervsdeltagelse, *Arbejdspapir 24/2011*, Finansministeriet.

6 Appendiks

6.1 Arbejderbevægelsens Erhvervsråd (AE) 2019

6.1.1 Konstruktion af data

Analysepopulationen er dannet på baggrund af alle 30-årige i befolkningen pr. 1. januar i det givne år (BEF registret). For replikeringen af AE (2019) analysen er det givne år 2017, og i genestimeringen samt i beregningerne af beskæftigelsesgrader er de givne år 2018 og 2019.

I analysepopulationen er der set bort fra personer, der er i gang med en uddannelse (baseret på Elevregistret og RAS), og personer der ikke var i Danmark, da de var 15 år (BEF registret). Derudover er der kun inkluderet personer, som er fuldt skattepligtige i Danmark (baseret på personindkomst registret) og ikke er udvandret i året (baseret på BEF registret).

Tilknytning til arbejdsmarkedet er bestemt vha. variablen SOC_STATUS_KODE, med formatet beskæftiget/ledig/uden for arbejdsmarkedet (se i øvrigt tabel i beskrivelsen af data til analysen baseret på Søgaard (2011)) og variablen STOETTE_BESK_KODE (begge fra RAS registret). En person betragtes som beskæftiget, hvis SOC_STATUS_KODE angiver personen som beskæftiget, og det ikke fremgår af STOETTE_BESK_KODE, at personen er i nogen form for støttet beskæftigelse. Endvidere tjekkes det via DREAM registret at personen i den tilsvarende uge for RAS-ugen ikke modtager en ydelse som indikerer, at de ikke er beskæftigede.

Beskæftigelsesgrad i november bestemmes ud fra variablen NOV_TILSTAND_GRAD fra RAS. For hver enkelt person summeres over alle november tilstande, som er klassificeret som "beskæftiget" (SOC_STATUS_KODE 110 til 136).

I analysen matches ufaglærte og faglærte på en lang liste af variable. Fra befolkningsregistret er hentet information om køn (mand/kvinde), herkomst (dansk/efterkommer eller indvandrer), opvokset hos enlig forælder målt ved 15 år (opvokset med enlig mor eller enlig far/alle andre kategorier), forældrenes uddannelsesniveau ved 15 år (grundskole, ukendt eller uoplyst/erhvervsfaglig eller gymnasial/videregående uddannelse), og bopæl i landsdele (Fyn/Københavns by/Københavns omegn/Nordjylland/Nordsjælland/Syddjylland/Vest og syd Sjælland/Vestjylland/Østjylland/Østsjælland og Bornholm). Fra RAS og DREAM registre er der konstrueret variable for forældrenes jobsituation (beskæftiget/ledig/uden for arbejdsmarkedet), og for hvorvidt personen har haft et ungdomsjob ved 14 år (havde job/havde intet job). På baggrund af oplysninger fra karakterregistret, er der dannet variable for afgangskarakterer i dansk (mangler eller ukendt/ikke opnået 2/opnået minimum 2) og matematik (mangler eller ukendt/ikke opnået 2/opnået minimum 2), for hvilken skoletype personen er afgang fra (folkeskole, øvrige eller ukendt/ fri- eller privatskole/efterskole), og for hvorvidt en person har gået i 10. klasse (ja/nej). I AEs analyse indgår også en variabel for boligforhold (eje/leje), da personen er 15 år. Denne information er ikke tilgængelig i BUVMs databeredskab og er derfor udeladt. I analysen på nye, sammenlagte data kontrolleres desuden for år.

6.1.2 Empirisk model

Der anvendes propensity score matching til estimering af effekt på såvel beskæftigelsesfrekvens som beskæftigelsesgrad. I praksis er der anvendt psmatch2¹⁰ i Stata16. Der er i analysen blevet matchet på de 5 nærmeste naboer. Da interessen er i uddannelseseffekten for ufaglærte, behandles ufaglærte som "treatment"-gruppe, og der estimeres effekt of treatment on the treated, ATT.

Foruden den overordnede gruppe af ufaglærte estimeres beskæftigelseseffekten for tre delpopulationer:

1. De "bedst egnede", defineret som de 25 pct. af de ufaglærte, der har størst estimeret sandsynlighed for at gennemføre en faglært uddannelse, baseret på alle personer i analysepopulationen.
2. Alle ufaglærte, med undtagelse af personer på førtidspension.
3. De "bedst egnede" jf. punkt 1, som ikke modtager førtidspension. Det er resultater for denne gruppe, der præsenteres i indeværende rapport.

6.2 De Økonomiske Råd (DØR) 2018

6.2.1 Konstruktion af data

For at konstruere datasæt for DØR analysen, er det først forsøgt at ramme deres analysepopulation fra 2016. Derefter er dataperioden blevet udvidet til 2018 og 2019 efter samme fremgangsmåde.

Populationen dannes på baggrund af alle 29-årige, som i befolkningen 31. december 2015 fra BEF registreret. Efterfølgende sorteres personer ud fra fremgangsmåden beskrevet i DØR (2018). Individer med manglende information omkring uddannelse er frasorteret i analysen. Hvis man i stedet satte disse til at være ufaglærte, ville den endelige population stige med 353 individer. Der er en forskel på 600 individer i analysepopulationen i replikationen sammenlignet med den oprindelige analysen. I DØR analysen er der dog en diskrepans på 300 personer ift. deres beskrevne frasortering og deres estimer. Denne forskel er ikke forklaret yderligere. Sammenlignet med antallet af individer de anvender i deres estimering er forskellen til vores population på 300 personer.

Tilknytningen til arbejdsmarkedet er bestemt ud fra variabelen BESKT13 (kode for personens væsentligste indkomstkilde) i AMR, idet personer med værdier, der dækker over selvstændig og lønmodtagerstatus (01-05), regnes for beskæftigede.

Analysen indeholder forklarende variable opdelt i demografiske, socioøkonomiske og evnemæssige baggrundskarakteristika. De demografiske karakteristika omhandler køn og kommune information som 16-årig. Socioøkonomiske karakteristika er målt, når individet er barn, og omhandler forældres uddannelsesbaggrund, indkomst, forsørgelsesgrad, beskæftigelsesstatus, kriminalitetshistorik, flytninger, anbringelse, antal børn i husstanden, eneforsørgelse, og om faderen er ukendt. Slutteligt omfatter de evnemæssige karakteristika folkeskolekarakterer, hvorvidt individet har gennemført folkeskolen som hhv. 17- og 20-årig, hospitalsindlæggelser, hvorvidt idømt frihedsstraf og erhvervsindkomst som 15- og 16-årig. Alle baggrundsvariable er dannet gennem en række forskellige registre fra Danmarks statistik. I analyser på nye, sammenlagte data kontrolleres desuden for år.

¹⁰ Leuven, E. og B. Sianesi (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. <http://ideas.repec.org/c/boc/bo-code/s432001.html>. Version 2.0.10

Beskæftigelsesgrad i november bestemmes ud fra variablen NOV_TILSTAND_GRAD fra RAS. For hver enkelt person summeres over alle november tilstande, som er klassificeret som "beskæftiget" (SOC_STATUS_KODE 110 til 136).

6.2.2 Empirisk model

Sammenhængen mellem beskæftigelse og karakteristika for henholdsvis faglærte og ufaglærte estimeres på baggrund af probitmodellerne:

$$\begin{aligned} P_{1,i} &: \text{Beskæftigelsesandsynlighed som faglært} = \Phi(\beta_f x_{f,i}) \\ P_{2,i} &: \text{Beskæftigelsesandsynlighed som ufaglært} = \Phi(\beta_{uf} x_{uf,i}) \end{aligned}$$

Hvor x er en vektor af karakteristika, i indekserer individer, og uf og f hhv. ufaglærte og faglærte. I praksis er anvendt *probit* i Stata16.

Herefter beregnes den individspecifikke beskæftigelseseffekt for alle individer som forskellen prædikeret beskæftigelsesandsynlighed på baggrund af parameterestimererne ($\hat{\beta}_f$ og $\hat{\beta}_{uf}$) fra de to estimerede probit modeller:

$$BE_i = \hat{P}_{1,i} - \hat{P}_{2,i} = \Phi(\hat{\beta}_f x_{f,i}) - \Phi(\hat{\beta}_{uf} x_{uf,i})$$

Den gennemsnitlige effekt for forskellige delpopulationer beregnes som gennemsnittet af de individspecifikke effekter for den respektive delpopulation.

For at beregne den forventede effekt af faglært uddannelse, for ufaglærte med estimeret 90 pct. sandsynlighed for at gennemføre en erhvervsuddannelse, såfremt de påbegynder den, estimeres sammenhængen mellem karakteristika og gennemførelse, på baggrund af alle de personer i analysepopulationen, der på et tidspunkt har påbegyndt en erhvervsuddannelse (det vil sige, alle faglærte og den delmængde af ufaglærte, der er påbegyndt og siden frafaldet en erhvervsuddannelse):

$$P_0: \text{Sandsynligheden for at fuldføre en erhvervsuddannelse, hvis den er påbegyndt} = \Phi(\beta_{pab} x_{pab,i})$$

Herefter tages gennemsnittet af BE_i , for ufaglærte med en prædikeret gennemførelsessandsynlighed på over 90 pct. på baggrund af parameterestimererne ($\hat{\beta}_{pab}$) fra denne model.

For at estimere den forventede effekt af faglært uddannelse for hypotetiske ufaglærte med 100 pct. gennemførelsessandsynlighed, bestemmes et tredje gradspolynomium for sammenhængen mellem den individspecifikke estimerede beskæftigelseseffekt og sandsynligheden for gennemføre en erhvervsuddannelse for den ufaglærte delpopulation. Dette gøres ved at regressere de individspecifikke beskæftigelseseffekter for ufaglærte på $\hat{P}_i$, $\hat{P}_i^2$, $\hat{P}_i^3$ og en konstant (estimeret ved OLS).

I udvidelsen af analysen til at omfatte beskæftigelsesgrad anvendes Cragg's Truncated Normal Hurdle Regression modeller (Cragg 1971) frem for en probit modeller i første skridt (dvs. til estimering af de individuelle beskæftigelseseffekter). Denne model er en udvidelse af en Tobit model for hjørne løsninger, hvor en stor del af populationen har observeret 0 som outcome, her timer. I praksis anvendes funktionen *churdle* i Stata16. Analysen følger derudover analysen af effekten på beskæftigelsesfrekvens.

6.3 Søgaard (2011)

6.3.1 Konstruktion af data

Analysepopulationen er baseret på RAS fra 1980 til 2018, og består af 30-59 årige med en gyldig ident. Fra 2008 betinges yderligere på, at man befinder sig i befolkningen på referencetidspunktet (I_BEFOKKNINGEN_KODE=1).

Tilknytning til arbejdsmarkedet i november bestemmes ud fra variablen ARBSTIL, NYARB, SOC_STATUS_KODE eller SOCSTIL_KODE. Der er anvendt samme kategorisering, som anvendes af Danmarks Statistik, jf. nedenfor. Det indebærer at arbejdsstyrken før 2008 defineres som beskæftigede og nettoledige, mens den fra 2008 defineres som summen af beskæftigede og bruttoledige¹¹. I analysen refererer erhvervsdeltagelse til den primære tilknytning til arbejdsmarkedet. Det vil sige, der betinges på PRIMAER_STATUS_KODE=1 eller NOVPRIO=1, hvor relevant.

Beskæftigelsesgrad bestemmes ud fra variablen NOV_TILSTAND_GRAD, som er tilgængelig fra 2008. For hver enkelt person summeres der over alle november tilstande, som er klassificeret som "beskæftiget" (SOC_STATUS_KODE 110 til 136). Det vil sige, for beskæftigelsesgrad tillades PRIMAER_STATUS_KODE=1 eller PRIMAER_STATUS_KODE=3.

Der dannes desuden en indikatorvariabel for modtagelse af orlovsydelse, som antager værdien 1, hvis enten variablen SOC_STATUS_KODE=315 eller NYARB/SOC_STIL_KODE=318. Endelig dannes en indikatorvariabel for modtagelse af overgangsydelse, som antager værdien 1 hvis NYARB/SOC_STIL_KODE=325.

Tabel A.1.: Definition af arbejdsstyrken

	I arbejdsstyrken	Udenfor arbejdsstyrken
ARBSTIL	11-20 (selvstændige og medarbejdende ægtefæller); 31-37 (lønmodtagere) og 40 (arbejdsløse)	60-92 (uden for arbejdsstyrken)
NYARB	11-20 (selvstændige og medarbejdende ægtefæller); 31-37 (lønmodtagere) og 40 (arbejdsløse)	310-331 (uden for arbejdsstyrken)
SOC_STIL_KODE	115-120 (selvstændige og medarbejdende ægtefæller); 130-135 (lønmodtagere); og 200 (nettoledige)	310-335 (uden for arbejdsstyrken)
SOC_STATUS_KODE	110-120 (selvstændige og medarbejdende ægtefæller); 131-136 (lønmodtagere) og 200 (arbejdsløse)	311-517 (uden for arbejdsstyrken)

Fra RAS registret anvendes i øvrigt variablene ALDER_RAS eller ALDER_AMR til at bestemme aldersgruppe, og variablen KOEN til at bestemme køn.

¹¹ Vi har estimeret modellen med en alternativ definition af arbejdsstyrken, hvor vi tilføjer personer, der i RAS er beskrevet som aktiverede (socstil_kode=320, 321, eller 335; nyarb/arbstil=320 eller 321) i et forsøg på at tilnærme os Søgaards definition. Dette påvirker ikke resultaterne væsentligt.

Til hver person tilknyttes oplysninger om herkomst fra registret IEPE, idet der skelnes mellem danskere, indvandrere og efterkommere, på baggrund af Danmarks Statistiks definitioner.

Herudover tilknyttes oplysninger om højest fuldførte uddannelse i hvert år, baseret på nyeste version (2019) af registret P4udd_hfudd, som er en del af BUVMs statistikberedskab. Uddannelser, der er gennemført i december i et givent år henføres til året efter, da arbejdsmarkedstilknytningen måles i november. Uddannelser klassificeres som enten grundskole, gymnasium, erhvervsfaglig, kort videregående, mellemlang videregående eller lang videregående, baseret på uddannelsesklassifikationen DI-SCED, jf. tabellen nedenfor. Manglende uddannelsesoplysninger slås sammen med kategorien "Grundskole"¹².

Tabel A.2: Klassifikation af uddannelser

Uddannelsesgruppe	DISCED koder (overgruppe)
Grundskole	10 (Grundskole) eller uoplyst
Gymnasium	20 (Gymnasiale uddannelser); 35 (Adgangsgivende uddannelsesforløb)
Erhvervsfaglig	30 (Erhvervsfaglige uddannelser)
Kort videregående	40 (Korte videregående uddannelser)
Mellemlang videregående	50 (Mellemlange videregående uddannelser); 60 (Bacheloruddannelser)
Lang videregående	70 (Lange videregående uddannelser); 80 (Ph.D. og forskeruddannelser).

6.3.2 Empirisk model

Direkte estimering

Ved den direkte estimering anvendes data, som er aggregeret til køn-alders-niveau (fx 30 årige mænd), idet personer af anden etnisk herkomst end dansk udelades (dvs., der betinges på IE_TYPE=1). Analysen anvender regressionsligningen:

$$dFaktisk_{it} = \beta_0 + \beta_1 dHypotetisk_{it} + \beta_2 dKontrol_{it} + \beta_3 D_{alder, køn} + \beta_4 D_{år, køn} + \beta_5 dOrlov_{it} + \beta_6 dOvergang_{it} + \mu_{i,t}$$

Hvor i indekserer grupper defineret ved køn og alder, t indekserer år, og:

$dFaktisk_{it}$ er beregnet som ændringen i gruppe i 's erhvervsfrekvens mellem år $t - 1$ og år t .

$$dFaktisk_{it} = Erhvervsfrekvens_{it} - Erhvervsfrekvens_{i,t-1}$$

Det vil sige, eksempelvis, forskellen mellem erhvervsfrekvensen for 40-årige mænd i år 2000 og erhvervsfrekvensen for 40-årige mænd i år 1999.

¹² Vi har forsøgt med to andre måder at håndtere uoplyst uddannelse: (1) vi har behandlet det som en selvstændig uddannelse kategori; og (2) vi har udeladt personerne fra analysen. (1) resulterer i gennemslag på over 1 for de ældste aldersgrupper og under 0,1 for hele gruppen af 30-59 årige; (2) resulterer i et gennemslag på over 0,9 for alle aldersgrupper.

$dHypotetisk_{it}$ er beregnet som den ændring i gruppe i 's erhvervsfrekvens der ville være fremkommet mellem år $t - 1$ og år t , såfremt de uddannelsesspecifikke erhvervsfrekvenser var uændrede, og det alene var uddannelsessammensætningen, der havde ændret sig. Altså, idet g indekserer uddannelser, og $N_{i,g,t}$ er antal personer i gruppe i med uddannelse g til tidspunkt t :

$$dHypotetisk_{it} = \sum_g (Erhvervsfrekvens_{i,t-1,g} \times p_{itg}) - Erhvervsfrekvens_{i,t-1}, \text{ hvor } p_{itg} = \frac{itg}{\sum_g N_{itg}}$$

Det vil sige, eksempelvis, forskellen mellem den erhvervsfrekvens 40-årige mænd *ville have haft* i år 2000, hvis de havde samme uddannelsesspecifikke erhvervsfrekvenser som 40-årige i år 1999, men et andet uddannelsesniveau; og *den faktiske* erhvervsfrekvens for 40-årige mænd i år 1999. Koefficienten på $dHypotetisk$ fortolkes i Søgaard (2011) som gennemslaget af uddannelse på den faktisk erhvervsfrekvens.

$dKontrol_{it}$ beregnes som forskellen mellem erhvervsfrekvensen for gruppe i i år t , og den et år yngre gruppe i år $t - 1$. Eller med andre ord, ændringen i erhvervsfrekvensen for kohorte k , defineret ved køn og fødselsår, mellem år $t - 1$ og år t :

$$dKontrol_{it} = Erhvervsfrekvens_{kt} - Erhvervsfrekvens_{k,t-1}$$

$D_{alder,køn}$ er alder-køn-dummyer

$D_{år,køn}$ er år-køn-dummyer

$dOvergang_{it}$ er ændringen fra år $t - 1$ til år t i andelen af gruppe i , som er på overgangsydelse. Dvs.:

$$dOvergang_{it} = Overgang_{it} - Overgang_{i,t-1}$$

$dOrlov_{it}$ er ændringen fra år $t - 1$ til år t i andelen af gruppe i , som er på arbejdsmarkedsorlov. Dvs.:

$$dOrlov_{it} = Orlov_{it} - Orlov_{i,t-1}$$

Indirekte estimering

Ved den indirekte estimering anvendes data, som er aggregeret til køn-alders-uddannelses-niveau (fx 30-årige mænd med erhvervsuddannelse). Personer af anden etnisk herkomst end dansk udelades fra estimeringen af udtyndingseffekten.

Udtyndingseffekten estimeres ved hjælp af følgende regressionsligning:

$$dFaktisk_{itg} = \alpha_0 + \alpha_1 dGE_{itg} + \alpha_2 dKontrol_{itg} + \alpha_3 D_{alder,køn,uddannelse} + \alpha_4 D_{år,køn} + \alpha_5 dOrlov_{itg} + \alpha_6 dOvergang_{itg} + \varepsilon_{itg}$$

Hvor i indekserer grupper defineret ved køn og alder, t indekserer år, g indekserer uddannelsesgrupper og:

dGE_{itu} er ændringen fra år $t - 1$ til år t i evneindekset for gruppe iu (fx 30-årige mænd med en erhvervsuddannelse). Det vil sige,

$$dGE_{itg} = GE_{itg} - GE_{i,t-1,g}$$

Resten af variablene er defineret som før med den ændring, at der nu er tale om ændringer inden for alder-køn-uddannelsesgrupper.

Evne-indekset er beregnet som:

$$GE_{itg} = \ln(p_{itg}^{lavere} + 0,5 \times p_{itg})$$

Hvor p_{itg} angiver andelen af køn-alders-gruppen, der har uddannelse g til tidspunkt t ; og p_{itg}^{lavere} angiver andelen af befolkningen med et lavere uddannelsesniveau. Det vil sige, p_{itg} kunne eksempelvis være andelen af 30-årige mænd med en erhvervsuddannelse i år 1990, og p_{itg}^{lavere} ville da være andelen

af befolkningen med enten grundskole eller en gymnasial uddannelse som højeste fuldførte uddannelse.

På baggrund af estimatet for udtyndingseffekten estimeres efterfølgende gennemslaget af uddannelsesfremgang på erhvervsfrekvensen. Dette gøres ved først, for grupper med dansk herkomst, at fremskrive uddannelsesspecifikke erhvervsfrekvenser fra et basisår, idet disse ændrer sig som følge af udtynding. Det vil sige, der beregnes:

$$Erhvervsfrekvens_{itg} = Erhvervsfrekvens_{i,t=0,g} + \hat{\alpha}_2(GE_{itg} - GE_{i,t=0,g})$$

Hvor $\hat{\alpha}_2$ er estimatet fra de indirekte estimationer. Basisåret er 1985 i de oprindelige estimationer og i replikationen. I opdateringen er basisåret 1995, når hele perioden 1980 til 2019 betragtes, og 2009 når perioden 2008 til 2019 betragtes.

Herefter fremskrives befolkningens samlede erhvervsfrekvens fra basisåret tre gange. I disse fremskrivninger inkluderes personer af anden etnisk herkomst end dansk, men for disse personer differentieres der ikke mellem forskellige uddannelsesniveauer. De tre gange kan beskrives således:

- (1) Hvor alder-køn-herkomst-uddannelsesspecifikke erhvervsfrekvenser holdes konstante på niveauet fra basisåret, og kun befolkningsandele af alder-køn-herkomst-uddannelsesgrupper ændrer sig. Dette benævnes den proportionale uddannelsesfremskrivning, $fremskrevet_t^{proportional}$
- (2) Hvor alder-køn-herkomst-uddannelsesspecifikke erhvervsfrekvenser ændrer sig som følge af udtynding (dvs. den beregnede $Erhvervsfrekvens_{itg}$ benyttes) og befolkningsandele af alder-køn-herkomst-uddannelsesgrupper ændrer sig. For personer af udenlandsk herkomst holdes erhvervsfrekvensen dog konstant på niveauet fra basisåret. Dette benævnes fremskrivning inklusiv udtynding, $fremskrevet_t^{udtynding}$
- (3) Hvor alder-køn-herkomst specifikke erhvervsfrekvenser holdes konstante på niveauet fra basisåret og kun befolkningsandele af alder-køn-herkomst-grupper ændrer sig. Dette benævnes den demografiske fremskrivning, $fremskrevet_t^{demografisk}$

Gennemslaget findes herefter som:

$$gennemslag_t = \frac{fremskrevet_t^{udtynding} - fremskrevet_t^{demografisk}}{fremskrevet_t^{proportional} - fremskrevet_t^{demografisk}}$$

I den oprindelige analyse og replikationen evalueres gennemslaget i år 2004. I opdateringen evalueres det i år 2018.



BØRNE- OG
UNDERVISNINGSMINISTERIET