

## Reestimation af uddannelsessøgende til modelversion Okt18

### Resumé:

*Relationen for uddannelsessøgende reestimeres til modelversion Okt18. Siden seneste reestimation er data blevet revideret for hele perioden, hvilket giver et niveauskifte før 2008, og en lidt anden udvikling efter 2008. Det har givet anledning til, at den ene logistiske trend er blevet rykket ét år tilbage, således at vendepunktet nu ligger i 2009 frem for i 2010.*

*Sammenlignet med estimationen til Okt16 er der imidlertid ikke sket meget med estimationsresultatet. Parameterestimatet til ændringen i ledigheden er steget lidt, mens selve estimationen klarer sig bedre målt på  $R^2$  og regressionens standardfejl. Fit og residualer viser, at relationen har fået lidt sværere ved at ramme den faktiske udvikling i uddannelsessøgende efter 2010.*

*Der er desuden foretaget en dekomponering af relationens forklarende variables bidrag til forklaringsgraden.*

---

BGS21n18 - Reestimation af uddannelsessøgende til modelversion Okt18

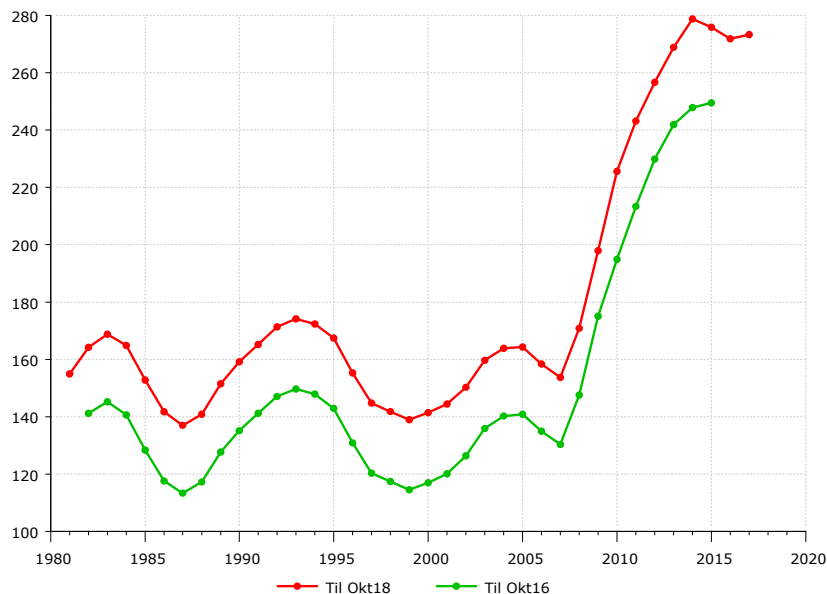
Nøgleord: Reestimation, uddannelsessøgende, Okt18

*Modelgruppepapirer er interne arbejdsrapporter. De konklusioner, der drages i papirerne, er ikke endelige og kan være ændret inden opstillingen af nye modelversioner. Det henstilles derfor, at der kun citeres fra modelgruppepapirerne efter aftale med Danmarks Statistik.*

## Dataændringer

Siden seneste reestimation af relationen for uddannelsessøgende er data for uddannelsessøgende uden for arbejdsstyrken, *Uuxa*, blevet revideret. Kilden til tallene er RAS-statistikken (Registerbaseret Arbejdsstyrke), og i sommeren 2017 blev denne statistik revideret for perioden 2008-2016 pga. en metodeomlægning. I den forbindelse blev uddannelsessøgende fordelt på flere forskellige grupper af uddannelsessøgende bl.a. personer under uddannelse (ordinær), kursister og produktionsskoleelever. Til brug i ADAM er disse grupper lagt sammen til den velkendte variabel *Uuxa*. Ændringen har givet et højere niveau i antallet af uddannelsessøgende. Før 2008 er der, pga. tilbageføring vha. kædning i 2008, kun sket et niveauskifte. Efter 2008 har udviklingen i data ændret sig, jf. figur 1.

**Figur 1: Antallet af uddannelsessøgende, 1000 pers.**



Da der estimeres en relation for *ændringen* i andelen af uddannelsessøgende blandt 15-34 årige, vil niveauskiftet tilbage i tid ikke få den store betydning for estimationsresultatet. Udviklingen efter 2008 kan derimod godt få betydning.

## Estimation af uddannelsessøgende

### Relationen

Relationen for uddannelsessøgende var i ADAM Okt16 givet ved:

$$\begin{aligned} \text{dif}\left(\frac{Uuxa}{U1534}\right) &= a_1 \cdot \frac{2 \cdot \text{dif}\left(\frac{Ulb}{U1564}\right) + \text{dif}\left(\frac{Ulb_{-1}}{U1564_{-1}}\right)}{3} \\ &+ a_2 \cdot \text{dif}(dtuuxa01) + a_3 \cdot \text{dif}(dtuuxa10) \end{aligned} \quad (1)$$

<i>Uuxa:</i>	<i>Antallet af uddannelsessøgende uden for arbejdsstyrken</i>
<i>U1534:</i>	<i>Befolkning i aldersgruppen 15-34 år</i>
<i>Ulb:</i>	<i>Bruttoledighed i alt</i>
<i>U1564:</i>	<i>Befolkning i aldersgruppen 15-64 år.</i>
<i>dtuuxa01:</i>	<i>Logistisk trend med vendepunkt i 2001</i>
<i>dtuuxa10:</i>	<i>Logistisk trend med vendepunkt i 2010</i>

I denne relation er det altså primært ændringen i andelen af bruttoledige blandt de 15-64 årige, som driver ændringen i andelen af uddannelsessøgende blandt 15-34 årige. Dette suppleres af to logistiske trends med vendepunkt i hhv. 2001 og 2010.

Sidstnævnte trend blev indført i forbindelse med reestimationen til modelversion Okt15 for at fange den markante stigning i andelen af uddannelsessøgende blandt de 15-34 årige efter 2007, jf. figur 1.

I forbindelse med reestimationen til modelversion Okt18 er det valgt at rykke den logistiske trend med vendepunkt i 2010 et år tilbage, således at den nu har vendepunkt i 2009. Det sker pga. revisionen af tallene for uddannelsessøgende, som har gjort, at andelen af uddannelsessøgende stiger lidt kraftigere frem til 2014, og derefter vender mere skarpt.

Relationen for uddannelsessøgende til Okt18 er derfor givet ved:

$$\begin{aligned} \text{dif}\left(\frac{Uuxa}{U1534}\right) &= a_1 \cdot \frac{2 \cdot \text{dif}\left(\frac{Ulb}{U1564}\right) + \text{dif}\left(\frac{Ulb_{-1}}{U1564_{-1}}\right)}{3} \\ &+ a_2 \cdot \text{dif}(dtuuxa01) + a_3 \cdot \text{dif}(dtuuxa09) \end{aligned} \quad (2)$$

<i>Uuxa:</i>	<i>Antallet af uddannelsessøgende uden for arbejdsstyrken</i>
<i>U1534:</i>	<i>Befolkning i aldersgruppen 15-34 år</i>
<i>Ulb:</i>	<i>Bruttoledighed i alt</i>
<i>U1564:</i>	<i>Befolkning i aldersgruppen 15-64 år.</i>
<i>dtuuxa01:</i>	<i>Logistisk trend med vendepunkt i 2001</i>
<i>dtuuxa09:</i>	<i>Logistisk trend med vendepunkt i 2009</i>

## Estimationsresultat

Ændringen af vendetangenten giver et bedre estimationsresultat målt på  $R^2$  og regressionens standardfejl, jf. tabel 1 kolonne (a) og (b). Derudover bliver parametrene til de to logistiske trends mindre, når den sidste trend har vendepunkt i 2009, mens parameteren til ændringen i bruttoledigheden bliver større. Det virker altså til, at den primære forklarende variabel får en større rolle, mens de to logistiske trends betyder mindre, hvis vendetangenten rykkes.

**Tabel 1: Estimationsresultat, uddannelsessøgende**

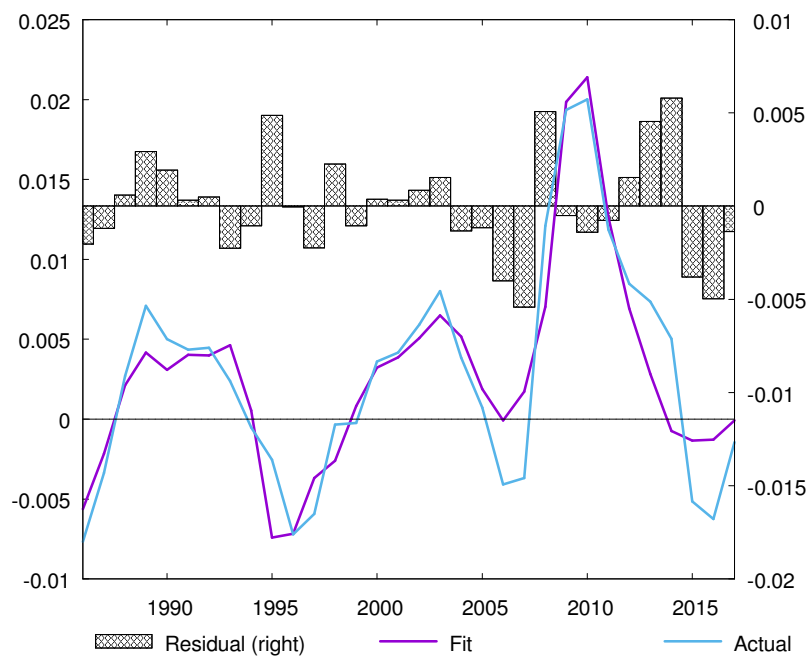
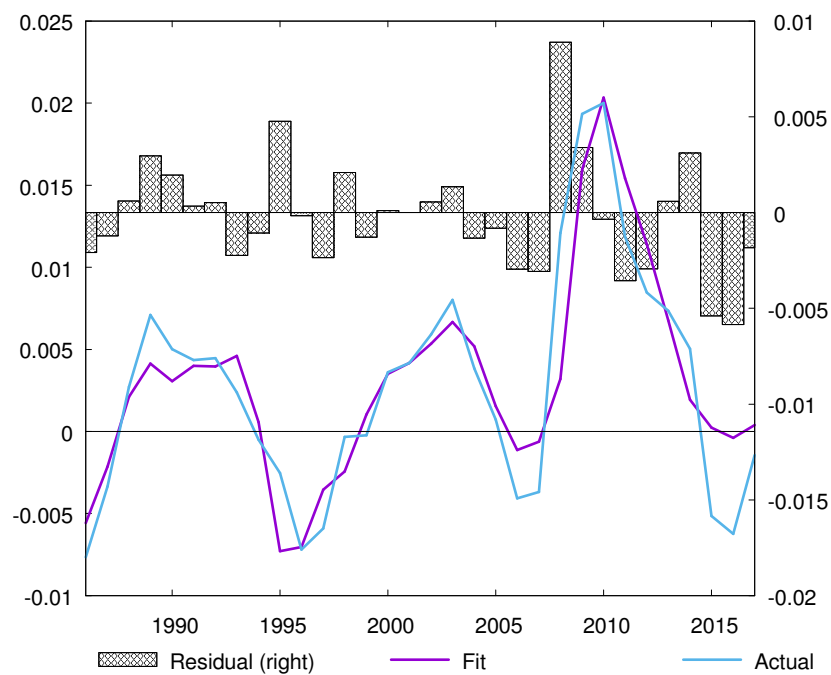
	(a)		(b)		(c)	
Forklaret variabel	Estimat	SE	Estimat	SE	Estimat	SE
Uddannelsessøgende Dlog(Uuxa/U1534)	Til Okt18 – dtuuxa09		Til Okt18 – dtuxxa10		Til Okt16	
1. a1 dif(Ulb/U1564)-led	0.543840	0.072787	0.539110	0.076666	0.504215	0.076951
2. a2 Dif(dtuxxa01)	0.037535	0.009336	0.039632	0.009803	0.034478	0.009396
3. a3 Dif(dtuxxa10)			0.084185	0.008742	0.088091	0.009474
4. a3 Dif(dtuxxa09)	0.080058	0.007790				
Loglikelihood $R^2 \setminus$ Std.fejl. for reg. Periode	135.414 0.850116/0.002795 1986 – 2015		133.883 0.833174/0.002941 1986 – 2015		125.269 0.826776/0.002920 1986 – 2013	

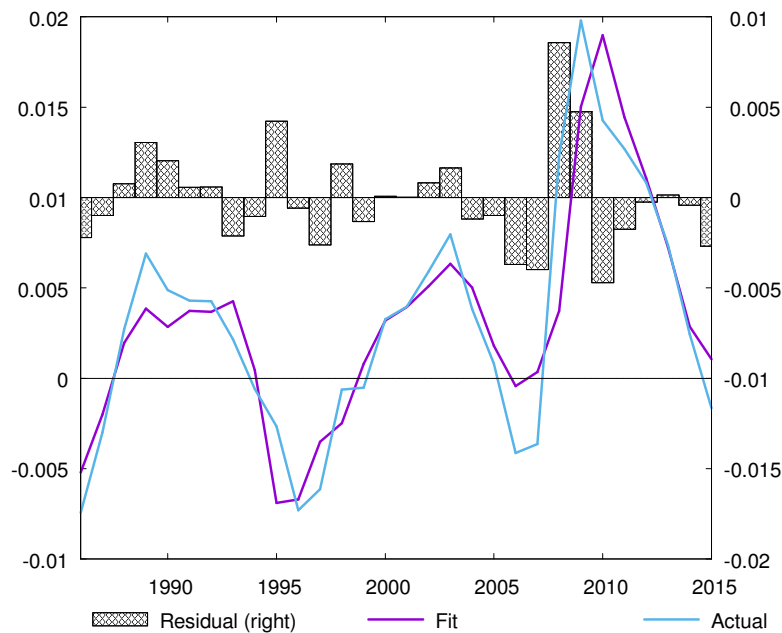
Estimationsresultatet til Okt16 er for sammenligningens skyld præsenteret i kolonne (c) i tabel 1. Fra estimationen til Okt16 til den nye estimation til Okt18 er estimationsperioden udvidet med 2 år, så der nu estimeres til og med 2015. Sammenlignes resultatet fra Okt16 med estimationsresultatet til Okt18 i kolonne (a) bemærkes først og fremmest, at estimationen klarer sig bedre nu end tidligere. Det gælder både ift.  $R^2$  og regressionens standardfejl. Derudover har ledighedsvariablen fået en større rolle i relationen, idet parameteren til denne variabel er steget.

## Fit og residualer

Sammenlignes fit og residualer for estimation (a) og (b), figur 2 og figur 3, kan man se effekten af at vendetangenten rykkes et enkelt år. Med vendepunkt i 2009 fittes bedre i årene lige omkring toppen i 2010, men lidt dårligere lige efter toppunktet. Hen mod slutningen af samplet reduceres residualerne en smule ved at vendetangenten ligger i 2009 i stedet for 2010.

Sammenlignes fit og residualer fra Okt16, figur 4, med fit og residualer fra estimationen til Okt18, figur 2, bemærkes at fittet er blevet dårligere i perioden efter 2010. Det skyldes primært dataændringerne, som har fået andelen af uddannelsessøgende til at falde mere i 2015 end tilfældet var da der blev estimeret til Okt16.

**Figur 2: Fit og residualer, estimation til Okt18, vendetangent i 2009****Figur 3: Fit og residualer, estimation til Okt18, vendetangent i 2010**

**Figur 4: Fit og residualer, estimation til Okt16, vendetangent i 2010**

## Dekomponering af relationen for uddannelsessøgende

Relationen for uddannelsessøgende kan dekomponeres, så det bliver klart, hvor meget hvert enkelt element i relationen bidrager med i forhold til at ramme den faktiske udvikling i antallet af uddannelsessøgende.

De 4 forklarende variable er konjunktur, demografi og de to logistiske trends med vendetangent i hhv. 2001 og 2009.

Udgangspunktet for dekomponeringen er 1985. Her fra måles den faktiske udvikling i antallet af uddannelsessøgende. I sidste halvdel af 1980'erne er der således lidt færre uddannelsessøgende end i 1985, jf. figur 5. I slutningen af samplet er antallet af uddannelsessøgende steget med ca. 120.000 personer ift. 1985-niveauet.

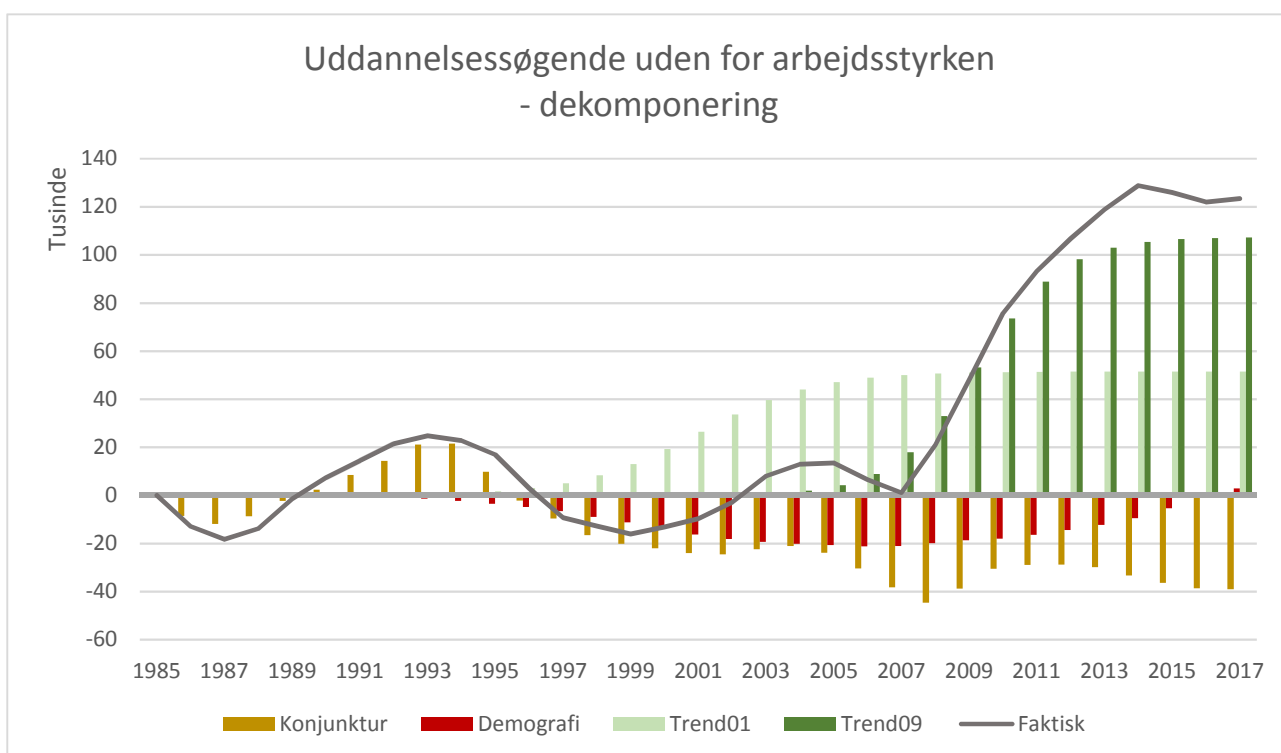
Frem til midten af 1990'erne er det primært konjunktur, som bidrager til en forklaring på udviklingen i uddannelsessøgende, jf. figur 5. Der er meget lidt bidrag fra den demografiske udvikling i denne periode, og de to trends er endnu ikke i brug.

Fra midten af 1990'erne og samplet ud er der et negativt bidrag fra konjunktur og demografi, jf. figur 5. Det kommer sig af, at ledighedsgraden og antallet af 15-34 årige ligger lavere i hele denne periode sammenlignet med 1985. Der er dog stadig fluktuationer i udviklingen af de to, og bl.a. bidraget fra den lave ledighedsgrad op mod finanskrisen ses tydeligt i 2008. Her giver konjunktursituationen, at antallet af uddannelsessøgende burde være mere end 40.000 personer mindre end i 1985.

Idet antallet af uddannelsessøgende fra 2002 og frem ligger over 1985-niveauet er der brug for et supplement til demografi og konjunktur, som de forklarende variable, da de som nævnt trækker i den modsatte retning i hele denne periode.

Dette bliver klaret ved de to logistiske trends. Den første trend er lagt ind for at bidrage til at forklare, hvad der ligner et generelt skifte i uddannelsesnivea- og længde fra årtusindeskiftet og frem. Den anden trend hjælper til at forklare den store stigning i antallet af uddannelsessøgende, som begyndte omkring 2007. Denne stigning kan have flere forklaringer. Et bud er et øget optag af internationale studerende. Et andet bud er, at der igen i denne periode er kommet et øget fokus på uddannelse. Et tredje bud er, at datakvaliteten i netop denne periode bliver forbedret, og derfor indgår der løbende flere grupper af studerende i statistikken fra 2008 og nogle år frem.

**Figur 5: Dekomponering af relationen for uddannelsessøgende**



## Konklusion

Relationen for antallet af uddannelsessøgende uden for arbejdsstyrken,  $U_{uxa}$ , er blevet reestimeret til modelversion Okt18. Siden seneste reestimation til modelversion Okt16 er datakilden, RAS-statistikken, blevet revideret, hvilket især har betydning for udviklingen i antallet af uddannelsessøgende efter 2008.

Denne dataændring har givet anledning til, at den ene af relationens to logistiske trends har fået ændret vendepunkt fra 2010 til 2009. Derudover er estimationsperioden blevet udvidet med 2 år.

Reestimationen til Okt18 giver en større parameter til den forklarende variabel for ændringen i ledigheden samt en mindre rolle for de to logistiske trends.

Desuden fås et bedre resultat for selve estimationen målt på  $R^2$  og regressionens standardfejl.

Fit og residualer viser, at ligningen rammer lidt bedre omkring 2008-2010, men har fået sværere ved at forklare de seneste års udvikling i andelen af uddannelsessøgende.