

Notat om den månedlige arbejdskraftundersøgelse

Indhold

1. Produktionen af månedlige AKU-tal	3
2. Den valgte metode: Det glidende gennemsnit.....	5
3. Publicering af data, revisioner og usikkerhed.....	6
4. Overvejelser i forbindelse med udgivelsen af månedstal	8

Sammenfatning

Dette notat giver en introduktion til og beskrivelse af den månedlige udgivelse af Arbejdskraftundersøgelsen (AKU), der første gang publiceres i oktober 2013. Måned-AKU'en udspringer af den allerede eksisterende AKU-kvartalsundersøgelse, der har eksisteret siden 1994.¹

Notatet vil gennemgå baggrunden for produktionen af månedstal i AKU, og hvilke serier, der bliver publiceret. Der vil være et særligt fokus på valget af et *tre måneders glidende gennemsnit med forecast*, som grundlag for månedstallet. Produktionen af månedstallene i AKU vil blive præsenteret ved at bryde den ned i enkeltdele; dvs. stikprøvens størrelse og indsamling heraf, den specifikke månedsopregning, forecast-metoden og dannelse af det glidende gennemsnit.

Valget af denne metode skyldes, at produktion af rene månedsestimater viste sig ikke at være holdbart, bl.a. grundet en systematik i tallene samt at det viste sig, at sæsonkorrektionen af de rene månedsestimater ikke var muligt. Den valgte metode får konsekvenser for hvilke tal, der bør sammenlignes, hvilket vil blive præsenteret efterfølgende i afsnit 2.

Efterfølgende vil den valgte publiceringscyklus blive præsenteret. Forecastet er en prognose, der giver et kvalificeret bud på den måned, der endnu ikke er indsamlet af de tre måneder i gennemsnittet. I næste måned vil den måned, der var dannet ved forecast, derimod være indsamlet. Anden gang et månedsgennemsnit publiceres, er forecastet for den tredje måned således erstattet med indsamlede data. Yderligere vil tallene også blive 'benchmarket' op mod kvartalstallene. Dette beskrives i afsnit 3.

Afslutningsvis vil afsnit 4 præsentere de til- og fravalg, der er gjort i løbet af processen. Det er dette kapitel, der går i dybden med, hvorfor et rent månedstal blev fravalgt og tre måneders glidende gennemsnit blev valgt. Tillige vil valget af forecast frem for tre fuldt indsamlede måneder blive beskrevet.

¹ AKU eksisterede som årlig undersøgelse fra 1983 frem til 1994. I Statistikbanken forefindes tal tilbage til 1996 på udvalgte serier.

Baggrund

Det europæiske statistikbureau, Eurostat, har i flere år publiceret månedstal for de lande, der indgår i den fælleseuropæiske arbejdskraftundersøgelse, Labour Force Survey (LFS). Månedstallene for Danmark har således ikke været produceret af Danmarks Statistik, men er dannet af Eurostat på baggrund af den kvartalsvise arbejdskraftundersøgelse, som Eurostat har fremskrevet med udviklingen i den registerbaserede ledighed. Eurostat har på baggrund af de publicerede kvartalstal i Arbejdskraftundersøgelsen revideret månedstal, så de svarer til kvartalsudviklingen. Idet der er relativt store forskelle mellem den registerbaserede ledighedsstatistik og AKU, både hvad angår populationen, definitioner og sæsonudviklingen, har revisionerne af Eurostats månedstal ofte været meget markante.

Der har derfor været et ønske hos både Eurostat og Danmarks Statistik om, at Danmarks Statistik selv påbegynder produktionen af egne månedstal.

Danmarks Statistik vil fremadrettet publicere to månedlige ledighedsstatistikker; den registerbaserede ledighedsstatistik og måneds-AKU'en. Grundet de nævnte forskelle mellem de to statistikker omkring population, definition og sæsonudvikling, er både niveauet såvel som udviklingen for de to statistikker forskellig.² De udkommer i en fælles nyt den sidste torsdag i den efterfølgende måned.

Hvilke serier publiceres fra månedstallene

Månedstallene i AKU vil blive publiceret med de opdelinger, som Eurostat allerede publicerer. Disse ses nedenfor i tabel 1.

Tabel 1: Serier, som månedstallene bliver publiceret på.

	Mænd	Kvinder	Mænd	Kvinder
	15-24 år		25-74 år	
AKU-ledige	X	X	X	X
Beskæftigede	X	X	X	X

Disse dataserier er opgjort tilbage til og med januar 2007, og tallene bliver således også revideret i Eurostats databaser tilbage til 2007.

² For yderligere information om forskelle mellem AKU og den registerbaserede ledighedsstatistik se [her](#)

1. Produktionen af månedlige AKU-tal

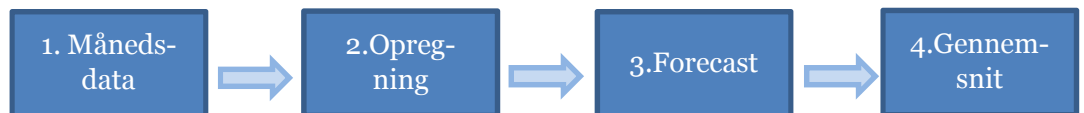
AKU er grundlæggende en kvartalsstatistik, hvilket betyder at produktionssystemet fortsat vil være optimeret til en kvartalsundersøgelse. Månedstatistikken skal ses som et supplement til kvartalstallene, og vil blive udgivet på udvalgte, aggregerede serier. Respondenter, der deltager i den månedlige AKU, vil således også være datagrundlag for den kvartalsvise AKU.

Ca. 22.000 interview gennemføres i AKU pr. kvartal. På månedsbasis er der således tale om ca. 7.500 interviews.

I det følgende afsnit beskrives produktionen af månedlige AKU-tal. Det gennemgås, hvordan Danmarks Statistik producerer ledigheds- og beskæftigelsestal fra Arbejdskraftundersøgelsen; fra indsamlingen af stikprøven til at tallene kan forefindes i statistikbanken. For eksemplificerings skyld er produktionen af september 2013 valgt som eksempel.

Det månedlige AKU-tal er baseret på et *glidende tre-måneders gennemsnit*, der i første omgang udkommer på baggrund af følgende fire trin.

Figur 1: Produktionen af månedstal



Måneddata

Stikprøven bliver indsamlet månedligt efter strukturen 4 uger – 4 uger – 5 uger, således at den seneste måned i kvartalet altid indeholder flere interviews end de to foregående måneder. Dette bliver der dog taget højde for i opregningen af månedstallene, således at respondenterne i vægtningen er fordelt ligeligt på de tre måneder i kvartalet.

Opregning

AKU opererer med en opregning, hvor de indkomne surveyresultater vægtes inden de publiceres, så de angiver resultater for hele befolkningen i alderen 15-74 år. Fordi der i Danmark er et relativt stort antal, der ikke svarer i undersøgelsen, er den kvartalsvise opregning rimelig avanceret, således at undersøgelsen bliver så repræsentativ som muligt. Der inddrages registerinformation som hjælpeinformation på områder som bl.a. alder, køn, region, uddannelse og socioøkonomisk status. Derudover inddrages den registerbaserede ledighed også i opregningen.

Den månedlige opregning er en simplificering af den kvartalsvise opregning, idet omfanget af hjælpeinformation såvel som grupperingen af hjælpevariable er reduceret. Det har været nødvendigt at simplificere modellen, fordi den månedlige stikprøve er væsentlig mindre end den kvartalsvise. Ved en for kompleks model og en lille stikprøve, vil modellen bryde sammen.

Idet vi har at gøre med et glidende gennemsnit og i eksemplet skal publicere september måned, vil det indsamlede data over august og september måned

blive opregnet til befolkningen. Modellen for månedsopregningen ses nedenfor i tabel 2.

Tabel 2 – Opregningen i den månedlige AKU

	Variabler	Grupperinger
Informationen krydses	-alder	6 grupper
	-køn	2 grupper
Informationen krydses	-alder	6 grupper
	-uddannelse	3 grupper
	-socioøkonomisk status	7 grupper
	-statsborgerskab	3 grupper
	-nettoledighed	4 grupper
	-bruttoindkomst	4 grupper

Forecast

Det næste led i produktionen vedrører appliceringen af et forecast af den tredje måned i det glidende gennemsnit – i dette tilfælde oktober måned. Når september måned er indsamlet bruges hele månedstidsserien fra 2007 og frem til at give en prognose af den kommende måneds estimater. Der forecastes på de fire underserier (køn og aldersgrupperne 15-24 og 25-74 år) for både ledige og beskæftigede. Forecast metoden er valgt for at gøre tallene aktuelle. Alternativt skulle man have indsamlet den sidste måned i gennemsnittet – oktober måned – hvilket ville betyde, at man først kunne udgive tallene en måned senere.

Glidende gennemsnit

Et tre måneders glidende gennemsnit kan herefter dannes på baggrund af de første to indsamlede måneder og den sidste forecastede måned, hvor data endnu ikke er indsamlet. Måneden, der publiceres på, er den midterste måned af de tre måneder, der danner gennemsnittet. Styrken ved denne model er, at månedstallet indeholder den seneste udvikling, uden at dette går ud over aktualiteten af tallene.

Figur 2: Det glidende gennemsnit med forecast

Tal for:	Gennemsnit af:		
September	August	September	Oktober
	Indsamlede data	Indsamlede data	Forecast

2. Den valgte metode: Det glidende gennemsnit

At benytte et glidende gennemsnit med forecast har dog nogle konsekvenser for, hvordan tallene bør anvendes og tolkes.

Hvordan tolkes et glidende gennemsnit?

Det glidende gennemsnit som metode betyder, at sammenligning af månedstal fra en måned til den næste er problematisk. Det skyldes, at et månedstal baseret på tre måneders glidende gennemsnit delvist har en fælles datamængde med månedstallene for de to måneder før og de to måneder efter den pågældende måned. Udviklinger mellem hinanden følgende måneder bliver derved vanskelige at fortolke. Ændringer i ledigheden mellem eksempelvis august og september vil således ikke nødvendigvis være et udtryk for reelle udviklinger, fordi det ikke er til at sige, om den ændrede ledighed skyldes, at juli ikke længere udgør en tredjedel af gennemsnittet, eller om det skyldes, at oktober nu indgår i gennemsnittet.³

Det anbefales derfor, at man går tre måneder tilbage, hvis man vil sammenligne udviklinger i serierne, for at finde det sidste datapunkt, der ikke har nogen fælles datamængde med den aktuelle måned. Dette illustreres i figur 3 nedenfor.

Figur 3: Data i det glidende gennemsnit

Tal	Gennemsnit					
Sep	Aug	Sep	Okt			
Okt		Sep	Okt	Nov		
Nov			Okt	Nov	Dec	
Dec				Nov	Dec	Forecast

Udvikling

Ved produktionen af månedstallet for september vil august, september og oktober indgå i det glidende gennemsnit. Som figuren illustrerer, har september måned en fælles datamængde med de opgjorte gennemsnit for månederne oktober og november. Data for måneden oktober er således en del af det glidende gennemsnit for alle tre måneder, hvilket gør sammenligninger mellem disse måneder vanskelig at tolke. Først når tallet for december dannes, får serien et nyt datapunkt, der ikke deler datamængde med september. Det er med dette spænd på tre måneder, at Danmarks Statistik anbefaler at kommentere på tallene.

Sæsonkorrektion

Det glidende gennemsnit bliver til sidst sæsonkorrigeret. Ved månedstallene er der tale om en indirekte sæsonkorrektion. Det vil sige, at underserierne sæsonkorrigeres mens de aggregerede hovedserier dannes ved at summere underserierne (beskæftigede og AKU-ledige fordelt på køn og på de to aldersgrupper 15-24 år og 25-74 år). Sæsonkorrektionen kommer ind som det sidste element,

³ Ydermere vil estimater på enkelte måneder også betyde, at man netop foretager sin vurdering på baggrund af 1/3 af det glidende gennemsnit. Dermed rammer man ind i de problemer, der er i det rene månedstal, den høje usikkerhed og den problematiske systematik, som vil blive uddybet i afsnit 4.

således at både forecastet og det glidende gennemsnit er dannet, når underserierne sæsonkorrigeres.

3. Publicering af data, revisioner og usikkerhed

Foreløbige og indsamlede tal

Det glidende gennemsnit med forecast betyder, at der løbende vil være et parallelt forløb i de tal, der publiceres fra den månedlige AKU. Tallet for september, der publiceres ultimo oktober er *foreløbigt tal*, da den sidste måned, der indgår i gennemsnitsberegningen, oktober, ikke er indsamlet, men er et forecast. Forecastet er en prognose, og denne prognose kan vise sig at tage fejl i større eller mindre grad. Når oktober måned er indsamlet publiceres et nyt tal for september. Dette tal vil først blive offentliggjort ultimo november i Statistikbanken, hvor tidsserien vil blive revideret på baggrund af de indsamlede tal.

Revision ift. kvartalstallet

For at opnå størst mulig sammenlignelighed med de månedstal, som Eurostat publicerer for Danmark, vil der ved afsluttede kvartaler forekomme en revision af månedstallene, hvor månedstallene 'benchmarkes' op mod kvartalstallene.

Dette betyder, at eksempelvis månedstallene for juli, august og september vil blive benchmarket op mod kvartalstallene for 3.kvartal på de relevante serier. Dette sker ved, at der korrigeres for differencen mellem månedstal og kvartalstal, så månedstallene kommer til at stemme overens med kvartalstallene. Tallene for 3.kvartal vil blive publiceret ultimo november måned. Sæsonkorrigerede serier vil dog ændre sig, idet man skal tre år tilbage plus indeværende år for at finde faste tal, der ikke længere påvirkes af sæsonkorrektionen. For de ikke-sæsonkorrigerede serier, kan det ses i tabel 3, hvornår der publiceres.

Tabel 3: Publicering af månedstal

Tal for	Publiceres 1.gang - inklusiv forecast	Publiceres 2.gang - data er indsamlet	Publiceres 3.gang – revision ift. kvartalstal
September	Oktober	November	November
Oktober	November	December	Februar
November	December	Januar	Februar
December	Januar	Februar	Februar

Hvis september tages som et eksempel, så vil september måned første gang blive publiceret i oktober måned. Her vil den sidste måned (oktober) være et forecast. Tallet vil således være foreløbigt. Når data er indsamlet for den sidste måned (oktober) i det tre-måneders glidende gennemsnit, der danner september tallet, sker den anden publicering (slut november).

September er den sidste måned i 3.kvartal, så ultimo november ligger kvartalstallet for 3.kvartal klar. Det kan her vise sig, at der er en mindre afvigelse mellem det indsamlede tal for september og kvartalstallet. Derfor revideres månedstallet med kvartalstallet som benchmark. Ultimo november publiceres

3.kvartal, og månedstallet kan publiceres for tredje gang på baggrund af revision i forhold til kvartalstallet.

Det vil være afhængigt af, hvordan en given måned er placeret i kvartalspubliceringen, hvor lang tid der vil gå før tallet publiceres tredje gang. Eksempelvis bliver oktober måned, som er den første måned i 4. kvartal, først revideret i forhold til kvartalstallene i februar måned, tre måneder efter det første gang blev publiceret.

Usikkerhed

Usikkerheden på de månedlige tal beregnes på en anden måde end ved kvartalstallene grundet den valgte metode - det glidende gennemsnit med forecast. Ved kvartalstallene bliver der beregnet en usikkerhed ud fra et 95 pct. konfidensinterval direkte på de indsamlede tal. Dette lader sig ikke gøre ved månedstallene, hvor det nyeste tal som bekendt indeholder en tredjedel element af forecast. Her vil usikkerheden blive beregnet på baggrund af de to indsamlede måneder og den forecastede måned. De vægter alle 1/3 af den samlede usikkerhed for månedstallene. Typisk vil usikkerheden være større ved måneden med forecast, da denne netop ikke er indsamlet, men er en prognose.

De usikkerhedstal, der publiceres i 'Nyt fra Danmarks Statistik', er således udregnet på ovenstående måde. Det er kun ikke-sæsonkorrigerede tal, hvor det er muligt at beregne usikkerheder.

Omfanget af revisioner

I perioden fra januar 2013 til december 2013 har revisionerne for AKU-ledige mellem foreløbige og indsamlede tal i gennemsnit ligget på 5.000 ledige personer. Revisionerne mellem de indsamlede tal og tallene, der benchmarkes op mod kvartalstallene, har i gennemsnit ligget på 1.000 AKU-ledige fra januar til december 2013. For begge revisioner gælder det, at tallene for ledighed er blevet justeret op såvel som ned.

I tabel 4 nedenfor er juni 2013 valgt som eksempel, eftersom denne måned er det seneste datapunkt i serien, hvor der foreligger endelige tal. Dette skyldes, at hele kvartalet skal være indsamlet før det er muligt skabe et nyt gennemsnit på baggrund af benchmarking.⁴ Her ses det, at AKU-ledigheden på baggrund af et foreløbigt gennemsnit, inklusiv et forecast over juli måned, lander på 187.000 AKU-ledige. Efter indsamlingen af juli måneden, er gennemsnittet blevet revideret opadgående med 5.000 AKU-ledige. Da kvartalet afsluttes og benchmarken foretages, er gennemsnittet for juni måned stadigvæk 192.000.

Tabel 4: Revisioner i gennemsnittet for AKU-ledige 15-74 år

	Foreløbigt gennemsnit	Indsamlet gennemsnit	Benchmarket gennemsnit
Juni 2013	187.000	192.000	192.000

Revisionerne i forhold til benchmarken med kvartalstallene indgår ikke i beregningen af usikkerhed.

⁴ Først når 3. kvartal udkommer d.20.november 2013, vil det være muligt at foretage samme beregninger for hhv. juli, august og september.

4. Overvejelser i forbindelse med udgivelsen af månedstal

Kvalitet og aktualitet centralt

Der har været to afgørende parametre, der har stået centralt i produktionen af månedsdata. Det ene har handlet om at sikre så høj kvalitet på månedsestimaterne som muligt, og det andet om at kunne publicere tallene så de havde den højeste grad af aktualitet.

Dette har haft betydning for, at Danmarks Statistik har valgt den ovenfor præsenterede model for publikation af månedstal. Vi vil derfor i dette afsnit belyse de overvejelser, der ligger til grund for den valgte metode, ved samtidigt at præsentere de metodiske fravalg, der er blevet truffet.

Forskellige typer af måneds-AKU i Europa

Helt overordnet har der på europæisk niveau vist sig typisk at være tre veje at gå, for lande, der har valgt selv at producere månedstal for AKU.

I Sverige har man omlagt hele produktionen fra at være en kvartalsstatistik til at være en månedsstatistik i forbindelse med publiceringen af månedstal. Dette skete allerede i 1970. Sverige producerer således hele AKU som en månedsstatistik, modsat Danmark der laver månedstal på få aggregerede serier, og har kvartalsstatistikken som hovedundersøgelsen.

At Sverige har haft mulighed for at kunne lave månedstal på de mange finopdelte grupper skyldes, at antallet af gennemførte interviews i den svenske AKU er på over 20.000 pr. måned, hvor den danske AKU til sammenligning har ca. 7000-8000 pr. måned. En omlægning til en månedsstatistik vil således kræve en betragtelig udvidelse af stikprøven, hvilket ikke været muligt for den danske AKU.

Dette betød, at der var to muligheder at arbejde videre med.

Den første mulighed var at publicere rene månedstal på baggrund af de indsamlede måneder. Et rent månedstal vil sige, at estimatet for en given måned er bygget udelukkende på det indsamlede data for denne måned. Et sådant månedstal publiceres i Finland og Østrig.

Den anden mulighed var at publicere tal for dannet på baggrund af et tremåneders glidende gennemsnit. Denne model anvendes i Norge og England.

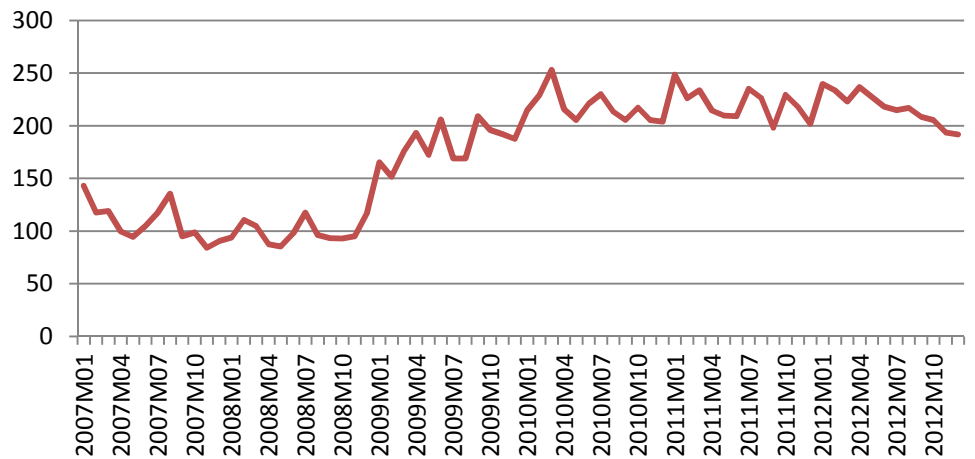
Udgangspunktet for arbejdet med den danske måneds-AKU var et ønske om at etablere en månedsstatistik på rene månedstal, idet dette tal er det klart mest anvendelige når en sammenligning mellem to på hinanden følgende måneder skal sammenlignes.

Udfordringer ved det rene månedstal

Det rene månedstal viste sig dog at skabe en række store udfordringer, som ikke alle lod sig løse.

Ser man på den udviklingen i den overordnede tidsserie på rene månedstal i figur 4 fra måned til måned, ses ganske store udsving i AKU-ledigheden. Eksempelvis ser man et fald fra marts 2010 til april 2010 på ca. 37.000 AKU-ledige og en stigning fra december 2010 til januar 2011 på ca. 44.000 AKU-ledige. Serien er således meget ustabil, grundet den relativt lille stikprøve.

Figur 4: Rå månedstal for AKU-ledigheden for 25-74-årige i tusinde

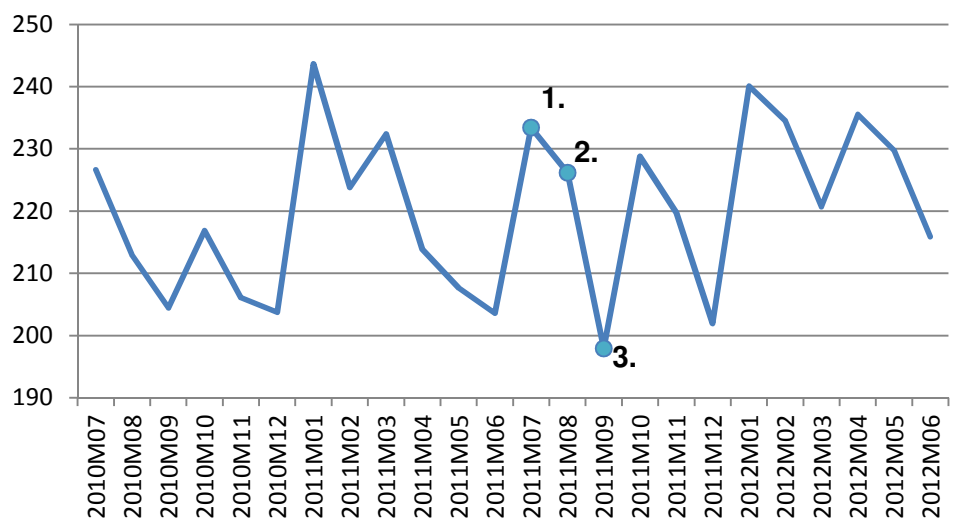


Et vigtigt element i at stabilisere tidsserien knytter sig til sæsonkorrektio n, da en sæsonkorrektio n vil stabilisere sæsonbetingede udsving mellem de enkelte måneder.

Her opstod det første store problem. Det viste sig, at det ikke var muligt at sæsonkorriger e, hverken underserierne opdelt på køn og alder eller de overordnede serier for ledighed og beskæftigelse.

Det mest centrale problem var dog en uforklarlig systematik i månedstallene, der især er udtalte i perioden 2010-2011. Dette kan genfindes i figur 5, hvor der er fokuseret ind på perioden omkring 2010-2011, hvor denne systematik er mest udtalt.

Figur 5: Systematik i AKU-ledigheden på rene månedstal i tusinde

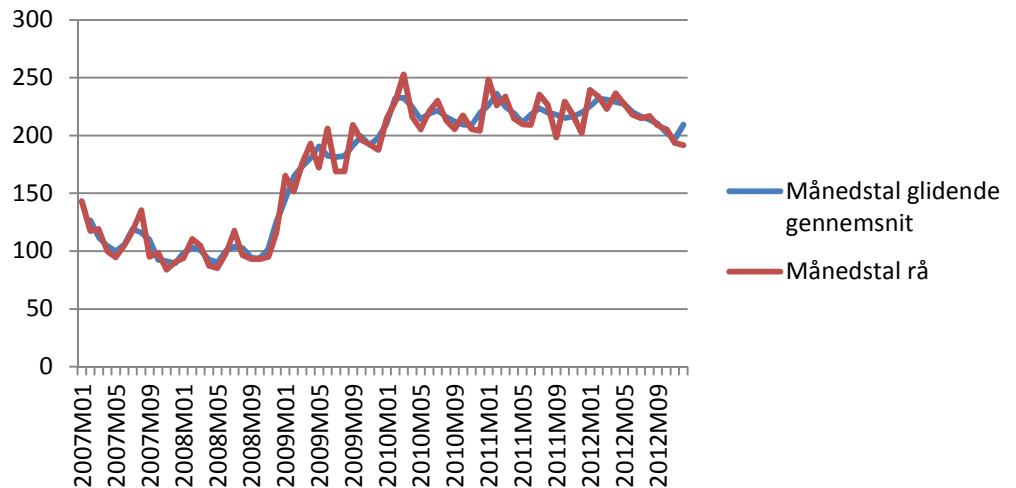


Som det ses har der været en tendens til at månedsestimaterne på de tre måneder, der tilsammen danner et kvartal, systematisk vurderer ledigheden højt i

den første måned, hvorefter ledigheden falder i anden måned, for slutteligt at falde markant i tredje måned, hvilket gentager sig flere kvartaler i træk. Systematikken er ikke problematisk over kvartaler, da der sker en udligning, men det satte grundlæggende spørgsmålstejn ved kvaliteten af de rene månedstal. Der blev foretaget grundige undersøgelser af både AKU-sektionen og Danmarks Statistiks metodeenhed uden at en god forklaring på systematikken blev fundet. Dette blev afgørende for, at det rene månedstal blev vurderet til at være af for dårlig kvalitet til, at man kunne udgive det.

Det viste sig, at det glidende gennemsnit, hvis metode beskrives i afsnit 2, løste flere af de problemer, der opstod ved de rene månedstal. Først og fremmest gjorde det glidende gennemsnit tidsserien langt mere stabil, men ydermere viste det sig muligt at sæsonkorrigere alle serierne, og den problematiske systematik blev fjernet. Dette ses i figur 6.

Figur 6: Glidende gennemsnit og rå månedstal for AKU i tusinde



Som det ses, så er det glidende gennemsnit langt mere stabilt end den mere volatile tidsserie for rå månedstal. Samtidigt fjernes den systematik i perioden 2010 -2011 som ses i de rå månedstal.

Valg af model til det glidende gennemsnit

Også i relation til et tre måneders glidende gennemsnit skulle der træffes valg omkring, hvilken model der skulle anvendes. Valget af model blev især en afvejning af datakvalitet i forhold til publiceringstidspunkt. Centralt var således, at tallene ikke skulle publiceres for lang tid efter publiceringsmåneden, ideelt set ultimo den følgende måned.

Det betød også, at det var udelukket, at indsamle alle tre måneder og først publicere data efter de alle tre var indsamlet. Dertil kom ønsket om at publicere på den midterste måned, så den seneste udvikling også indgik i gennemsnittet. Hvis man i stedet valgte at publicere på den sidste måned i det glidende gennemsnit, ville man ikke få den seneste udvikling med.

For den valgte model publiceres der på den midterste måned af de tre måneder, der danner gennemsnittet. Den sidste måned, oktober, i denne model er et forecast, dvs. en prognose, der med baggrund i de tidligere måneder søger at give et bud på udviklingen i den sidste måned, der endnu ikke er indsamlet.