

# Stokastiske befolkningsprognoser for Norge 2012-2060

Aslaug Hurlen Foss\*

Norges befolkning passerte nylig 5 millioner mennesker. Dette skjedde mye tidligere enn de offisielle befolkningsframskrivingene anslo i hovedalternativet fra 2005. Dette illustrerer at befolkningsutviklingen er svært usikker. Det kan derfor være nyttig å knytte mål for usikkerhet til prognosene i form av sannsynligheter. De offisielle befolkningsframskrivingene lager mange alternativer, blant annet såkalt høyalternativ og lavalternativ, men det er ikke knyttet sannsynligheter til disse. Her presenteres en alternativ metode for befolkningsframskrivninger, en stokastisk metode for å lage befolkningsprognoser. Resultatene fra denne metoden vil bli sammenlignet med de offisielle befolkningsframskrivingene.

## Innledning

Anslag for fremtidig befolkning er beheftet med usikkerhet. Til tross for at befolkningsframskrivingene lager flere alternative baner for utviklingen vil det fortsatt kunne komme overraskende endringer i de underliggende demografiske forholdene. Et eksempel på dette har vært den uventede høye arbeidsinnvandringen til Norge de siste årene. Nettoinnvandringen har ligget på rundt 40 000 de siste årene og var oppe i hele 47 000 i 2011. Dette har ført til en høyere befolkningsvekst enn forventet og 19. mars passerte vi 5 millioner innbyggere. Den høye nettoinnvandringen har ligget på et nivå som er mye høyere enn forutsetningene som var lagt til grunn for befolkningsframskrivingene for noen år siden. I tillegg har fødselsoverskuddet<sup>1</sup> også ført til at befolkningen vokser. I de siste årene har fødselsoverskuddet ligget på rundt 20 000, med drøyt 60 000 fødte og 40 000 døde. Endringene i mønsteret for dødelighet og fruktbarhet har vært mye langsommere enn endringene i migrasjon. Men også i dødelighets- og fruktbarhetsutviklingen kan det komme overraskende endringer. I mange år ble det antatt at forventet levealder kom til å stabilisere seg på et gitt nivå, mens det nå har vist seg at den øker jevnt. I dagens offisielle befolkningsframskrivninger illustreres usikkerheten gjennom at det presenteres ulike baner for befolkningsutviklingen, basert på ulike antagelser om utviklingen i sentrale størrelser. Det angis imidlertid ikke hvor sannsynlige de ulike alternativene er. Til hjelp i anvendelsen av befolkningsframskrivninger kan det være nyttig å sette tall på usikkerheten. I denne artikkelen ser vi nærmere en metode for å gjøre dette.

## Deterministiske befolkningsframskrivninger

De offisielle befolkningsframskrivingene er laget etter "hvis"-prinsippet. Det blir tatt utgangspunkt i en observert folkemengde fordelt på alder og kjønn. Det blir laget antagelser om utviklingen i henholdsvis fruktbarhet, dødelighet og migrasjon. Befolkningsframskrivingene viser resultatene av hva som skjer med befolkningen hvis disse antagelsene inntreffer. I tillegg kan befolkningen bli delt opp i regioner og etter innvandringsstatus, og det kan gjøres separate antagelser for hver enkelt gruppe. For nærmere beskrivelse av metoden, se Brunborg mfl. (2011). Det alternative for fruktbarhet, dødelighet og migrasjon som man tror mest på blir kalt middelalternativet M, det høyeste alternativet H og det laveste alternativet L. Ved å kombinere forskjellige alternativer for hver av komponentene gis det noen anslag på hvordan befolkningen kan variere i framtiden. Hvis man for eksempel har valgt lavalternativet for fruktbarhet, så vil det i prognosene være lav fruktbarhet i hele perioden. Dette er en streng forutsetning, som kan være urealistisk. I virkeligheten vil fruktbarheten variere fra år til år, og kan i prinsippet gå fra lavalternativet til høyalternativet i løpet av noen få år. En annen restriktiv, og kanskje urealistisk, antagelse er at man forutsetter at kombinasjonen dødelighet, fruktbarhet og migrasjon er konstant for hele perioden. I virkeligheten kan alle disse komponentene variere mellom lave og høye nivåer uavhengig av hverandre. Dette blir gjerne illustrert ved å lage mange alternative befolkningsframskrivninger. Når fruktbarhet, dødelighet og migrasjon alle har 3 nivåer vil det si at det er mulig å lage opptil 27 kombinasjoner, og hvis vi tar med innenlandske flyttinger blir det enda flere. Ut over at man anser middelalternativet som det mest troverdige, er det dermed vanskelig å vurdere sannsynligheten av de ulike kombinasjonene opp mot hverandre.

Aslaug Hurlen Foss er statistikkrådgiver ved Seksjon for statistiske metoder og standarder (aslaug.hurlen.foss@ssb.no)

\* En takk til Helge Brunborg, Nico Keilman, Øyvind Langsrud og Elin Halvorsen for nyttige kommentarer til denne artikkelen.

<sup>1</sup> Antall fødte minus antall døde

## Stokastiske befolkningsprognoser

Dersom man i befolkningsprognoser ønsker å knytte sannsynligheter til ulike utfall, må man bruke en annen framgangsmåte enn den som brukes i befolkningsframskrivingene i dag. Det finnes mange modeller og metoder å lage en stokastisk prognose på. Stokastiske befolkningsprognoser kan bli laget ved å beregne sannsynlighetsfordelinger og baner for fruktbarhet, dødelighet og migrasjon. Banene for fruktbarhet, dødelighet og migrasjon kan lages med utgangspunkt i formelle modeller eller basert på hva noen tror er mulig utvikling. Ofte blir de laget som en kombinasjon av dette. Hovedpoenget er at forutsetningene for de ulike komponentene angis som sannsynlighetsfordelinger, og man kan eventuelt spesifisere om for eksempel fordelingene for fruktbarhet og dødelighet avhenger av hverandre. En stokastisk befolkningsmodell forsøker å etterligne den virkelige verden, der utviklingen kan hoppe litt opp og ned, og utviklingen i fruktbarhet, dødelighet og migrasjon kan gå i forskjellige retninger. Ved å simulere mange forskjellige befolkningsutviklinger kan vi lage prognoser for befolkningsutviklingen, og angi hvor stor sannsynlighet det er for et bestemt resultat.

Stokastiske befolkningsprognoser er også laget etter "hvis"-prinsippet. Hvis banene og sannsynlighetsfordelingene rundt banene for dødelighet, fruktbarhet og migrasjon er slik, da vil sannsynligheten for befolkningen bli slik. Vi kan lage prediksjonsintervall<sup>2</sup> uttrykt med sannsynlighet for befolkningsprognosene. Prediksjonsintervallet vil være avhengig av valg av modell for å estimere variansen og valg av banene for dødelighet, fruktbarhet og migrasjon. I stokastiske prognoser blir brukerne tvunget til å tenke i intervaller med sannsynligheter. Det er ikke mulig å velge ett enkelt alternativ, slik som det er mulig i deterministiske befolkningsframskrivinger.

## Modell

Beregningene i denne artikkelen baserer seg på en metode for å lage stokastiske prognoser som er utviklet av Alho mfl. (2008). Denne modellen benytter middelalternativet<sup>3</sup> i den deterministiske befolkningsframskrivingen og legger til usikkerhet rundt denne prognosen. Dette blir gjort ved å benytte de aldersspesifikke ratene fra middelalternativet i befolkningsframskrivingen for dødelighet og fruktbarhet. I tillegg blir antallet nettoinnvandrere fra middelalternativet brukt, fordelt etter alder og kjønn. Antagelser om fruktbarhet og innvandring ble i 2011-framskrivingen gjort per landgrupper og ikke for hele landet under ett. I den stokastiske modellen er det derfor brukt fruktbarhetsratene fra middelalternativet i 2010-framskrivingen, som har omtrent samme bane som middelalternativet i 2011. I tillegg ble totaltall for nettoinnvandring fra 2011-framskrivingen brukt, men fordelt etter kjønn og et fast aldersmønster. Dødelighetsratene som er brukt er middelalternativet i 2011-framskrivingen.

### Scaled model for error

Modelleringen blir gjort i log-skalaen. La  $R(j, t)$  være

aldersspesifikk rate,  $j$  representerer alder og  $t$  tid.  $\hat{R}(j, t)$  er prognosen for raten fra den deterministiske befolkningsframskrivingen.

$$\log(R(j, t)) = \log(\hat{R}(j, t)) + X(j, t) \quad j=1 \dots J \text{ og } t=1 \dots T$$

$X(j, t)$  er et feilledd

$$X(j, t) = \sum_{q=1}^t \varepsilon(j, q)$$

$$\varepsilon(j, t) = S(j, t)(\eta_j + \delta_t)$$

$S(j, t) > 0$  er en fast ikke-tilfeldig vekt som varierer over alder og tid.

$\eta_j$  er en tilfeldig variabel som varierer med alder og den er antatt å være normalfordelt med forventning 0 og varians  $\kappa$ :  $N(0, \kappa)$

$\delta_t$  er en tilfeldig variabel som varierer med tid og alder og den er antatt å være normalfordelt med forventning 0 og varians  $1 - \kappa$ :  $N(0, 1 - \kappa)$  og ukorrelert over tid

Variablene  $\eta_j$  og  $\delta_t$  er antatt å være ukorrelerte.

$$\text{corr}(\eta_j, \eta_i) = \text{corr}(\delta_j, \delta_i) = \rho^{|i-j|} \quad \text{autoregressiv struktur}$$

Nettoinnvandring og de aldersspesifikke ratene tillegges et støyledd. Ved å gjøre dette kan disse nye ratene bli sett på som tilfeldige variabler. Støyleddet som er lagt til, består av en deterministisk vekt etter alder og tid, og av to tilfeldige variabler som styrer korrelasjon mellom aldersgrupper og tid, se faktaboks. Estimerer for vekten og korrelasjonen for fruktbarhetsratene, dødelighetsratene og nettoinnvandring er hentet fra Alders mfl. (2007). Dette forskningsprosjektet "Uncertain population of Europe" ble gjennomført i perioden 2001-2004 og finansiert av EU's femte rammeprogram. Selv om det er noen år siden dette prosjektet ble gjennomført, er det antatt at beregningen av usikkerheten er fortsatt gyldige.

For å lage de stokastiske prognosene for Norge ble dataprogrammet "Program for error propagation, PEP" benyttet.<sup>4</sup> Dette programmet tar utgangspunkt i befolkningen per 1. januar 2011 og trekker så verdier fra sannsynlighetsfordelingen for fruktbarhet, dødelighet og migrasjon. Ut fra dette blir en fremtidig bane for befolkningen laget. I denne analysen er det laget 3 000 slike baner som så er brukt til å beregne prediksjonsintervaller. Hvor mange prosent prediksjonsintervallet skal dekke er valgfritt. Det er valgt prediksjonsintervall som ligger nær høy- og lavalternativet i befolkningsframskrivingen. For totaltall for hele landet er det valgt å bruke et 90 prosents prediksjonsintervall, mens det

<sup>2</sup> Konfidensintervall for predikerte fremtidige verdier

<sup>3</sup> MMMM - middelalternativet i befolkningsframskrivingene

<sup>4</sup> Det er utviklet av Juha Alho, for beskrivelse av programmet, se <http://joyx.joensuu.fi/~ek/pep/pepstart.htm>.

for befolkningen fordelt etter alder og kjønn er valgt å bruke et 80 prosents prediksjonsintervall.

## Resultater

### Totalt for hele befolkningen

Prediksjonsintervallet øker jo lenger frem i tid vi går. I 2020 vil befolkningen være mellom 5,4 og 5,6 millioner med 90 prosent sikkerhet ut fra denne stokastiske modellen. Prediksjonsintervallet er da på 200 000 personer, mens i 2060 vil dette intervallet øke til hele 4,5 millioner personer. Norge vil da med 90 prosent sannsynlighet ha en befolkning på mellom 5,5 og 9,0 millioner personer. Prediksjonsintervallet ligger innenfor lav- og høyalternativene fra befolkningsframskrivingene<sup>5</sup>. I dette tilfellet fungerer lav- og høyalternativet godt som et slags prediksjonsintervall for å illustrere usikkerheten i framskrivingene.

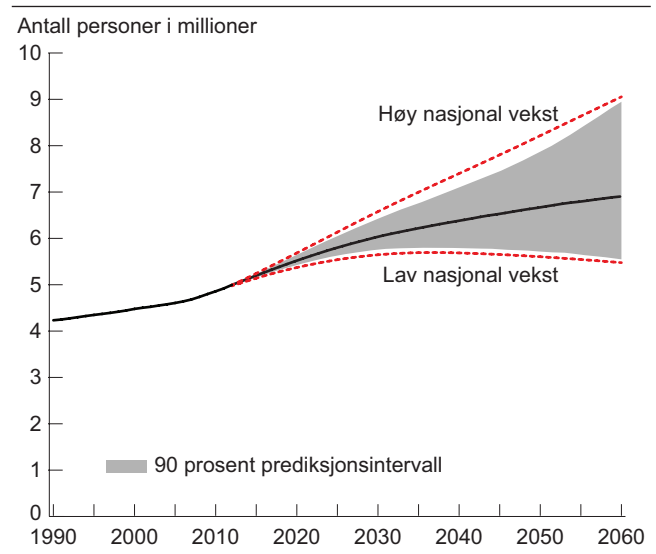
Uansett om man baserer seg på stokastiske befolkningsprognoser eller deterministiske lav- og høyalternativer slik som i dagens offisielle framskrivinger, vil anslagene selvsagt avhenge av de grunnleggende antagelsene man gjør seg om sentrale størrelser. En illustrasjon på dette er den stokastiske prognosen presentert i Keilman mfl. (2002), laget ut fra befolkningen i 1996. Den viste at viste at med 2,5 prosent sannsynlighet ville befolkningen være 5 millioner eller mer i 2012. Grunnen til at den observerte folkemengden havnet i randsonen av prognosen var i hovedsak at innvandringen ble langt høyere enn det som ble antatt i prognosen.

### Befolkning etter alder og kjønn

Når vi ser på aldersfordelingen i den stokastiske prognosene og framskrivingene, blir forskjellene mellom de to metodene større. Figur 2 viser befolkningspyramider med 80 prosent prediksjonsintervall i grått for de stokastiske prognosene og røde linjer for de deterministiske befolkningsframskrivingene. For personer over 80 år, er prediksjonsintervallet større enn forskjellen mellom lav- og høyalternativet fra befolkningsframskrivingene, for de fleste årene 2012-2060. Det samme gjelder for de yngste, der prediksjonsintervallet også er klart større enn forskjellen mellom lav- og høyalternativet fra framskrivingen. Prediksjonsintervallet øker med prognoseperioden, men i de deterministiske framskrivingene er fruktbarheten satt til å være konstant etter en viss periode<sup>6</sup> slik at intervallet bare varierer med antall kvinner i fruktbar alder. I 2060 tilsvarende forskjellen mellom lav- og høyalternativet for aldersgruppen 0-4 år om lag et 60 prosent prediksjonsintervall.

Å modellere og framskrive inn- og utvandring godt på en god måte, er krevende. Forutsetningene for migrasjon i befolkningsframskrivingene bygger på en økonomisk modell. Innvandringen bestemmes i denne modellen av inntektsnivået i Norge relativt til andre lands inntektsnivå og arbeidsløsheten i prosent av arbeids-

Figur 1. Stokastiske befolkningsprognoser og befolkningsframskriving. Hele landet. 2012-2060



Kilde: Statistisk sentralbyrå

styrken i Norge og andre land. For nærmere beskrivelse av modellen se Brunborg mfl (2011). Migrasjon blir påvirket av politikk og økonomi både i Norge og resten av verden, og slike endringer kan ha en rask innflytelse på migrasjonen. I de stokastiske prognosene er det brukt middelalternativet for migrasjon og lagt til variasjon ut fra historiske data. I framskrivingene i 2011 ligger nettoinnvandringen i høyalternativet på et svært høyt nivå. Dette ser vi tydeligst i aldersgruppen 20-65 år, som er den aldersgruppen som blir sterkest påvirket av ulike forutsetninger om innvandring. Høyalternativet ligger i 2060 utenfor 99-persentilen i de stokastiske prognosene. Innvandringen de siste årene har vært svært høy og tallene kan variere svært mye. Beregningene av usikkerhet i migrasjon for denne stokastiske prognosene kan dermed se ut til å være underestimert i forhold til de siste årenes utvikling.

Når befolkningen er fordelt etter alder og kjønn, kan vi ikke tolke variasjonsbredden mellom høy- og lavalternativet i framskrivinger som et prediksjonsintervall på ett bestemt nivå. Ut fra de stokastiske prognosene tilsvarende variasjonsbredden fra framskrivingen fra 60 til 100 prosent i et prediksjonsintervall, alt etter hvilken aldersgruppe og hvilket år prognosen er for.

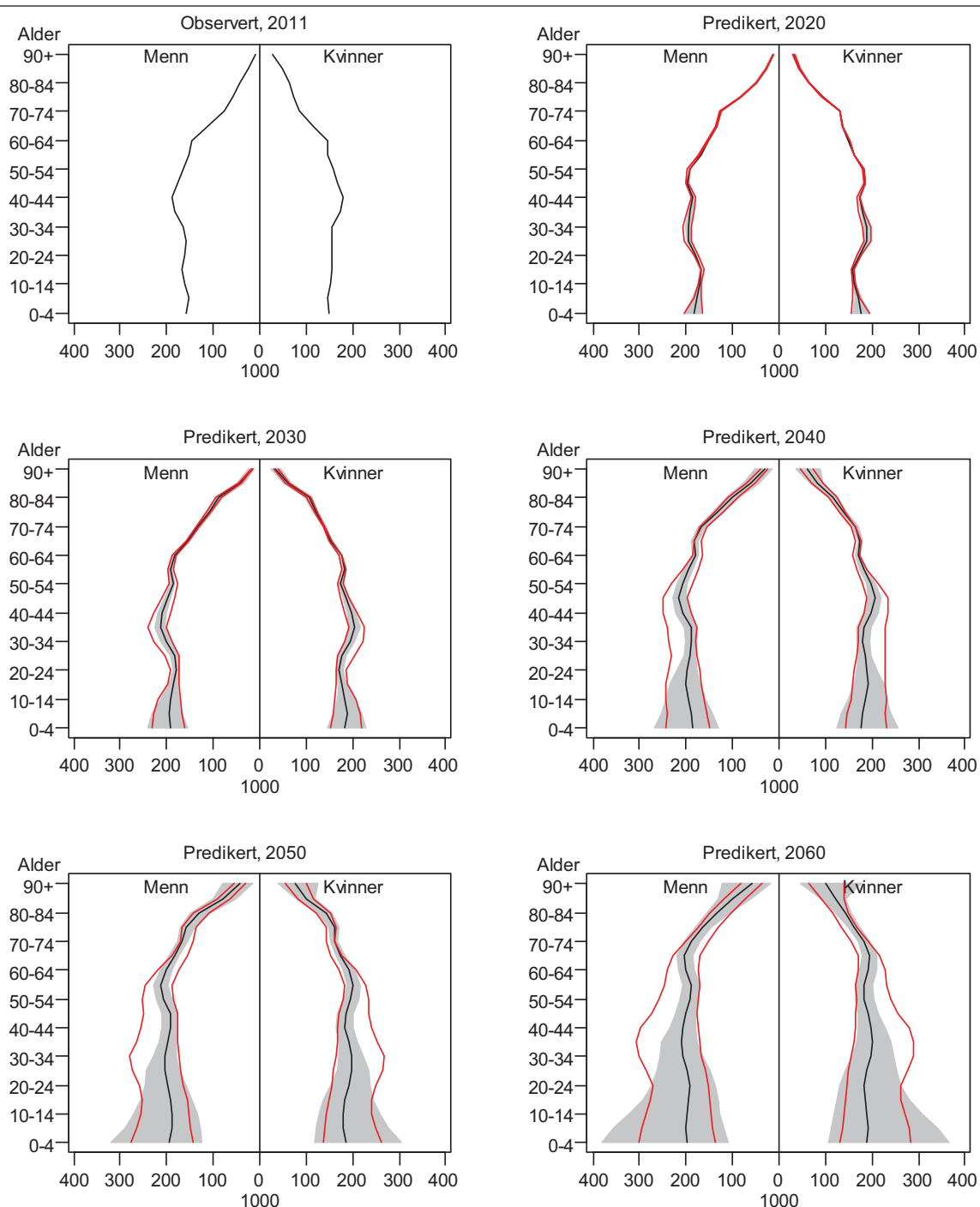
### Forsørgerbyrde

Antallet eldre i forhold til personer i yrkesaktiv alder gir en indikasjon om samfunnets forsørgerbyrde til enhver tid. Den antas å øke kraftig i løpet av de neste femti år, se figur 3, og det er knyttet stor interesse til prognosene for akkurat dette forholdstallet. I den nye pensjonsreformen er det fleksibelt uttak av pensjon i alderen 62 til 75 år. Den reelle pensjonsalderen ble i 2005 målt til å være 63,1 år (Bø, 2007). I 2011 var forholdstallet mellom antall personer 65 år og eldre og personer i alderen 20-64 år på 0,25, eller en til fire. Ut fra de stokastiske prognosene vil det i 2060 ligge mellom 0,41 og 0,58 med 80 prosent sannsynlighet. Når vi går mer enn 20 år

<sup>5</sup> HHMH- og LLML- alternativene.

<sup>6</sup> I høyalternativet er samlet fruktbarhetstall, SFT, antatt å øke til 2,05 i 2015 for deretter å holdes fast, mens i SFT i lavalternativet antas å synke til 1,71 i 2020 og deretter være konstant.

Figur 2. **Befolkningspyramider med 80 prosent prediksjonsintervall i grått og røde linjer for lav- og høyalternativet i befolkningsframskrivingen. Observert 2011, predikert 2020, 2030, 2040, 2050 og 2060**



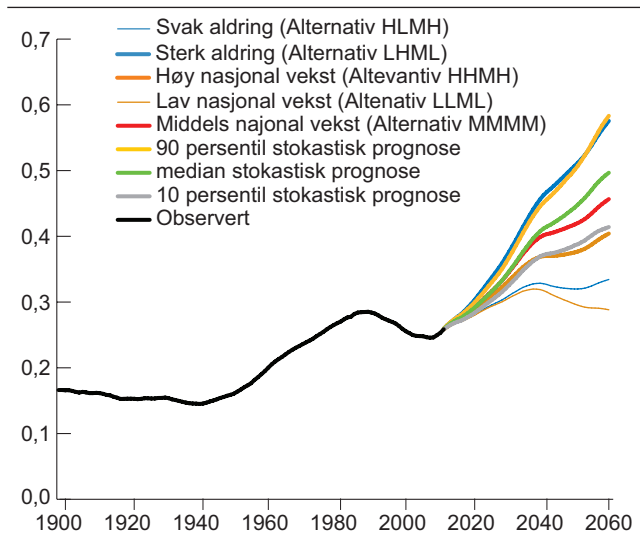
Kilde: Statistisk sentralbyrå

frem i tiden vil usikkerheten ikke bare skyldes usikkerhet i dødelighet og innvandring, men også endringer i fruktbarhet. Likevel skyldes usikkerheten i prognosene for forsørgerbyrden hovedsakelig usikkerhet i dødelighet og migrasjon. Endringene i dødelighet går langsomt, der levealderen har vært jevnt stigende de siste årene. Nettoinnvandringen har vært sterkt økende de siste årene og er generelt svært varierende. Prediksjonsintervallet for forsørgerbyrden kan være noe urealistisk da usikkerheten i nettoinnvandringen kan se ut til å være underestimert i denne stokastiske modellen. Beregningene av usikkerheten i migrasjon er her gjort før

den siste innvandringsbølgen. Med større innvandring vil forsørgerbyrden bli mindre.

Ved bruk av lav- og høyalternativet fra befolkningsframskrivingene får vi andre resultater, middelalternativet ligger høyere enn høyalternativet. Grunnen til dette er at høy innvandring fører til lavere forsørgerbyrde. Lav- og høyalternativet fra befolkningsframskrivingene kan dermed ikke bli brukt som et prediksjonsintervall for selve forsørgerbyrden. For å si noe om hvordan utviklingen i forsørgerbyrde kan bli, vil det være mer nyttig å bruke kombinasjoner av høy levealder og lav

Figur 3. **Forsørgerbyrde for eldre. Forholdet mellom antall eldre (65+) og antall i yrkesaktiv alder (20-64 år). Observert 1900-2011, predikert 2012-2060**



Kilde: Statistisk sentralbyrå

nettoinnvandring, det vil si alternativet kalt sterk aldring eller lav levealder og høy nettoinnvandring, det vil si alternativet kalt svak aldring.

### Oppsummering

Lav- og høyalternativet i befolkningsframskrivingene blir trolig brukt av noen som et slags prediksjonsintervall. Det er derfor viktig å vite at desto mer vi deler opp befolkningsframskrivingene etter alder, kjønn, region eller innvandrerkategori, jo mindre sannsynlighet vil det være for at befolkningen faktisk vil ligge i dette intervallet. Dessuten vil det for forholdstall slik som forsørgerbyrde ikke lenger være mulig å bruke lav- og høyalternativet i befolkningsframskrivingen som et prediksjonsintervall. Ved å bruke stokastiske prognoser kan vi knytte sannsynligheter til hvordan framtidens befolkning kan bli, ut fra hva vi på prognosetidspunktet tror om utviklingen i fruktbarhet, dødelighet og migrasjon. Trendsift i innvandring, dødelighet eller fruktbarhet kan ha stor innvirkning på befolkningsutviklingen, men kan være vanskelig å forutse. Bruk av stokastiske prognoser istedenfor deterministiske befolkningsframskrivinger endrer ikke denne usikkerheten. Sammenliknet med de deterministiske befolkningsframskrivingene har denne metoden likevel den fordelen at utviklingen i komponentene tillates å variere over tid og at det er mulig å sette tall på hvilken befolkningsutvikling som er mer sannsynlig enn en annen, gitt forutsetningene som ligger til grunn.

### Referanser

Alders, M., Keilman, N. og Cruijisen, H. (2007): Assumptions for long-term stochastic population forecasts in 18. European countries. *European Journal of Population* (2007) 23:33-69

Alho, J. M., Jensen, S. E. H., Lassila, J. (2008): *Uncertain Demographics and fiscal Sustainability*. Cambridge University Press

Brunborg, H. og Texmon, I. (2011): *Befolkningsframskriving 2011-2100: Modell og forutsetninger*. Økonomiske analyser nr 4/2011. Statistisk sentralbyrå

Bø, T. P. (2008): *Arbeidstiden - mønstre og utviklings-trekk*. kapittel 5. Rapporter 2008/12. Statistisk sentralbyrå

Keilman, N., Pham, S. Q., Hetland, A. (2002): Why population forecast should be probabilistic – illustrated by the case of Norway. *Demographic research*. Volume 6. Article 15. p 409-454