

Tidsrekkebasert modellering og framskriving av Covid-19 i Norge—nye tilfeller, innleggelser og R-tall

Ragnar Nymoen *
7. april 2021 (oppdateres)

Sammendrag

I dette notatet presenteres et opplegg for å lage prognoser for nye tilfeller (insidens) og for innlagte på sykehus med påvist Covid-19. Metoden og erfaringsgrunnlaget er hentet fra framskriving av makroøkonomiske tidsrekker, og er derfor komplementært til framskrivningsmetoden som benyttes av FHI. Kombinasjon av prognosemetoder kan bidra til økt prognoseøyaktighet, men dette forutsetter at komplementære metoder utvikles og benyttes i praktisk framskriving. Dette har ikke skjedd i særlig grad i Norge under pandemien. Metoden som presenteres tillater samtidig beregning av det tidsavhengige R-tallet, som er en endogen variabel i modellen, ikke en parameter. Denne tilnærmingen til beregning av R-tallet er vanlig i utlandet, men har fått mindre oppmerksomhet i Norge

1 Innledning

Etter utbruddet av Covid-19 i Norge har allmenheten blitt godt kjent med reproduksjonstallet, R . Begrepsmessig viser R hvor mange personer én person med koronasmitte smitter videre. Hvis $R = 1$, vil hver smittede person smitte én annen person. Et høyere R-tall kan varsle mulig ny kraftig spredning av viruset, som igjen kan utfordre helsevesenet. Om tallet ligger under $+1$, er det derimot mindre smittespredning, og man kan si at man har en forløpig kontroll over utviklingen. Mediene omtaler ofte R-tallet i saker om utsiktene framover. Onsdag 24. mars rapporterte [NRK](#) at FHI i sin siste modelleringsrapport hadde tallfestet R-tallet til 1,3 (nasjonalt). Anvendt på en daglig tidsskala vil dette R-tallet mangedoble antall nye smittede svært raskt, men i samme rapport framskrev FHI at tallet på nye tilfeller til å være 2500 søndag 28. mars 2021¹ Prognosetallet var mer enn en dobling fra nivået 24. mars, men var likevel mindre dramatisk enn det nasjonale R-tallet på 1,3 alene skulle tilsi. Forklaringen er at FHI benytter en kompleks prognosemodell som tar hensyn til at det er regionale forskjeller i smittesituasjone og en rekke andre faktorer som er med å bestemme hvordan Covid-19 sprer seg i Norge.

Den operasjonelle definisjonen av R-tallet varierer fra land til land. I Norge er R-beregningen en integrert del av den epidemiologisk modelleringen og simuleringen av pandemien². I flere andre

*Universitetet i Oslo, Økonomisk institutt. e-post: ragnar.nymoen@econ.uio.no

¹Situational awareness and forecasting for Norway. FHI COVID-19 modelling team. Week 11, 24 March 2021.

²<https://www.fhi.no/sv/smittsomme-sykdommer/corona/koronavirus-modellering/>

land beregnes det tidvarierende R-tallet mer direkte fra tidsrekke­data for nye tilfeller, eller for antallet nye innleg­gelser på sykehus. På nasjonalt nivå har dette vist seg å gi relevant informasjon. Med den mer direkte tilnærmingen kan R-tallet finnes som et biprodukt fra en modell som allerede framskriver nettopp antallet nye smittede. Med et slikt opplegg er R-tallet ikke en parameter i modellen (som driver resultatene), men en endogen variabel. Dette notatet beskrives en slik modell, og det vises eksempler på framskrivninger som også inkluderer R-tall.

2 Modellbeskrivelse

Framskrivninger ble publisert på [Normetrics](#) under den første bølgen vinteren 2020. I juni i fjor var imidlertid tallene for nye smittede og sykehusinnleg­gelser blitt så lave at det ikke var noe poeng å fortsette med framskrivningene, inntil et eventuelt nytt utbrudd.³ Nå befinner vi oss midt i en tredje bølge, med smittetall som var langt høyere enn under den første bølgen, og med stigende antall sykehusinnleg­gelser. Det har derfor vært interessant å ta fram modellen fra ett år tilbake og respesifisere den ved å bruke de forlengede og oppdaterte tidsrekke­ne.

Modellen framskriver daglig antall nye bekreftede smittede med Covid-19. Data­kilden er FHI. Tallene er ordnet etter dato på avlagt test. Dette betyr at det vil være avvik mellom disse tallene og de hyppig refererte tallene for nye smittede på VG.no.

Modellen framskriver i tillegg antall nye pasienter innlagt på sykehus med bekreftet Covid-19, og det daglige antallet som er innlagt på sykehus i Norge («sykehussenger»). Men også framskrivningene av disse variablene drives av prognosen for nye tilfeller.

Den sentrale variabelen som bestemmes i modellen er altså antallet nye smittede per dag, NS_t . Når NS_t er bestemt i modellen, oppdateres det kumulative antallet SS_t med definisjons­ligningen:

$$SS_t = SS_{t-1} + NS_t. \tag{1}$$

Siden det er rimelig å anta at NS_{t+1} er korrelert med SS_t , er det rimelig å inkludere «lags» av SS_t i modell­ligningen som bestemmer NS_t . Mer presist har vi benyttet:

$$NZ_t = SS_{t-k_1} - SS_{t-k_1-k_2}$$

som en indikator på smitten i populasjonen. Et rimelig anslag er at det går fire dager fra en person blir smittet til han eller hun kan smitte andre, som innebærer $k_1 = 4$. Etter estimering av sammenhengen (under første bølge) viste det seg at vi like gjerne kunne sette $k_1 = 1$. Forklaringen

³<http://normetrics.no/covid-19-and-norwegian-jobs-crisis/>

på at «timingene» ikke trenger være så presis når formålet med modellen er prognoser, kan være at de ulike kandidatene til indikatorer for smittepotensiale er innbyrdes korrelerte, og korrelerte med den ikke-observerbare reelle smitten i befolkningene. Smittepotensialindikatoren NZ_t avhenger også av «cut-off» parameteren k_2 , kanskje særlig tidlig i epidemien (eller i starten av ny bølge). Basert på erfaringene med framskrivningene under første bølge, har vi i mars 2021 fortsatt med å bruke $k_2 = 14$. Men å velge et noe bredere tidsintervall (vindu) har ikke avgjørende innvirkning på framskrivningene.

Dersom vi setter $k_1 = 1, k_2 = 13$:

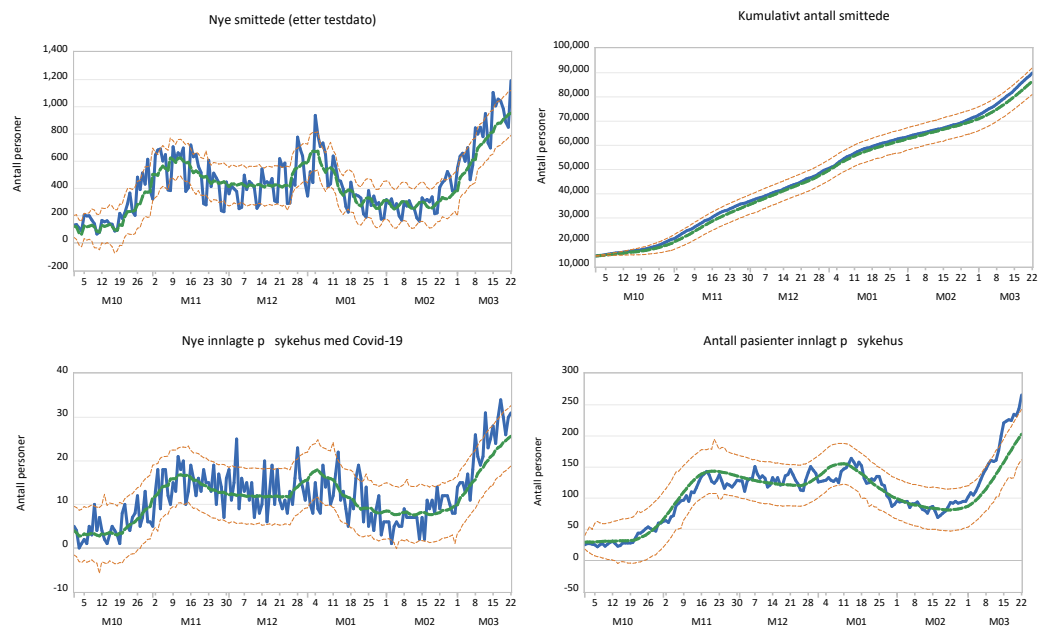
$$NZ_{t-1} = (SS_{t-1} - SS_{t-14}) = \sum_{j=1}^{14} NS_{t-j} \quad (2)$$

som er positiv så lenge epidemien varer. I tillegg til NZ_{t-1} inneholder modelleringen for NS_t lags av $\Delta NS_t (= NS_t - NS_{t-1})$ av høy orden. Dette er gjort for blant annet å modellere mønsteret med færre positive tester på lørdager og søndager enn på de andre ukedagene.

De øvrige variablene i modelleringen kan tolkes som effekter av importert smitte, effekter av ikke-farmasøytiske tiltak og endringer i testkapasitet (og i befolkningens utnyttelse av den kapasiteten). Disse representerer intervensjoner, eller mer generelt strukturelle brudd i den økonometriske ligningen. Fra makroøkonomisk framskrivninger er vi godt kjent med at strukturelle brudd i økonomien skjer hyppig og at det er et nesten uoverstigelig problem å få brakt slike strukturelle brudd inn i prognosemodellen raskt nok til at store prognosefeil kan unngås, [Nymoen \(2019, Kap.12\)](#). Det er grunn til å tro at de som framskriver forløpet av epidemier står ovenfor lignende utfordringer.

Når modellen brukes til sporing (føyning) av pandemien, vil variasjon i disse variablene (f.eks. en indikatorvariabel for vinterferien i 2020) være med å påvirke simuleringresultatene over den historiske perioden. Når modellen brukes til rene framskrivninger vil derimot disse variablene holdes konstant på sine gitte nivåer. Framskrivningene bygger dermed på en forutsetning om ingen nye utbrudd pga importsmitte, og ingen nye intervensjoner av typen for å «slå ned» ved hjelp av nedstegning, innreiseforbud osv. En tredje bruk av modellen er til å simulere scenarier, og da kan indikatorer for politikkbruk eller importsmitte aktiveres i framskrivningsperioden, for nettopp å simulere forskjellige mulige, men ikke spesielt sannsynlige, trender for spredning av Covid-19 i befolkningen.

Antallet nye pasienter innlagt med Covid-19, $NINL_t$, er modellert som en funksjon av antall nye smittede (i periode t og med lag). Det er estimert en noe sterkere sammenheng mellom smitteindikatoren og $NINL_t$ fra midten av mars 2021 enn tidligere i epidemien. Antallet innlagte, INL_t er modellert som en funksjon av $DINL_t$, med koeffisient 1 pålagt, og antallet innlagte dagen før, (INL_t , med koeffisient mindre enn +1.



Figur 1: Dynamisk simulering av Covid-19 i Norge, fra 1. oktober 2020 til 22. mars 2021. Blå grafer er faktiske tall, grønne stiplede grafer er simuleringer (gjennomsnitt av 1000 simuleringer) og de avtandene mellom de røde stiplede grafene angir konfidensintervall (90 prosent).

I påvente av vaksinerings av en tilstrekkelig andel av befolkningen er det ikke-farmasøytiske tiltak som har vært brukt til å forhindre en eksplosiv utvikling, $R > 1$. Vi hatt en lang rekke slike tiltak, nasjonale, regionale (også forskrifter for enkeltkommuner). Tiltakene har tatt form av både anbefalinger og påbud/forbud. Detaljene i denne virkemiddelbruken er det ikke mulig å representere i en empirisk framskrivningsmodell for nasjonalt nivå. Som nevnt ovenfor inneholder modellen likevel noen indikatorvariable (også kalt «dummier») som sammenfaller med det mange vil forbinde med viktige endringer i måten Covid-19 har spredd seg på i Norge.

Figur 1 viser hvordan modellen simulerer utviklingen fra oktober 2020 til slutten av mars 2021. Tallene for nye smittede inneholder mye kortsiktig variasjon. En del av dette er som nevnt «sesongvariasjon» som et resultat av mindre omfattende testing i helgene. Selv om modellen ikke fanger opp de kortsiktige variasjonene når vi simulerer den over en så lang periode som her (det er bedre håp om når vi bare simulerer en ukes tid framover) «tracker» den trenden i antall nye smittetilfeller. Dermed blir det kumulative antallet smittede også ganske nøyaktig simulert.

To variabler som det viktig å følge med på er nye sykehusinnleggelses og antallet som til enhver tid er innlagt på sykehus med Covid-19 («sykehussenger»). Det høyeste antallet innlagte hittil i pandemien var 325 (1. april 2020). Modellen følger de to pasient-trendene bra helt fram til starten av mars, da stigningen er noe brattere enn det simuleringene henger med på. I alle fall når vi starter simuleringene fem måneder tidligere.

3 R-tall

I Norge beregnes reproduksjonstallet R av FHI, ved hjelp av avansert epidemiologisk modellering. Imidlertid er ikke den norske metoden representativ for det som gjøres i andre land. I f.eks Tyskland er det nasjonale R tallet gjennomsnittet av nye tilfeller de siste fire dager, delt på gjennomsnittet de forgående fire dagene, jf. [Harvey and Kattuman \(2020\)](#). Mer generelt:

$$R_{t,r,p} = \frac{\sum_{j=0}^{p-1} NS_{t-j}}{\sum_{j=r}^{p+r-1} NS_{t-j}} = \frac{\sum_{j=0}^{p-1} NS_{t-j}}{\sum_{j=0}^{p-1} NS_{t-j}} \quad (3)$$

hvor summen i nevneren begynner på lag r og summene i teller og nevner kan overlape hverandre. Etter litt algebra kan $R_{t,r,p}$ skrives som:

$$R_{t,r,p} = 1 + r\hat{g}_{NS,t} \quad (4)$$

der $\hat{g}_{NS,t}$ er en implisitt estimator på vekstraten til NS_t . Den er gitt ved:

$$\hat{g}_{NS,t} = \frac{1}{r} \sum_{j=0}^{r-1} \frac{\sum_{j=0}^{p-1} \Delta NS_{t-j}}{\sum_{j=r}^{p+r-1} NS_{t-j}} \quad (5)$$

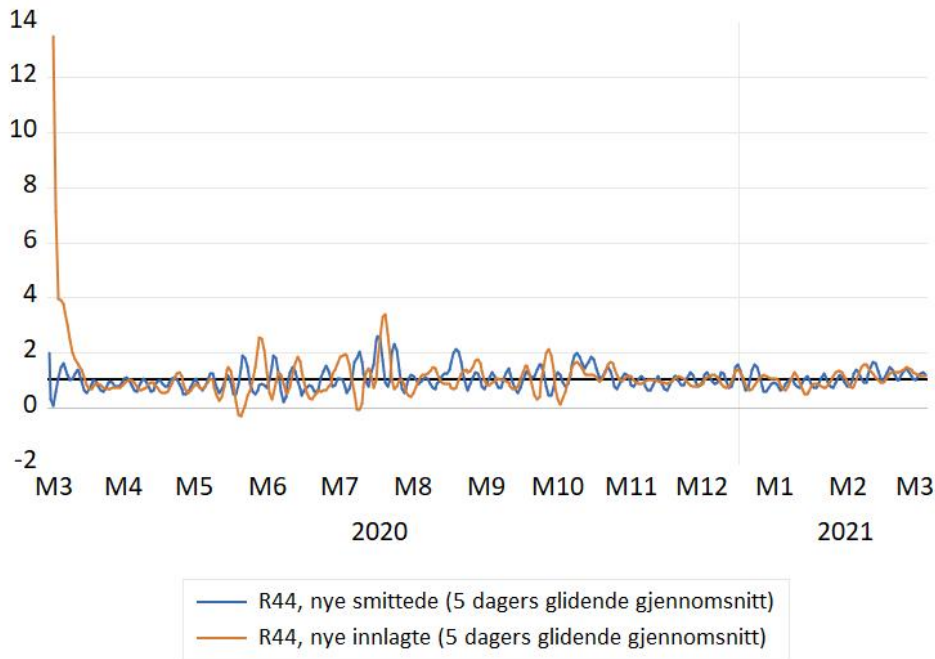
hvor $\Delta NS_t = NS_t - NS_{t-1}$. Parameteren r representerer perioden det tar før en smittet person kan bringe sykdommen videre. For Covid-19 har $r = 4$ vært ansett som et fornuftig valg. p bestemmer hvor mange lags som tas med i estimeringen av vekstraten. Det er vanlig å velge $p = 4$, og denne operasjonaliseringen referer vi til som R_{44} nedenfor.

Ved å betrakte uttrykkene ovenfor ser vi at R -tallet vil være nær 1 når den målte smitteøkningen er liten, og høy når smitteøkningen er høy. Det er derfor litt uklart om selve beregningen og rapporteringen av R -tallet kan hjelpe oss særlig mye til å si noe om utviklingen av pandemien framover. Men satt inn en mer epidemiologisk modell kan det kanskje det.

Det er også klart at det beregnede R -tallet vil variere, nærmest fra dag til dag. Dette vil spesielt være sant når det er sesongvariasjoner i smitteregistreringen og tallene er påvirket av «stop-go» politikk for å slå smitten ned samtidig som befolkningen og økonomien må få seg noen pusterom så lenge pandemien får herje uten motstand fra effektive vaksiner.

Et R -tall som beregnes ved å bruke en større p , for eksempel $p = 7$ vil bidra til å stabilere estimatet på vekstraten, og blir benevnt med R_{47} nedenfor.

Det finnes en litteratur om estimering av $R_{t,r,p}$ ved hjelp av tidsrekkeметoder, se [Harvey and Kattuman \(2020\)](#) og referansene der. Poenget i vår sammenheng er imidlertid at når vi først har en modell som sporer nye smittede, er det lett å beregne et slikt R -tall, nærmest som et biprodukt av modellarbeid som uansett må gjøres.



Figur 2: R-tall for perioden 15/3-2020 til 20/3-2021. Basert på tidsseriene NS_t og $NINL_t$, med $r = p = 4$. Glidende 5-dagers gjennomsnitt.

Det kan også beregnes R-tall basert på nye sykehusinnleggelser, $NINL_t$ i vår notasjon. I en slik beregning kan det være det være gode grunner til å sette (r, p) til andre tall enn $(4, 4)$, men for sammenlignings skyld viser Figur 2 R_{44} for både nye smittede og nye innlagte. Vi ser at R-tallet for nye smittede var 2 helt i starten av perioden (15. mars 2020). Grafen til R-tallet basert på nye sykehuspasienter starter svært høyt, nettopp fordi lå på sykehus helt i begynnelsen av epidemien. Imidlertid kommer også grafen for dette R-tallet raskt ned på et nivå som svinger rundt 1.

4 Framskrivning

Modellen kan brukes til å framskrive variablene i Figur 1 på daglig basis. Dette vil være kort-siktsprognoser som bare vil utnytte informasjonen i selve tidsseriene og «gjenværende effekter» av ikke-farmasøutiske tiltak som er blitt representert i modellen. Effekt av vaksineringsprogrammet må da selvsagt framskrives.

Den konkrete framskrivningen skjer i følgende steg.

1. På dag T oppdateres databasen med nye tall for nye smittede (NS_{T-1}), nye innlagte ($NINL_{T-1}$) og innlagte (INL_{T-1}). Det hentes også inn data for vaksineringsprogrammet, som vil bli tatt i bruk til å forbedre modellen senere i pandemien.

2. I og med at NS_t er ordnet etter testdato vil de siste observasjonene, særlig $NS_{T-1}, NS_{T-2}, NS_{T-3}$ være foreløpige tall som vil bli revidert over de nærmerste dagene. Fordi målet er å framskrive $NS_{T+h}, h = 1, 2, \dots, H$ skaper dette et sagbladproblem (velkjent for prognosemakere). Løsningen som er valgt er å oppdatere modellens koeffisienter på en utvalgsperiode som slutter i $T - 3$ og bruke modellen til å fylle inn sagbladet for NS_{T-3}, NS_{T-1}, NS_T .
3. SS_t beregnes ved akkumulering av NS_t fram til og med SS_{T-1}
4. Betinget på resultatene i steg 2 og 3, simuleres modellen dynamisk for $T, T + 1, \dots, T + H$, der H for eksempel settes til 7.

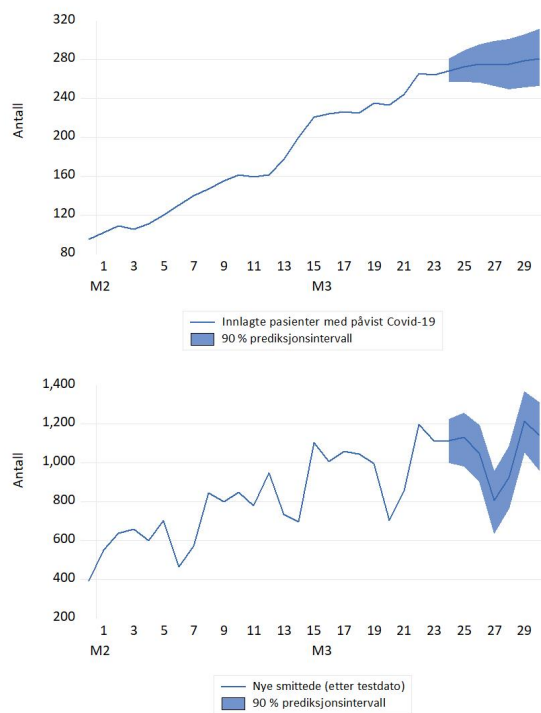
Det gjøres 5000 (stokastiske) simuleringer - Punktprognosene $\widehat{NS}_{T+h}, \widehat{SS}_{T+h}, \widehat{NINL}_{T+h}, \widehat{INL}_{T+h}$ beregnes som gjennomsnittet av de 5000 simuleringene. Fordi det er avvik fra normalitet i residualen til de estimerte ligningene trekkes restleddene til de stokastiske simuleringene ved hjelp av «bootstrapping». Dette vil redusere tendensen til at de beregnende usikkerhetsintervallene (vifte-
ne) er underestimeringer av prognoseusikkerheten (pga uholdbar forutsetning om normalfordeling).

Figur 3 viser et eksempel på prognoser for antallet nye smittede og antallet pasienter med Covid-19 som er innlagt på sykehus. Prognoseperioden er 24.-30. mars 2021 og prognosen er vist sammen med tall fra de foregående 24 dagene (merk at tallet for 20, 21. og 22. mars er «now-casted» i steg 2 i prognosemetoden som ble beskrevet ovenfor. Vi ser at prognosen er at antallet sykehuspasienter vil fortsette å øke, men det holder seg under rekorden på 325 innlagte fra februar i 2020. Sammenhengen med smittespredningen er at antallet nye smittede framskrives på et høyt nivå og fortsatt er økende, selv om en viss utflating kan anes. Effekter av de strenge (og gradvis strengere) smitteverntiltakene fra tidligere i mars er ikke presist estimert i modellen ennå. Dersom de blir sterkere i tiden framover er det håp om at prognosen er for pessimistisk.

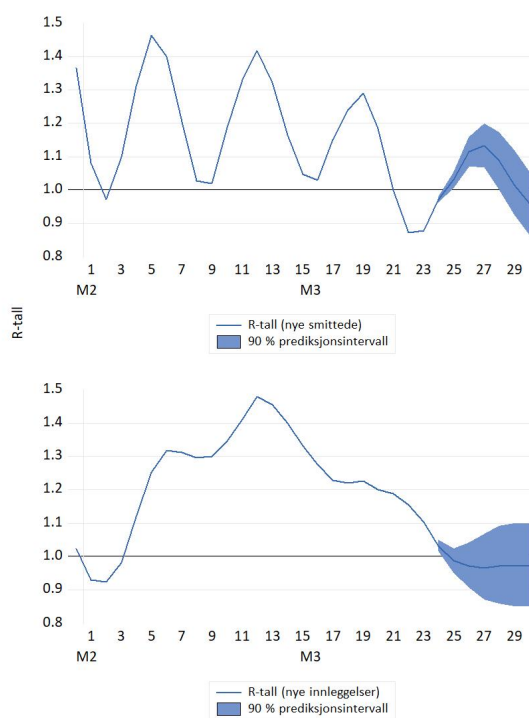
Figur 4 viser et eksempel på framskrivning av R-tall som er basert på antall nye smittede og på antallet nye pasienter innlagt på sykehus. Begge R-tall har stort sett vært over +1 i mars, og dette er også tilfellet for de framskrevne tallene. Men +1 ligger til dels innenfor fanen som angir prediksjonsintervallene.

5 FHIs prognoser fra uke 11 og uke 12

Tabell 1 viser prognosene for nye tilfeller og innlagte som ble publisert av Folkehelseinstituttets [modelleringsrapport](#) fra 24/3-2021. FHI presenterer prognoser hele tre uker fram i tid. Siden de tidsrekkebaserte prognosene bare går 7 dager framover, kan vi bare sammenligne med FHIs prognoser for 28. mars som er ca 2500 nye tilfeller mens tidsrekke modellen (Figur 1) gir «bare» 714. Noe av denne forskjellen kan skyldes at modellen framskriver tallene slik de er registrert (etter



Figur 3: Nye smittede med Covid-19 (etter testdato) og antallet pasienter innlagt på sykehus. Faktiske tall (27/2- 23/3 2021) og prognoser for perioden 24/3-30/3 2021.



Figur 4: R-tall basert på nye tilfeller og nye sykehusinnlagte med Covid-19. Beregnede R-tall (27/2-22/3 2021) og prognoser for perioden 24/3-30/3 2021. Grafene viser 5 dagers glidende gjennomsnitt.

testdato) FHIs framskrivninger ser ikke ut til å innholde særlig sesongvariasjon. Siden 28. mars er en søndag, og dersom kan dette forklare noe av forskjellen. Dersom vi tar et fem dagers sentret gjennomsnitt av tidsrekkeprognosen blir anslaget for 28. mars 900 nye smittede. Det er verdt å merke seg, slik det er anført i Tabell 1, at FHIs prognoser for 21 mars var omlag 2000 nye tilfeller, det vil si omtrent det doble av det faktiske tallet (Figur 5 i modelleringsrapporten). Prognosene for antall pasienter innlagt på sykehus er deet større sammenfall. Der har FHIs modelleringsrapport ca 260, mens tidsrekkeprognosen i Figur ga 293. Tabell 2 viser FHI prognoser fra uke 12. Vi

Tabell 1: Utdrag fra Table 3 i *Situational awareness and forecasting for Norway FHI COVID-19 modelling team Week 11, 24 March 2021*

Publiseringsdato	Første dag	Én uke fram, 28. mars	2 uker, 4. april	3 uker 11. april
24. mars	22. mars	Median/Gj.snitt (PI)	Median/Gj.snitt (PI)	Median/Gj.snitt (PI)
Nye tilfeller	2000*	2561/2506 (1791-3872)	3235/3129 (2063-5250)	4080/3884 (2349-7132)
Innlagte	210*	261/260 (196-334)	333/332 (232-445)	417/409 (283-628)

* Avlest fra grafen for “Daglig Insidens” i Figure 5.

ser at prognosen for 28. mars da var blitt nedjustert med en faktor på 0,6, 1490. Det er et bra egenskap ved en prognemodell at den responderer på ny informasjon (oppdaterte parametre og nye initalbetingelse). Likevel var justeringen langt fra stor nok, siden antallet nye smittede som var testet på denne dagen (som var Palmesøndag) var 473. FHIs prognoser for antallet innlagte («sykesenger») for 28. mars ble ikke endret så mye fra uke 11 til uke 12. Den var fortsatt cirka 260, som er bare var en titalls «sykesenger» unna det faktiske tallet på 273.

Tabell 2: Utdrag fra Table 3 i *Situational awareness and forecasting for Norway FHI COVID-19 modelling team Week 12, 31 March 2021*

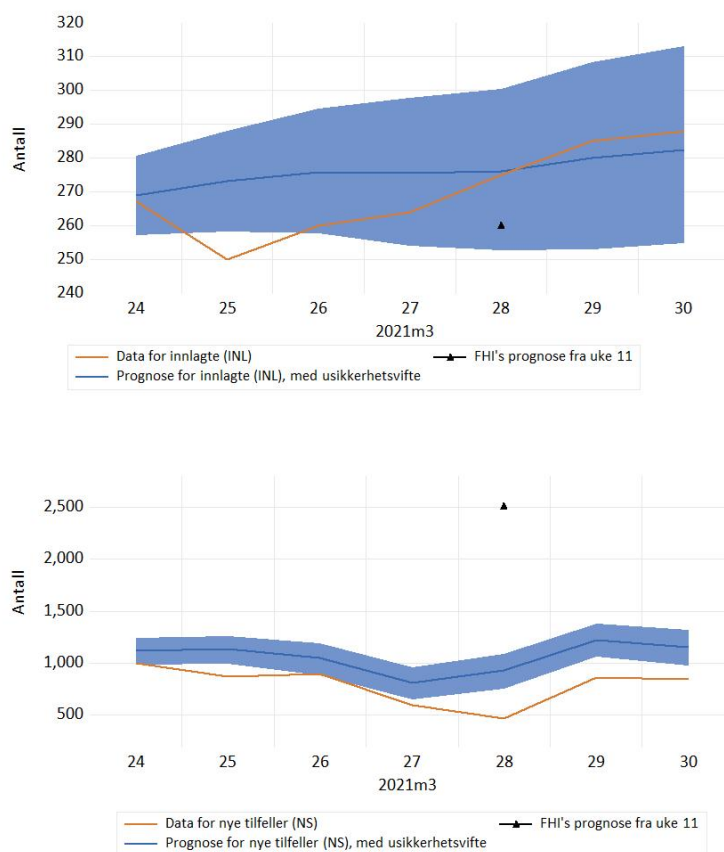
Publiseringsdato	Første dag	Én uke fram, 4. april	2 uker, 11. april	3 uker 18. april
31. mars	28. mars	Median/Gj.snitt (PI)	Median/Gj.snitt (PI)	Median/Gj.snitt (PI)
Nye tilfeller	1490*	1491/1446 (1039-2033)	1534/1478 (1007-2232)	1578/1516 (923-2391)
Innlagte	260*	272/270 (217-336)	288/288 (208-373)	301/302 (214-411)

* Avlest fra grafen for “Daglig Insidens” i Figure 5.

6 Prognoseevaluering

Prognosen i figur 3 viste seg å overvurdere antall nye smittede (NS) i 7-dagersperioden 24-30. mars. I figur 5 er grafen for prognosen vist sammen med grafen for faktisk antall nye smittede, NS . Vi ser at overvurderingen er vesentlig mindre i den tidsrekkebaserte prognosen enn i FHI’s prognose i *Situational awareness and forecasting for Norway FHI COVID-19 modelling team Week 11, 24*

March 2021—FHI anlaget for 28. mars er angitt med et triangelsymbol. Når det gjelder antall



Figur 5: Prognosene vist i figur 3 sammen med faktiske tall (data). FHI's prognoser fra uke 11 for 28. mars angitt med triangelsymbol

pasienter innlagt på sykehus med påvist Covid-19, viser tidsrekkemodellens prognoser en viss undervurdering i starten av 7-dagers perioden, og en liten overvurdering mot slutten. Med unntak for 25. mars er ikke feilene signifikante (data ligger innenfor usikkerhetsviften).

Siden prognosen oppdater alle virkedager, er det mulig å lage tilsvarende grafer for alle prognoserunder (ettthvert som faktiske data blir tilgjengelig). Dette vil ganske sikkert gi eksempler på prognoser som både treffer bedre, men også dårligere, enn prognosen fra 24. mars. For å foreta en formell statistisk evaluering av prognosene må det først «samles opp» datasett med éndagsprognoser, todagsprognoser og så videre. Når utvalgene er blitt store nok kan prognoseskjevheten («bias») estimeres, og det kan gjøres statistisk sammenligning av nøyaktighet med konkurrerende prognosemodeller. Slike tester vil bli inkludert i en senere versjon av notatet.

Referanser

Harvey, A. C. and P. Kattuman (2020). A Farewell to R: Time series models for tracking and forecasting epidemics. *Covid Economics*, (51), 36–73.

Nymoen, R. (2019). *Dynamic Econometrics for Empirical Macroeconomic Modelling*. World Scientific, Boston.