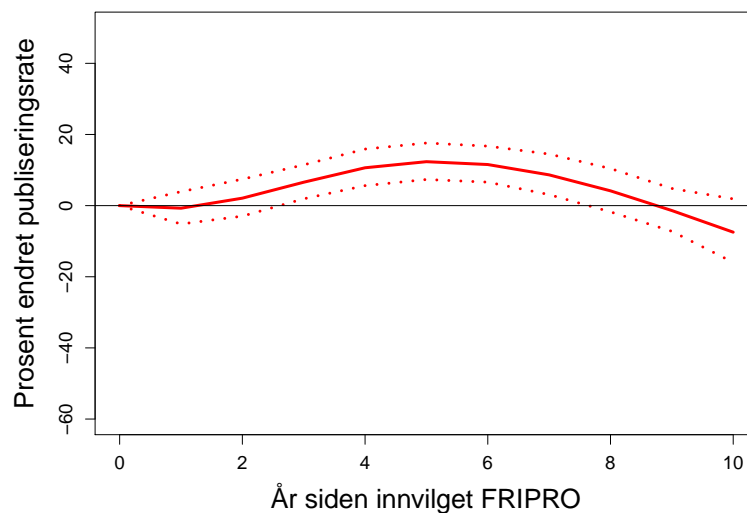


Analyse av hvilken effekt tildeling av FRIPRO-midler har på forskernes vitenskapelige produksjon

Effekt av tid siden innvilget FRIPRO



Notatnr
Forfattere

SAMBA/45/17
Olav Nikolai Breivik
Magne Aldrin

Dato

24. januar 2018

Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Kristen Nygaards hus ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistisk-matematisk modellering og har et senter for forskningsdrevet innovasjon, Big Insight, med finansiering fra Norges forskningsråd, bedrifter og offentlige partnere. Innen statistikk jobbes det med et bredt spekter av problemstillinger, for eksempel finansiell risiko, jordobservasjon, estimering av fiskebestander, helse og beskrivelse av geologien i petroleumsreservoarer. NR er ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Innen IKT-området har NR innsatsområdene e-inkludering, informasjonssikkerhet og smarte informasjonssystemer.

NRs visjon er forskningsresultater som brukes og synes.

Tittel

Forfattere

Olav Nikolai Breivik <olavbr@nr.no>
Magne Aldrin <Magne.Aldrin@nr.no>

Dato

24. januar 2018

Publikasjonsnummer

SAMBA/45/17

Sammendrag

Denne rapporten oppsummerer en studie utført av Norsk Regnesentral (NR) på oppdrag fra Norges forskningsråd (NFR). Studien undersøker i hvilken grad en forsker, eller nærmere bestemt en prosjektleder, som får tildelt et Fri prosjektstøtte (FRIPRO)-prosjekt får en økt eller bedret vitenskapelig produksjon under eller i etterkant av prosjektet i forhold til en sammenlignbar forsker som har søkt, men ikke fått tildelt et FRIPRO-prosjekt.

I forhold til en normal årlig publiseringsrate for en sammenlignbar forsker får en forsker som har fått tildelt et FRIPRO-prosjekt en økt publiseringsrate fra og med to år etter tildeling med en topp mot omkring 12% økt publiseringsrate fem år etter tildeling. Deretter blir effekten mindre, og åtte år etter tildeling er den estimerte effekten 0.

Når det gjelder årlig siteringsrate per artikkel får en forsker som har fått tildelt et FRIPRO-prosjekt en økt siteringsrate etter tildeling i forhold til en normal siteringsrate for en sammenlignbar forsker. Åtte år etter tildelingsåret utgjør dette omkring 10% høyere siteringsrate.

Emneord

Målgruppe

NFR

Tilgjengelighet

Åpen

Prosjekt

FRIPRO-analyse

Prosjektnummer

Satsningsområde

Publikasjons- og siteringsanalyse

Antall sider

36

© Copyright

Norsk Regnesentral

Innhold

1	Utvidet sammendrag	5
2	Introduksjon	9
3	Data	10
4	Metode	14
4.1	Modell for publiseringsrate	14
4.2	Modell for siteringsrate	17
4.3	Alternativ, forenklet modell for publiseringsrate	19
4.4	Estimering	19
5	Resultater	20
5.1	Publikasjonsanalysen	20
5.2	Siteringsanalysen	27
6	Diskusjon	34
	Referanser	36

1 Utvidet sammendrag

Forskningsrådets ordning for fri prosjektstøtte (FRIPRO) er en åpen, nasjonal konkurransearena som favner alle fag og disipliner, uten at det legges tematiske føringer eller settes krav om anvendelse og umiddelbar nytteverdi. FRIPRO skal fremme vitenskapelig kvalitet i internasjonal forskningsfront, dristig og nyskapende forskning, karriere for unge forskertalenter og mobilitet for forskere tidlig i sin karriere. FRIPRO er et av Forskningsrådets sentrale virkemidler for å fremme fri, forskerinitiert grunnleggende forskning og har i sin nåværende form, organisert samlet innenfor Divisjon for vitenskap, eksistert siden utlysningen i 2004. FRIPRO er delt inn i tre fagkomiteer. De tre fagkomiteene er FRINATEK (matematikk, naturvitenskap og teknologi), FRIMEDBIO (medisin, helse og biologi) og FRIHUMSAM (humaniora og samfunnsvitenskap). Søknadene inkludert i denne studien tilhører en av de tre fagkomiteene. I tillegg er det inkludert noen få søknader fra 2004 av typen Store forskerinitierte prosjekter (STORFORSK), som ble behandlet under FRIPRO.

Formålet med denne studien har vært å måle hvilken effekt tildeling av FRIPRO har på de tildelte forskernes vitenskapelige produktivitet og kvalitet. Dette innebærer ikke bare å måle de direkte resultatene i form av vitenskapelige publikasjoner generert i de konkrete prosjektene, men å se om slik tildeling har langsiktig effekt på forskernes framtidige vitenskapelige resultater. To typer analyser er utført: En som måler effekt i vitenskapelig produktivitet i form av antall fagfellevurderte vitenskapelige publikasjoner, og en som måler bruk (som uttrykk for kvalitet) gjennom siteringer av slike. Det er mange faktorer utover tildeling av en bestemt type forskningsmidler som potensielt påvirker forskernes vitenskapelige resultater. Hensikten med denne analysen har vært å prøve å isolere effekten av tildelinger fra FRIPRO fra disse øvrige faktorene ved å sammenlikne forskere med og uten slik tildeling.

Studien er basert på data for 2553 forskere som har søkt et forskerprosjekt på minst 2,5 millioner kroner i perioden 2004-2012. Data fra nyere søknader er ikke utnyttet da vi har liten historisk informasjon etter nyere prosjekter. Av disse fikk 562 innvilget minst ett FRIPRO-prosjekt. Dataene er samlet inn fra registre uten medvirkning fra de enkelte forskerne. I publiseringsanalysen har vi basert oss på publikasjonsdata registrert i Current research information system in Norway (CRISTIN) for den enkelte forsker. Vi har for hver forsker inkludert årlig antall publikasjoner i kategorier som typisk oppnår publikasjonspoeng i Norsk vitenskapsindeks (NVI). Det innbefatter bl.a. kategoriene "Vitenskapelig artikkel" og "Vitenskapelig monografi".

Vi har analysert årlig publiseringsrate for hver enkelt forsker og korrigert for faktorer som

- at forskeren er registrert i Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU) sitt forskerpersonalregister (hvis ikke kan forskeren være pensjonert eller ikke lenger være ansatt i en forskningsinstitusjon),
- kjønn,

- alder,
- alder ved avlagt doktorgrad,
- forskerens fagfelt og hvordan publiseringsraten generelt er endret over tid innen fagfeltet,
- type forskningsinstitusjon forskere er ansatt ved, og hvordan den registrerte (i CRIS-TIN) publiseringsraten er endret over tid for den enkelte type institusjon,
- en effekt for den enkelte forsker, slik at forskeren fungerer som sin egen kontroll.

Basert på faktorene over har vi modellert hvordan forventet årlig publikasjonrate vil variere over tid for en forsker i løpet av karrieren. I tillegg har vi modellert effekten av å få tildelt et FRIPRO-prosjekt som en prosentvis endring i publikasjonraten *utover normal publiseringsrate* i tida etter tildeling av et FRIPRO-prosjekt. Resultatene er vist i figur 1, hvor innvilgelsesåret refererer til året før prosjektstart. Fra og med tre år etter innvilgelsesåret er det en økning i publiseringsraten med en topp mot omkring 12% økt publiseringsrate fem år etter tildeling. Deretter blir effekten mindre, og åtte år etter tildeling er den estimerte effekten 0. Hvis vi antar at et prosjekt varer i tre eller fire år, dvs. slutter i år fire eller fem, tyder resultatene på at tildeling av et FRIPRO-prosjekt er assosiert med en økning av publiseringsraten under og i noen år i etterkant av prosjektet, men at det ikke er noen langtidseffekt som varer videre i karrieren.

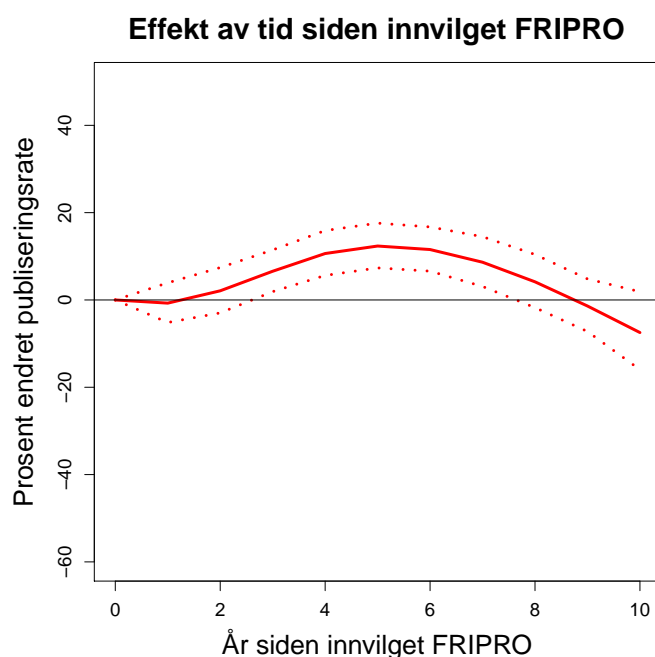
For å kvantifisere hva dette har å si totalt sett har vi for de 562 forskerne som fikk tildelt et FRIPRO-prosjekt beregnet deres forventede antall publikasjoner med et FRIPRO-prosjekt, og sammenlignet det med forventede antall publikasjoner for de samme forskerne om de ikke hadde fått et FRIPRO-prosjekt. Beregningen viser at 1275 ekstra artikler (95% konfidensintervall fra 414 til 2111) kan tilskrives tildeling av FRIPRO-prosjektene. Det tilsvarer i snitt 0.27 (0.09, 0.46) flere publikasjoner per år per forsker i perioden etter FRIPRO-tildeling.

Ved å dele effekten av å få tildelt FRIPRO inn i de enkelte fagkomiteer, finner vi at meste parten av FRIPRO-effekten på publiseringsrate kan tilskrives prosjekter innenfor FRINATEK.

Som et mål på kvalitet har vi analysert årlige antall siteringer per artikkel publisert av den enkelte forsker. Denne siteringsanalysen er basert på publikasjoner i kategorien "Article" registrert i publikasjonsdatabasen Web of Science (WoS). Det gjør at analysen blir noe begrenset. Blant annet blir det ikke tatt hensyn til monografier og lignende publikasjoner. Videre er det kjent at WoS har hatt dårligere dekning for tidsskrifter innen humaniora og samfunnsfag enn medisin, matematikk, naturvitenskap og teknologi.

I modellen for årlig siteringsrate per artikkel skrevet av den enkelte forsker har vi korrigert for faktorer som

- kjønn,
- tid siden publikasjonsår,



Figur 1. Estimert effekt på publiseringsrate av tildeling av FRIPRO-midler som funksjon av tid siden innvilgelsesår, hvor stiplet linje viser 95% konfidensintervall.

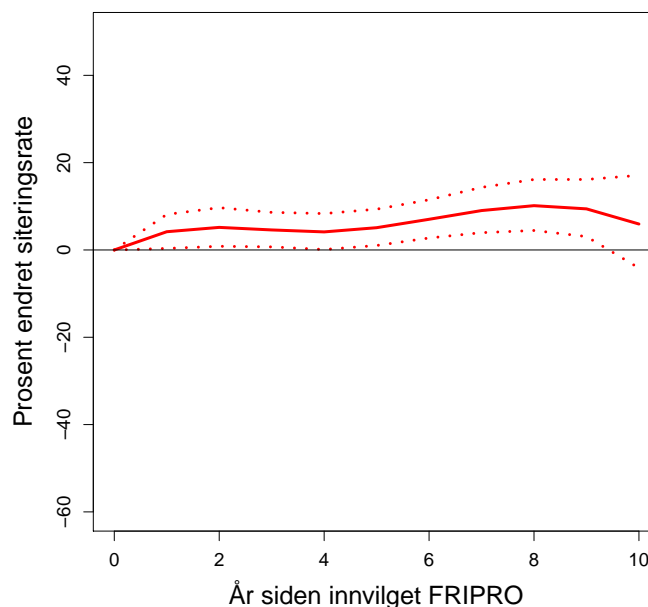
- alder,
- alder ved avlagt doktorgrad,
- forskerens fagfelt og hvordan publiseringsgraden generelt er endret over tid innen fagfeltet,
- artikkelens fagfelt,
- en effekt for den enkelte forsker, slik at forskeren fungerer som sin egen kontroll,
- en effekt for den enkelte artikkel, for å ta høyde for at enkelte artikler siteres ofte og andre sjeldent.

Vi har så modellert effekten av å få tildelt et FRIPRO-prosjekt som en prosentvis endring i siteringsraten *utover normal siteringsrate* i tida etter tildeling av et FRIPRO-prosjekt. Figur 2 viser at det er en signifikant økt siteringsrate for artikler publisert etter tildeling av et FRIPRO-prosjekt. Allerede for en artikkel publisert ett år etter tildeling er det en forventet høyere siteringsrate på omkring 5%, og denne er økt til omkring 10% for artikler publisert åtte år etter tildelingsåret. Arbeidet til en artikkel publisert året etter tildeling vil imidlertid vanligvis være utført i forkant av prosjektet, slik at det er vanskelig å fortolke denne delen av kurven som effekt av en FRIPRO-tildeling.

Ved å dele effekten av å få tildelt FRIPRO inn i de enkelte fagkomiteer, finner vi også for siteringsraten at mesteparten av FRIPRO-effekten kan tilskrives prosjekter innenfor FRINATEK.

I analysene vi har utført har vi så langt vi har kunnet korrigert for faktorer som kan tenkes

Effekt av tid fra innvilget FRIPRO til publisering



Figur 2. Estimert effekt på siteringsrate av tildeling av FRIPRO-midler som funksjon av tid fra innvilgelsesår til publikasjonsår for en artikkel som er publisert i etterkant av FRIPRO-tildeling. Stiplet linje viser estimert 95% konfidensintervall.

å påvirke en forskers publiseringrate og siteringsraten per artikkel, for på den måten å isolere effekten av å få en FRIPRO-tildeling fra de andre faktorene. Imidlertid er det en del momenter en bør ha i bakhodet når en fortolker resultatene.

For det første er det slik at observert assosiasjon ikke nødvendigvis impliserer en kausal sammenheng. Videre er det slik at vi har analysert effektene for kun prosjektlederen i søknaden. Selv om vi gjennomgående har brukt ordet forsker, gjelder altså analysen ikke for øvrige forskere som har deltatt i prosjektene. Til sist er det verdt å huske på at de fleste forskere har alternative finansieringskilder. En forsker som får avslag på en FRIPRO-søknad vil kunne få midler fra annet hold, enten til et identisk eller lignende prosjekt, eller til andre typer prosjekter. Dermed er det bare tilleggseffekten utover alternativ finansiering vi har kunnet tallfeste. Da analysen kun omhandler prosjektlederne, og det eksisterer andre finansieringskilder, vil trolig den totale effekten av et FRIPRO-prosjekt være større enn det vi har tatt sikte på å estimere i vår analyse. Vi ønsker dermed å påpeke at resultatene i analysen ikke kan direkte brukes til å måle nytten av FRIPRO.

2 Introduksjon

Denne rapporten oppsummerer en studie utført av Norsk Regnesentral (NR) på oppdrag fra Norges forskningsråd (NFR). Studien undersøker i hvilken grad en forsker som får tildelt et forskerprosjekt i FRIPRO får en endret vitenskapelig produksjon under eller i etterkant av prosjektet i forhold til en sammenlignbar forsker som har søkt, men ikke fått tildelt et slikt prosjekt. Vi har kun opplysninger om prosjektlederen som søkte, slik at resultatene ikke nødvendigvis gjelder for øvrige forskere som har arbeidet i, eller ville ha arbeidet i prosjektene. Når vi bruker betegnelsen forsker er det dermed en prosjektleder som menes. Vi utfører separate analyser av publiseringsrater og siteringsrater, videre kalt henholdsvis publikasjonsanalysen og siteringsanalysen, til prosjektledere som søkte FRIPRO i perioden 2004-2012. Publiseringsraten er antatt å være et mål på forskernes produktivitet, og siteringsraten et mål på forskningens kvalitet. Publikasjonsanalysen undersøker dermed om FRIPRO-støtte endrer forskerens produktivitet, mens siteringsanalysen undersøker om FRIPRO-støtte endrer forskningens kvalitet.

Hovedfokus for analysene var opprinnelig å undersøke langtidseffekten av det å få tildelt FRIPRO-midler, dvs. av det arbeidet forskeren utfører etter at et eventuelt FRIPRO-prosjekt er over, og som dermed typisk vil bli publisert flere år etter at et prosjekt er ferdig. Imidlertid er tidsperioden for de tilgjengelige dataene noe kort til at vi kan uttale oss bastant om effekter lang tid etter prosjektslutt, og vi har derfor endret fokus til å gjelde hele tidsforløpet fra eventuell oppstart av et prosjekt til noen år etter antatt prosjektslutt.

Dataene som er brukt i rapporten er samlet inn fra flere databaser, uten at hver enkelt forsker har aktivt bidratt med informasjon. Følgende databaser er utnyttet: NFR sin database med informasjon om FRIPRO-søknader, NIFU (Nordisk institutt for studier av innovasjon, forskning og utdanning, <https://www.nifu.no/>) sitt forskerpersonalregister, CRISTIN (Current Research Information System in Norway, <http://www.cristin.no/>), og Web of Science (WoS), en internasjonal publikasjonsdatabase eid av Clarivate Analytics (<https://clarivate.com/>). Detaljert informasjon om data er gitt i seksjon 3.

I publikasjonsanalysen modellerer vi hvordan antall publikasjoner per år utvikler seg i løpet av dataperioden for den enkelte forsker. Modellen tar hensyn til i) kjønn, ii) forskerens alder det enkelte år og tid siden avlagt doktorgrad, dvs. hvor forskeren befinner seg i karrieren, iii) hvilket fagfelt forskeren tilhører og hvordan den rapporterte publiseringsraten innen fagfeltet har endret seg over tid og iv) hvilken type institusjon forskeren arbeider på og hvordan den rapporterte publiseringsraten for denne institusjonstypen har endret seg over tid, samt en individuell effekt for den enkelte forsker. Dette gjøres for å korrigere for flest mulige effekter som påvirker en forskers publikasjonrate. Vi har med en tidsvarierende effekt av det å ha fått tildelt et FRIPRO-prosjekt, hvor vi modellerer forskjellen i publiseringsrate mellom en forsker som har fått tildelt et FRIPRO-prosjekt og en sammenlignbar forsker (ved at modellen korrigerer for ulike faktorer) som ikke har fått tildelt et slikt prosjekt. Denne forskjellen beskrives som en funksjon over tid siden tildelt FRIPRO-prosjekt. Denne FRIPRO-effekten dekomponeres videre i de fire katego-

riene, FRINATEK, FRIMEDBIO, FRIHUMSAM og STORFORSK.

I siteringsanalysen modellerer vi hvordan antall siteringer per år for den enkelte publisering endres over tid etter publisering. Igjen korrigerer vi for en rekke faktorer knyttet til den enkelte forsker, men i tillegg korrigerer vi for en del faktorer knyttet til den enkelte artikkel. Vi har med en tidsvarierende effekt av det å ha fått tildelt et FRIPRO-prosjekt, hvor vi modellerer forskjellen i siteringsrate mellom en forsker som har fått tildelt et FRIPRO-prosjekt og en sammenlignbar forsker som ikke har fått tildelt et slikt prosjekt. Som for publikasjonanalysen beskrives denne forskjellen som en funksjon over tid siden tildelt FRIPRO-prosjekt. Også FRIPRO-effekten for siteringer dekomponeres i de fire kategoriene.

Begge analysene nevnt over studerer endringer over tid siden en eventuell tildeling av et FRIPRO-prosjekt. For publiseringsraten har vi utført ytterligere en analyse, basert på en forenklet problemstilling og en forenklet modell. Vi studerer da antall publikasjoner for den enkelte forsker i gjennomsnitt over det sjette og sjuende året etter at FRIPRO-prosjektet har startet, eller var tenkt startet hvis forskeren ikke fikk innvilget FRIPRO-søknad. Antar vi at et FRIPRO-prosjekt varer i tre år, vil dette være tredje og fjerde året etter avsluttet prosjekt. Det vil da være rimelig å anta at publikasjoner i dette tidsintervallet er publikasjoner basert på forskning utført i etterkant av et eventuelt FRIPRO-prosjekt. For et prosjekt med seinere oppstart eller et fire-årig prosjekt vil derimot publikasjonene også kunne omfatte forskning som er utført i løpet av prosjektet, fordi det tar tid å publisere vitenskapelig arbeid. Antall publikasjoner i etterkant av prosjektet blir så modellert som funksjon av antall publikasjoner i forkant av prosjektet og faktorer knyttet til den enkelte forsker, samt om forskeren fikk tildelt et FRIPRO-prosjekt eller om han eller hun søkte, men fikk avslag. Hensikten med denne analysen er å undersøke om vi ved en annen og enklere tilnærming får tilsvarende resultater som den mer komplekse og detaljerte analysen beskrevet tidligere.

Detaljert beskrivelse av modellene og annen metodikk er gitt i seksjon 4. Resultatene av analysen er gitt i seksjon 5. Til slutt følger en diskusjon av fortolkning av resultatene i seksjon 6.

3 Data

Av NFR fikk vi utlevert en liste med informasjon om alle søknader til forskerprosjekter i FRIPRO med søkt beløp over 2.5 millioner kroner i perioden 2004-2012. Det utgjør 5346 søknader, med 2650 unike prosjektledere. Se tabell 1 for en kort beskrivelse av data fra NFR.

Basert på listen fra NFR, leverte vi en liste med navn på forskerne, deres fødselsdato og tilhørighet ved søknadstidspunkter til NIFU. Av de 2650 unike prosjektlederne ble 2622 funnet i NIFU sitt forskerpersonalregister. Se tabell 2 for en kort beskrivelse av data utnyttet fra forskerpersonalregisteret. De 28 forskerne som ikke ble funnet i forskerper-

Data fra NFR, for hver enkelt søknad 2004-2012	Beskrivelse
Fødselsdato	Årstall, måned og dag, mangler for 371 forskere
Søkt prosjektperiode	Fra år og til år
Kategori	FRINATEK/FRIMEDBIO/FRIHUMSAM/STORFORSK
Utfall søknad	Avslag/tildelt støtte
Navn på prosjektleder	Etternavn og fornavn
Kjønn	Mann/kvinne

Tabell 1. *Beskrivelse av data mottatt av NFR.*

Data fra NIFU, for hver enkelt forsker	Beskrivelse
Navn	Fornavn og etternavn
Fødselsdato	Årstall, måned og dag
Institusjonstilhørighet hvert enkelt år	F.eks. instituttsektor, helsesektor eller UiO
Tid for oppnådd doktorgrad	Årstall
Fagfelt hvert enkelt år	Humaniora, samfunnsvitenskap, matematikk/naturvitenskap, teknologi, medisinske fag, helsefag, administrasjon og Landbruksfag/fiskerifag/veterinærmedisin

Tabell 2. *Beskrivelse av data fra NIFU sitt forskerpersonalregisteret som er utnyttet i analysen. Informasjonen er registrert annen hvert år i perioden 2001 til 2007, og hvert år i perioden 2007-2015. Tid for oppnådd doktorgrad manglet for 4.3% av forskerne.*

sonalregisteret ble tatt ut av analysen.

Basert på data vi fikk av NIFU, leverte vi en liste med forskernes navn og fødselsdato til CRISTIN. I retur fikk vi informasjon om e-postadresser registrert for hver enkelt forsker, og deres vitenskapelige arbeid registrert i CRISTIN. Arbeidet registrert i CRISTIN utgjør hovedgrunlaget for artikler brukt i publikasjonsmodellen. Totalt ble 2588 av de 2622 forskerne fra forrige trinn identifisert i CRISTIN. De øvrige 34 forskerne ble tatt ut av analysen. Hver forsker hadde rett til å reservere seg fra undersøkelsen. Vi informerte alle forskerne om reservasjonsretten gjennom epost. Det var da 35 forskere som reserverte seg, og det er dermed inkludert 2553 forskere i undersøkelsen, dvs. 96% av de opprinnelige 2650 forskerne i utvalget.

Vi ønsker at datagrunlaget for analysene skal bestå av publikasjoner av relativt høy vitenskapelig verdi. For å oppnå et slikt datagrunlag utnyttet vi kun data fra CRISTIN med hovedkategorier og underkategorier som typisk oppnår publikasjonspoeng i Norsk vitenskapsindeks (NVI). Merk at vi kun har NVI-data fra og med år 2011. Ved å undersøke andelen publikasjoner som fikk NVI-poeng etter 2011, endte vi opp med å basere datagrunlaget vårt på publikasjoner som var: i) under hovedkategori "Tidsskriftspublikasjon" med underkategori "Vitenskapelig artikkel" eller "Vitenskapelig oversiktsartikkel", eller ii) under hovedkategori "Bok" med underkategori "Vitenskapelig monografi", eller iii) under hovedkategori "Del av bok/rapport" med underkategori "Vitenskapelig kapittel/artikkel/conference proceedings". Vi tok med publikasjoner på alle språk, men norsk og engelsk utgjør mer enn 98% av disse. Dette utgjør 100 909 unike publikasjoner, men 129 951 publikasjoner med gjentakelser, fordi en publikasjon kan være forfattet av

Data fra CRISTIN, for hver enkelt publikasjon knyttet til den enkelte forsker	Beskrivelse
Navn på forfatter	Fornavn og etternavn
Hovedkategori	Se http://www.cristin.no
Underkategori	Se http://www.cristin.no
Årstall publikasjon	Årstall
Tittel	Tittel til artikkel/arbeid
Språk	Engelsk/norsk
DOI	DOI til 59% av publikasjonene
E-postadresser	Opp til fem epostadresser per forsker
NVI	Informasjon om publikasjon ga publikasjonspoeng (kun gitt fra år 2011)

Tabell 3. *Beskrivelse av data fra CRISTIN.*

flere av forskerne i undersøkelsen.

Vi utleverte videre til Clarivate en liste med informasjon om de 2588 forskerne funnet i CRISTIN. Følgende informasjon ble utlevert til Clarivate: e-postadresser, historiske arbeidssteder (uten tidshenvisning) og tidligst aktuelle tidspunkt for første publikasjon for den enkelte forsker. Årstallet for tidligst aktuelle publikasjon er definert som året forskeren fylte 24 år. Vi utleverte en slik variabel, uten informasjon om definisjonen, i stedet for fødselsdato for å minimere utlevering og kobling av informasjon. Clarivate knyttet opp forskerne til deres database brukt i WoS. Se tabell 4 for informasjon om data fra WoS. Dataene fra WoS brukes i publikasjonsanalysen til å supplere publikasjonene fra CRISTIN, og utgjør hele datagrunnlaget for siteringsanalysen. For WoS har vi bare plukket ut publikasjoner i kategorien "Article". Dermed er siteringsanalysen basert kun på artikler.

Vi er sikre på at publikasjoner i CRISTIN-data er skrevet av forskerne av interesse. Når det gjelder artikler i WoS-data er vi derimot i utgangspunktet mer usikre på om de er skrevet av forskerne vi studerer eller av noen med lignende navn eller initialer. Vi knyttet artikler i WoS til de aktuelle forskerne med følgende prosedyre: For *ite* forsker fant vi alle artikler knyttet til forskeren i WoS som og var knyttet til forskeren i CRISTIN. Denne sammenligningen ble gjort ved å bruke DOI-informasjon og ved å sammenligne titlene til artiklene. Vi oppdaget at titlene til artiklene kunne være stavet litt annerledes i CRISTIN og WoS, på grunn av denne mulige ulikheten godtok vi mindre variasjoner i titlene med bruk av R-funksjonen *agrepl*. Deretter fant vi alle artikler i WoS-data hvor relevante forsker var knyttet med e-postadresse til samme forsker i CRISTIN. Totalt endte vi opp med 75 379 artikler i WoS-data vi er sikre på at vi har knyttet til riktig forsker (totalt 54 065 unike artikler).

Datagrunnlaget i publikasjonsanalysen består av 2553 forskere og 130 351 publikasjoner, medregnet gjentakelser. De fleste av disse publikasjonene er registrert i CRISTIN, men de er supplert med 400 ekstra funnet i WoS. Utgivelsesåret for hver publikasjon tas i bruk slik at datagrunnlaget for hver forsker består av årlig antall publikasjoner hvert enkelt år i dataperioden, samt relevant tilleggsinformasjon, deriblant år for søkt og eventuelt innvilget FRIPRO-søknad. Vi har informasjon om første planlagte prosjektår, og vi antar

Data fra WoS, for hver enkelt publikasjon	Beskrivelse
Navn på forfatter	Fornavn eller initialer og etternavn
Årstall for publikasjon	Årstall
Årlige siteringer	Årlige siteringer i perioden f.o.m. 2000 t.o.m. 2016.
Fagfelt	WoS's inndeling av fagfelt, totalt 250 ulike fagfeltnavn.
Tittel	Tittel til artikkel
DOI	DOI til 88% av artiklene
Score	Indikerer på hvor sikker WoS er at artikkelen hører til den gitte forsker.
Kluster ID	Hvert kluster representerer en mengde av artikler WoS mener er sannsynlig skrevet av samme forsker.
E-postadresse	Eventuell e-postadresse knyttet mellom forsker og artikkel

Tabell 4. *Beskrivelse av data fra WoS. Totalt ble det funnet 1 558 406 potensielle artikler i WoS medregnet gjentakelser (1 051 260 unike artikler).*

søknaden er innvilget året før og betegner dette året med innvilgelsesåret. Grunnen til at vi inkluderer innvilgelsesåret er at dette året brukes som referanse når vi kvantifiserer effekten av FRIPRO. Det er 562 av forskerne som har blitt tildelt ett eller flere FRIPRO-prosjekter. I alt har disse blitt tildelt 705 prosjekter, men i analysene bruker vi kun første FRIPRO-tildeling for hver forsker. Disse 562 tildelingene fordeler seg med 158 på FRI-NATEK, 157 på FRIHUMSAM, 232 på FRIMEDBIO og 15 på STORFORSK.

Datagrunnlaget for siteringsanalysen består av 68 954 publikasjoner (inkludert gjentakelser), publisert i 2015 eller tidligere og årlige antall siteringer av hver av disse etter publiseringsåret, eller med andre ord 68 954 siteringstidsserier som begynner i år 2016 eller tidligere. Til sammen består tidsseriene i datagrunnlaget for siteringsanalysen av 456 100 år. For hver publikasjon eller siteringstidsserie bruker vi relevant tilleggsinformasjon om den enkelte forsker som i publikasjonsanalysen, men vi bruker i tillegg informasjon om den enkelte artikkel, slik som fagfelt.

I data fra WoS er det gitt en score til hver artikkel, som beskriver hvor sikre WoS mener de er på at den er skrevet av den gitte forsker. Score-variabelen er definert på intervallet fra 0 til 1. WoS har videre delt artiklene inn i klustere, hvor de mener å ha grunnlag til å tro at et kluster er tilknyttet kun en ekte person. Måten WoS har laget klusterne på er ikke offisielt kjent, og baserer seg på deres algoritmer og data de har samlet inn. Ved manuell gjennomgang, oppdaget vi at selv om score-verdien var høy, var det usikkert i om personen var personen av interesse. Det var spesielt tilfelle for personer med vanlig navn og fra større enheter, slik som UiO. Vi konkluderte dermed med å ikke utnytte score-verdiene da det kan medføre at vi inkluderer et ukjent antall feil artikler.

Se tabell 5 for deskriptiv statistikk av data utnyttet i analysen. Fra tabell 5 ser det ut som at forskere tildelt FRIPRO typisk publiserer flere artikler per år, og hver artikkel blir typisk mer sitert. I rapporten prøver vi å isolere effekten av å få tildelt FRIPRO, og kvantifisere den i årene etter en eventuell tildeling.

Data	Forsker tildelt FRIPRO	Forsker ikke tildelt FRIPRO
Gjennomsnittlig publikasjoner per år per forsker	4.4 (0.8, 11.4)	2.8 (0.3, 10)
Gjennomsnittlig siteringer per år per artikkel	3.8 (0.1, 12.0)	3.2 (0, 9.6)
Alder ved prosjektstart	48 (36, 63)	48 (35, 65)
Årstall ved oppnådd PhD	1993 (1977, 2001)	1994 (1979, 2005)
Registrert årstall for PhD	74%	65%
Kvinne	31.5 %	30.6%

Tabell 5. Deskriptiv statistikk av datagrunnlag brukt i analysen. Tallene viser gjennomsnittlig verdi og 90% dekningsgrad. Gjennomsnittlig publikasjoner er utregnet i intervallet 2000 til 2016, og gjennomsnittlig siteringer er utregnet i intervallet fra året etter publikasjon til 2016.

4 Metode

I denne seksjonen går vi gjennom modellene brukt for å analysere publiseringsratene og siteringsratene, samt en enklere modell for publiseringsrate. Denne siste er inkludert for å undersøke om vi får forenlige resultater med hva vi får med den mer komplekse og detaljerte publikasjonsmodellen som vi oppfatter som hovedmodellen for publiseringsraten.

4.1 Modell for publiseringsrate

La $Y_{i,t}$ betegne antall publikasjoner av *ite* forsker i år t . Vi antar at årlig antall publikasjoner er negativt binomisk fordelt, dvs.:

$$Y_{i,t} \sim \text{NB}(\mu_{i,t}, \nu), \quad (1)$$

hvor $\mu_{i,t}$ er forventet antall publikasjoner av *ite* forsker i år t , og ν er en overdispersjonsparameter som sammen med forventninga angir variansen som blir $\mu_{i,t} + \mu_{i,t}^2/\nu$. Vi antar videre at forventet antall publikasjoner av *ite* forsker i år t kan bli beskrevet av kovariater (eller forklaringsvariabler) og tilfeldige effekter ved:

$$\log(\mu_{i,t}) = f(X_{i,t}) + \gamma_i = \beta_0 + \sum_j \beta_j x_{i,t,j} + \sum_k s_k(x_{i,t,k}) + \gamma_i. \quad (2)$$

Siden modellen er additiv på logaritmisk skala vil alle effekter være multiplikative på original skala. For en kovariat som inngår lineært med koeffisient β kan den prosentvise effekten beregnes som $100 \cdot (\exp(\beta) - 1)$.

I ligning (2) representerer $X_{i,t}$ kovariatene til *ite* forsker i år t . Videre er f en additiv funksjon av de individuelle kovariateffektene, som kan skrives som lineærkombinasjon av kovariater (her $x_{i,t,j}$) og en sum av kovariater som inngår ikke-lineært (her $x_{i,t,k}$), hvor β_j er parametre som må estimeres fra dataene og s_j er ikke-lineære funksjoner som også må estimeres. Til sist er γ_i en såkalt tilfeldig effekt på individnivå, som er ment å korrigere for at en enkelt forsker kan publisere mer eller mindre enn hva en kan forklare med kovariatene som er inkludert i modellen. Den tilfeldige effekten er antatt normalfordelt med forventning null og varians σ_f , og betinget på data estimeres σ_f og individeffekter for hver enkelt forsker. Tabell 6 angir hvilke kovariater som er inkludert.

Kovariater i publikasjonsmodellen, se ligning (2)			
Parameter	Type	Verdier	Referanseverdi
Kjønn	Kategorisk (2 nivåer)	mann/kvinne	kvinne
Inkludert i NIFU per tid	Kategorisk (2 nivåer)	ja/nei	ja
Forskerfagfelt realfag	Referanse for fagfelt		
Forskerfagfelt humaniora og samfunnsfag	Kategorisk (2 nivåer)	ja/nei	nei/realfag
Forskerfagfelt medisin og helse	Kategorisk (2 nivåer)	ja/nei	nei/realfag
“Forskerfagfelt” administrasjon	Kategorisk (2 nivåer)	ja/nei	nei/realfag
Ansatt på universitet	Referanse for ansettelse		
Ansatt i instituttsektoren	Kategorisk (2 nivåer)	ja/nei	nei/universitet
Ansatt på høyskole	Kategorisk (2 nivåer)	ja/nei	nei/universitet
Ansatt i medisin- og helsesektoren	Kategorisk (2 nivåer)	ja/nei	nei/universitet
Oppnådd doktorgrad	Kategorisk (2 nivåer)	ja/nei	ja
Alder	Kontinuerlig	[30-72]	50 år
Alder ved avlagt doktorgrad	Kontinuerlig	[23,60]	40 år
Årstallseffekt forskerfagfelt realfag	Kontinuerlig	[2000,2016]	2011
Årstallseffekt forskerfagfelt humaniora og samfunnsfag	Kontinuerlig	[2000,2016]	2011
Årstallseffekt forskerfagfelt medisin og helse	Kontinuerlig	[2000,2016]	2011
Årstallseffekt ansatt i instituttsektoren	Kontinuerlig	[2000,2016]	2011
Årstallseffekt ansatt på høyskole	Kontinuerlig	[2000,2016]	2011
Årstallseffekt ansatt i medisin og helse	Kontinuerlig	[2000,2016]	2011
Tid siden innvilget FRIPRO	Kontinuerlig	[0,10]	0, innvilgelsesår
Tilfeldige effekter, se ligning (1) og (2)			
Forskereffekt	$\gamma_i \sim N(0, \sigma_f^2), i = 1, \dots, 2553$		

Tabell 6. Kovariater og tilfeldige effekter brukt i publikasjonsmodellen.

Alle kovariater har en referanseverdi hvor effekten settes lik 0 på logaritmisk skala, slik at det er kun er konstantleddet β_0 som angir normalverdien for en forsker og et år hvor alle kovariater har referanseverdien. Referanseforskeren er en kvinne, i øyeblikket registrert i forskerregistret, er 50 år gammel, tok doktorgrad som 40-åring, har fagfelt realfag, er ansatt på et universitet og årstallet er 2011. Forventet antall publikasjoner i 2011 for en slik forsker vil være $\exp(\beta_0) \cdot \exp(0.5\sigma_f^2)$, hvor det siste leddet skyldes at forventinga til den tilfeldige effekten γ_i på original skala er $\exp(0.5\sigma_f^2)$.

Kjønnseffekten beskriver hvor mye mer eller mindre en mann publiserer i forhold til en kvinne, gitt at alle andre faktorer er like.

Kovariaten “Inkludert i NIFUs forskerpersonalregister per tid” angir om forskeren av interesse har hatt en forskerjobb nært i tid, se mer detaljert definisjon under. Hvis denne har verdien “nei” kan forskeren ha blitt pensjonert eller gått over i en annen type stilling, og vil derfor forventes å publisere mindre.

Kovariaten for forskerens fagfelt er ment å ta hensyn til at forskere fra ulike fagfelt har typisk ulik publiseringsrate, og at de også kan ha ulik dekningsgrad i databasene. For eksempel kan en tradisjonelt forvente at en forsker innen humaniora eller samfunnsfag publiserer sjeldnere enn en forsker innen realfag. Kovariaten for fagfelt er opprinnelig for hvert enkelt år forskernes tilknytting til åtte grupper av fagfelt gitt i tabell 2. Vi har slått disse sammen til gruppene realfag (utenom medisin og helse) som er referansecate-

gorien, humaniora og samfunnsfag, medisin og helse, samt administrasjon. Vi har videre brukt et tre-års vindu av de årlige verdier og latt verdien i år t være majoriteten av fagfeltet i år $t - 2$, $t - 1$ og t . Hvis det er tre ulike fagfelt brukes fagfeltet i år t . Merk at de estimerte effektene vil gjelde for referanseåret som er 2011. Endringer over tid håndteres ved hjelp av årstallseffektene, se under.

Kovariaten for ansattforhold korrigerer for ulik dekningsgrad i CRISTIN for de ulike typer institusjoner. Universitetene har hatt god dekning i lang tid, mens andre grupper av institusjoner tidligere hadde dårligere dekning. Ansettelsesforhold kan variere over tid, og vi har definert denne tidsvarierende kovariaten på samme måte som fagfelt, det vil si at den i år t får det vanligste ansettelsesforholdet i år t og de to foregående år. De estimerte effektene vil også her gjelde for referanseåret som er 2011, mens endringer over tid håndteres ved hjelp av årstallseffektene, se under.

Alderseffekten er ment å ta hensyn til at forskere typisk publiserer mer etterhvert som de blir eldre og mer erfarne, men med en utflating og muligens en nedgang etterhvert som de nærmer seg pensjonsalderen.

“Alder ved avlagt doktorgrad” justerer for når i sin karriere forskeren har tatt doktorgrad, etter at alder og andre kovariater allerede er inkludert i modellen. Det er ikke alle forskere som har doktorgrad, men de kan likevel være kvalifisert for et FRIPRO-prosjekt, f.eks. hvis de har professorkompetanse. Disse er håndtert ved at “Alder ved avlagt doktorgrad” settes til referanseverdien som er 40 år, og at det i tillegg innføres en ekstra indikatorvariabel som er 0 for de som har tatt doktorgrad og 1 for de som ikke har tatt doktorgrad. Forskere med ukjent årstall for avlagt doktorgrad er behandlet som forskere uten doktorgrad.

Årstallseffektene for fagfelt og for arbeidssted eller institusjonstype er inkludert for å korrigere for at effekten av disse endres over tid. De er alle satt til 0 i referanseåret som er 2011, og forskjellen i dette året er gitt av effektene av indikatorvariablene som er beskrevet over. Årstallseffektene angir altså endring i publiseringsrate i forhold til 2011, innen det enkelte fagfelt eller innen den enkelte type arbeidssted. For eksempel har universitetene hatt god dekning i CRISTIN lengre tid, og vi forventer derfor at de øvrige sektorene har en økende registrert publiseringsgrad over tid. Når det gjelder fagfelt har vi ikke inkludert en egen årstallseffekt for kategorien administrasjon, fordi det gjelder ganske få forskere. I stedet har vi tildelt forskere registrert med administrativt fagfelt gjennomsnittet av årstallseffektene for de tre øvrige fagfelt.

Hovedfokus i studien er effekten av å få tildelt et FRIPRO-prosjekt. Det håndteres med kovariaten “Tid siden innvilget FRIPRO”. Denne er alltid 0 for forskere som aldri har fått innvilget FRIPRO. For forskere som har blitt tildelt et FRIPRO-prosjekt er den 0 til og med innvilgelsesåret (dvs. året før planlagt prosjektstart) og øker deretter med en per år. Effekten av denne kovariaten måler dermed den forventede forskjellen mellom en forsker som har fått innvilget et FRIPRO-prosjekt og de som ikke har, justert for alle øvrige faktorer i modellen. Den er dermed 1 i startåret, og 3 i sluttåret for et prosjekt som varer i tre år. Deretter kommer en periode på et par år hvor en kan anta at hovedarbeidet bak de

utgitte publikasjonene helt eller delvis er gjort i prosjektperioden, og deretter kommer en til en periode hvor en kan snakke om langtidseffekter av prosjektet. Vi estimerer FRIPRO-effekten som en kontinuerlig funksjon med bruk av splines, hvor usikkerheten uttrykkes ved en lineærkombinasjon av usikkerheten i gitte knutepunkter jevnt fordelt over område av interesse. Dette medfører at usikkerheten i områder med lite data er vanskelig å tolke, og vi har på grunn av det latt 10 år etter prosjektstart være øvre grense for hvor FRIPRO-kovariaten påvirker publikasjons- og siteringsraten. Vi utfører også en tilleggsanalyse hvor FRIPRO-effekten deles inn etter hvilken fagkomité søknaden går inn under. For forskere som har fått innvilget flere FRIPRO-prosjekter bruker vi alltid det første av disse, og betrakter de kommende som en mulig effekt av at de har fått det første av dem.

Kovariaten “Inkludert i NIFUs forskerregister per tid” er definert som en indikatorfunksjon. Indikatorfunksjonen er én hvis gjeldende tidspunkt er inneholdt i et tidsvindu nært første og siste året forskeren var registrert i forskerpersonalregisteret. Nært betyr her at forskeren må ha vært registrert i forskerpersonalregisteret i inneværende år eller året før. Merk at forskerpersonalregisteret var kun oppdatert annen hvert år før 2007, og hvert enkelt år fra og med 2007. Vi har på grunn av oppdateringstidspunktene videre definert indikatorfunksjonen til å være én hvis både forskeren var inkludert i forskerpersonalregisteret året etter og året er 2006 eller tidligere.

4.2 Modell for siteringsrate

I det følgende presenterer vi siteringsmodellen. Legg merke til at mens publikasjonsmodellen utnytter én tidsserie per forsker utnytter siteringsanalysen én tidsserie per artikkel. Husk også at siteringsanalysen kun omhandler artikler, siden det var det vi hentet ut fra WoS. Vi antar en relativt lik modell for siteringsanalysen som for publikasjonsanalysen, men for siteringsanalysen arbeider vi på artikkelnivå, og innfører derfor noen ekstra kovariater på artikkelnivå.

La nå $Y_{i,j,t}$ være antall siteringer i år t av j te artikkel av i te forsker. Tilsvarende som i publikasjonsmodellen antar vi at

$$Y_{i,j,t} \sim \text{NB}(\mu_{i,j,t}, \nu), \quad (3)$$

hvor $\mu_{i,j,t}$ er forventet antall siteringer av j te artikkel i år t av i te forsker, og ν er en overdispersjonsparameter. Selv om vi delvis bruker samme notasjon for publikasjonsmodellen er parameterverdiene ikke de samme. Vi antar videre at forventningen til antall siteringer av j te artikkel i år t av i te forsker kan bli beskrevet med kovariater og tilfeldige effekter ved:

$$\begin{aligned} \log(\mu_{i,j,t}) &= f(X_{i,j,t}) + \gamma_i + \delta_{i,j} \\ &= \beta_0 + \sum_l \beta_l x_{i,j,t,l} + \sum_k s_k(x_{i,j,t,k}) + \gamma_i + \delta_{i,j}, \end{aligned} \quad (4)$$

hvor, analogt med ligning (4), forklaringsvariablene $x_{i,j,t,l}$ inngår lineært og $x_{i,j,t,k}$ ikke-lineært, γ_i er en forskereffekt, og $\delta_{i,j}$ er en artikleffekt. Se tabell 7 for detaljer om kovariater brukt i siteringsanalysen som ikke er inkludert i publikasjonsanalysen.

Parameter	Kovariater i ligning 4		
	Type	Verdier	Referanseverdi
Artikkelkategori realfag	Referanse for artikkelkategori		
Artikkelkategori humaniora og samfunnsfag	Kategorisk (2 nivåer)	ja/nei	nei/realfag
Artikkelkategori medisin og helse	Kategorisk (2 nivåer)	ja/nei	nei/realfag
Artikkelkategori annet	Kategorisk (2 nivåer)	ja/nei	nei/realfag
Tid siden utgivelse av publikasjon	Kontinuerlig	[1,16]	1 år etter utgivelse
Tid fra innvilget FRIPRO til publisering	Kontinuerlig	[0,10]	0, innvilgelsesår
Tilfeldige effekter i ligning (3) og (4)			
Artikkeleffekt	$\delta_{i,j} \sim N(0, \sigma_a^2)$		

Tabell 7. *Kovariater og tilfeldige effekter brukt i modell for siteringer i tillegg til de som er nevnt i tabell 6, bortsett fra at kovariatene for inkludering i NIFUs forskerpersonalregister og ansettelsessted ikke er med.*

For siteringsanalysen inkluderer vi ikke kovariatene som går på om forskeren er registrert i NIFUs forskerpersonalregister eller ikke, fordi vi antar det er irrelevant for antall siteringer på en artikkel som er skrevet. Vi inkluderer heller ikke kovariatene for ansettelsessted. Vi unnlater dette da vi ikke har grunnlag for å tro at WoS har ulik historisk dekningsgrad av siteringer til artikler skrevet av forskere ansatt i ulike sektorer utover at de kan ha ulike fagfelt. Forøvrig er alle kovariater som var tatt med i publikasjonsanalysen også inkludert i siteringsanalysen, med samme fortolkning bortsett fra at de gjelder antall årlige siteringer per artikkel.

I tillegg har vi inkludert en kovariat for tid siden utgivelse av artikkelen, hvor det første året etter utgivelse er referanseåret. Vi ser bort fra siteringer i samme år som utgivelsesåret, da det er svært få.

Videre har vi inkludert en kovariat for fagfelt for artikkel, hvor realfag er referanse. Vi har etter beste skjønn gruppert de opprinnelige 250 fagfelt i WoS inn i kategoriene realfag, humaniora og samfunnsfag, medisin og helse, og annet. En artikkel kan ha flere tilhørende fagfelt, og dette er tatt hensyn til ved å la andelen tilhørighet til de ulike fagfeltene inngå som en lineær effekt i (4).

Det er inkludert to tilfeldige effekter i siteringsmodellen. De to effektene er ment å fange opp strukturer som ikke er forklart gjennom kovariatene. Forskereffekten tar hensyn til at ulike forskere typisk publiserer artikler som blir sitert i ulik grad. Artikkeleffekten tar videre hensyn til at ulike artikler blir sitert i ulik grad. Merk at hvis det kun er knyttet én artikkel til en spesifikk forsker i datagrunnlaget, så vil artikkeleffekten være blandet sammen med forskereffekten. Vi har dermed kun definert en forskereffekt for forskere som har to eller flere artikler knyttet til seg.

Vi ønsker å se på effekten av å få tildelt FRIPRO på siteringsraten, og dette er undersøkt ved å inkludere variabelen "Tid fra FRIPRO-tildeling til publikasjon". For hver artikkel en forsker får publisert i etterkant av en FRIPRO-tildeling angir denne antall år det går mellom tildeling og publisering av artikkelen, uansett om artikkelen er skrevet før, i løpet av, eller etter selve prosjektet. For artikler utgitt i samme år som innvilgelsesåret eller

tidligere er denne variabelen 0. Det er den også for alle artiklene til forskerne som aldri har fått tildelt et FRIPRO-prosjekt.

4.3 Alternativ, forenklet modell for publiseringsrate

Vi skal nå introdusere en ny modell for publikasjonsanalysen. Modellen kan bli sett på som en forenklet versjon av publikasjonsmodellen definert i 4.1. Dette er en mer standard statistisk modell hvor vi kan bruke standard programvare til estimering. En motivasjon for å bruke denne enklere metoden er for å se om resultatene peker i samme retning som resultatene fra modellen vi har presentert over.

Anta i te forsker søker om FRIPRO-prosjekt med oppstartsår t . Definer $Z_i^{(1)}$ til å være antall publikasjoner av i te forsker i år $t - 2$ og $t - 1$. Definer $Z_i^{(2)}$ til å være antall artikler i år $t + 5$ og $t + 6$. Anta så at

$$Z_i^{(2)} \sim \text{NB}(\mu_Z, \nu_Z), \quad (5)$$

der μ_Z er forventet $Z_i^{(2)}$ og ν_Z er en overdispersjonsparameter. Antar videre at $\log \mu_Z = f(\mathbf{X}_i)$, der \mathbf{X}_i representerer forklaringsvariablene til i te forsker: $\log(Z_i^{(1)} + 1)$, FRIPRO-tildeling, kjønn, fagfelt, alder ved år t og om forskeren er inneholdt i forskerpersonalregisteret i tid $t - 1$ og $t + 6$. Alle forklaringsvariablene, bortsett fra alder og $Z_i^{(1)}$, blir her behandlet som kategoriske variable. Forklaringsvariabelen om forskeren er inneholdt i forskerpersonalregisteret er definert som i seksjon 4.1. Merk at siden vi inkluderer tidligere antall publikasjoner i modell 5, vil hver enkelt forsker fungere som sin egen kontroll.

Forskere kan søke om FRIPRO flere ganger. Da dette var tilfelle definerte vi t som prosjektstartsår i siste søknad før 2010, eller prosjektstartsår i første innvilget søknad. Vi prøvde å definere t som prosjektstartsår for første søknad før 2010, og det endret resultatene i liten grad.

Merk at vi kun kan utnytte forskere som har søkt FRIPRO i perioden 2004-2009 med denne modellen, da vi krever at vi har data opp til syv år etter prosjektstart og vi kun har data frem til 2016. Det medførte at det er 2013 forskere inkludert i den forenklete publikasjonsanalysen.

4.4 Estimering

Parameterne i de ulike modellene er estimert ved tilpasning til de tilgjengelige dataene ved å maksimere likelihood-funksjonen. Usikkerheten til parameterestimaterne og avledede størrelser er kvantifisert ved å normaltilnærme likelihood-funksjonen i maksimum, basert på standard maksimum likelihood-teori.

For de to hovedanalysene er modellen estimert ved å bruke R-pakken TMB (Kristensen et al., 2015) kombinert med optimeringsfunksjonen `nlnmb` i R. Det er relativt mye data som skal analyseres med relativt komplekse modeller, spesielt for siteringsanalysen. Fordelen med å bruke TMB i denne analysen er at TMB benytter seg av automatisk derivering av likelihood-funksjonen, og Laplace-tilnærming av latente Gaussiske felt (det vil si personeffektene og artikkel-effektene). Å utnytte slik teori danner et godt utgangspunkt

for å estimere komplekse modeller med mange parametre innen rimelig tid. En bærbar pc med prosessor intel(R) Core(TM) i5-6300 CPU @ 2,40 GHz, brukte omtrent 20 minutter å tilpasse publikasjonsmodellen og fire timer på å tilpasse siteringsmodellen.

Effekten av alle kontinuerlige forklaringsvariabler som inngår i modellene (2) og (4) er modellert som ikke-lineære funksjoner ved bruk av splines (Hastie og Tibshirani, 1990). Vi har brukt såkalte kubiske B-splines som basisfunksjoner. Ved å bruke mange basisfunksjoner i en funksjon kan vi få en fleksibel kurve med god tilpasning, men vi bruker da også mange parametre, og det er en fare for overtilpasning. Vi har brukt Akaikes informasjonskriterium (AIC) for å velge antall basisfunksjoner for hvert sett av ikke-lineære funksjoner. AIC er et mye brukt kriterium for hvordan man veier modellens tilpasning og kompleksiteten mot hverandre for modellseleksjon, og består av et ledd som angir hvor godt modellen passer til dataene og et såkalt straffelegg for antall parametre i modellen, og begge ledd bør være minst mulig. Videre er signifikansnivåene til regresjonsparametrene beregnet ved bruk av likelihood-ratio-test (Devore og Berk, 2007). Signifikansnivået med likelihood-ratio-testen vil videre bli betegnet med L.R.-signifikansnivå.

Kun forskere i alderen f.o.m. 30 år t.o.m. 72 år på publikasjonstidspunkt gir bidrag til likelihooden i både publiseringsanalysene og i siteringsanalysen. Alderen kan variere utenfor det intervallet i datasettet, men vi forventer stor individuell variasjon i antall publikasjoner og siteringer for forskere som er svært unge eller ganske gamle. Ved estimering har vi derfor kun tatt med bidrag fra den enkelte forsker så lenge hun eller han er minst 30 år og maksimum 72 år.

Den alternative, forenklete modellen for publiseringsrate er en mer standard statistisk modell, en såkalt generalisert additiv modell med en negativt binomisk fordelt responsvariabel. Estimering i en slik modell er i prinsippet likevel komplisert, men det finnes nå velutprøvd programvare som gjør dette. Vi har brukt funksjonen *gam* i R-pakken *mgcv* (Wood, 2017).

En fordel med den alternative analyse er både at vi bruker en enklere modell og at vi bruker en R-funksjon som grundig testet. Det vil dermed være en indikasjon på at vår hovedanalyse er gjort riktig med TMB hvis den alternative publikasjonsmodellen gir resultater som er sammenlignbare med de vi får med den mer komplekse publikasjonsmodellen.

5 Resultater

I det følgende presenterer vi resultatene fra publikasjonsanalysen, inkludert resultater fra den enklere analysen, og deretter resultatene fra siteringsanalysen.

5.1 Publikasjonsanalysen

Tabell 8 viser maximum likelihood-estimatene av parameterne i modell 1, samt p-verdi med bruk av likelihood-ratio-test for en test på om effekten er signifikant forskjellig fra 0. Alle p-verdier er svært lave, hvilket vil si at alle sett av variable som er inkludert i

modellen har signifikante effekter. Hyperparameterne i modellen er gitt i nedre del av tabell 8. Verdien på standardavviket σ_f til forskereffekten er relativt stor, og tilsier at det er store individuelle forskjeller mellom forskerne utover det øvrige kovariater forklarer.

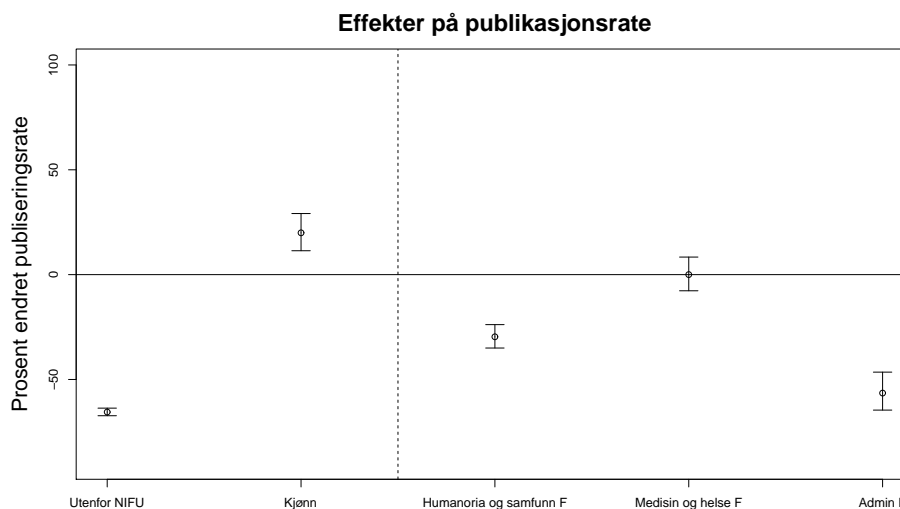
For referanseforskeren, som er kvinne, er registrert i forskerregistret, er 50 år gammel, tok doktorgrad som 40-åring, har fagfelt realfag, er ansatt på et universitet er forventet antall publikasjoner i 2011 lik $\exp(\beta_0) \cdot \exp(0.5\sigma_f^2) = \exp(1.077) \cdot \exp(0.5 \cdot 0.815^2) = 4.1$. Merk at størrelsesordenen stemmer godt overens med gjennomsnittlig antall publikasjoner per forsker gitt i tabell 5. Øvrige resultater fortolkes relativt til denne verdien, og det gjøres best ut fra de følgende figurene. Vi går først gjennom effekten av de ulike kovariater som er inkludert for å justere for forskjeller mellom forskere og mellom år, og avslutter med å presentere resultatene for effekten av å få tildelt et FRIPRO-prosjekt.

Kovariater i publikasjonsmodellen (2)		
Parameter	Estimat med 95% K.I.	L.R. signifikansnivå
Konstantledd, β_0	1.077 (0.961, 1.193)	ikke relevant
Utenfor NIFU	-1.065 (-1.092, -1.013)	$< 10^{-5}$
Kjønn (mann)	0.183 (0.109, 0.257)	$< 10^{-5}$
Forskerfagfelt	Se figur 3	$< 10^{-5}$
Alder	Se figur 4a	$< 10^{-5}$
Doktorgrad og eventuell alder ved avlagt doktorgrad	Se figur 4b	$< 10^{-5}$
Årstallseffekter fagfelt	Se figur 5	$< 10^{-5}$
Årstallseffekter institusjon	Se figur 6	$< 10^{-5}$
Tid siden innvilget FRIPRO	Se figur 8	$< 10^{-5}$
Hyperparametre i publikasjonsmodellen.		
σ_f	0.815 (0.790, 0.842)	
ν	3.83 (3.69, 3.98)	

Tabell 8. Estimerte parametre med 95% konfidensintervall og tilhørende p-verdier for publikasjonsmodellen. Merk at dette gjelder på den logaritmiske skalaen i modell (2), mens alle figurer vises på original skala som prosentvise effekter.

Figur 3 viser estimert prosentvis effekt av de kategoriske kovariatene på original skala, beregnet som $100 \cdot (\exp(\hat{\beta}) - 1)$, hvor $\hat{\beta}$ er en estimert koeffisient på logaritmisk skala. En forsker som ikke lenger er registrert i forskerpersonalregistret forventes å publisere 60% mindre enn en som er registrert, gitt at alt annet er likt. En mann publiserer i forventning omkring 20% flere artikler enn en kvinne. For forskerfagfelt må vi huske at referansen er realfag og at referanseåret er 2011. Resultatene tilsier at i 2011 publiserte forskere innen humaniora og samfunnsfag (i figuren forkortet til humaniora og samfunn F, hvor F står for forskerfagfelt) omkring 25% færre artikler enn en forsker innen realfag, mens en forsker innen medisin og helse (i figuren forkortet til medisin og helse F) publiserte like mange artikler som en realfagsforsker. Forskere som er registrert med en administrativ stilling publiserte mindre enn 50% av realfagsforskeren.

Figur 4a viser at forventet antall publiserte artikler per år mer enn dobles når forskerens alder øker fra 30 til 50 år, for deretter å flate ut og etterhvert sees en nedgang. Økningen



Figur 3. Estimerte effekter av kategoriske variable på publiseringssrate. Disse fortolkes i forhold til referansekategoriene som for de ulike grupper av variable som er i) forskerstatus: inneholdt i NIFUs forskerpersonalregister, ii) kjønn: kvinne, iii) forskerfagfelt: realfag, iv) institusjon: universitet.

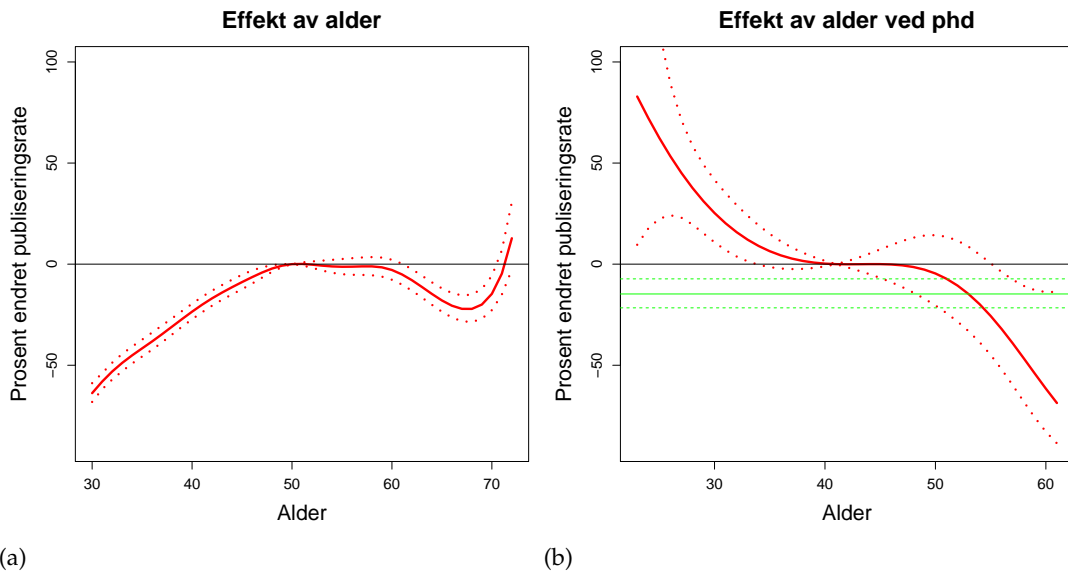
ved alder over 70 år kan skyldes at forskere som er aktive så lenge er en spesiell gruppe med høy publiseringssrate.

Figur 4b viser at forskere som tar doktorgrad tidlig jevnt over har en høyere publiseringssrate enn forskere som disputerer seint i livet. Forskere uten doktorgrad (grønn, heltrukken horisontal linje i figuren) har forventet publiseringssrate omtrent som en forsker som disputerte i en alder av 50 år.

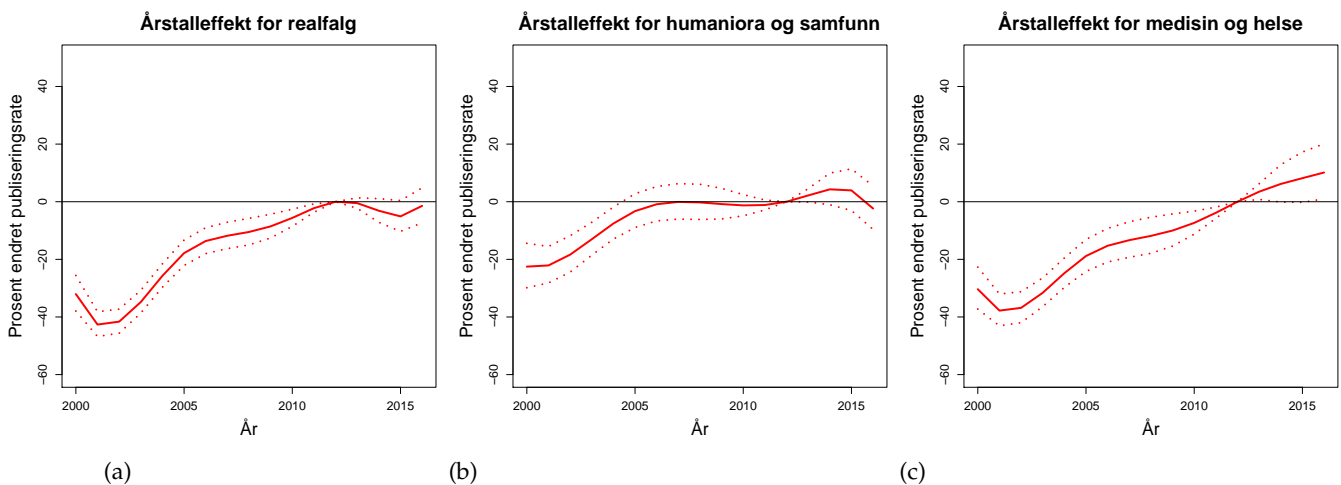
Figur 5 viser hvordan publiseringssraten innen de enkelte forskerfagfelt har blitt endret over tid. For alle fagfelt er publiseringssraten økt siden 2000, men innen humaniora og samfunnsfag har den flatet ut de ti siste årene. Siden universitet er referansekategori for institusjon gjelder disse kurvene for universitetsansatte.

Figur 6 viser endringer i registrert publiseringssrate over tid for de enkelte typer forskerinstitutioner *relativt til universitetene*. Det vi si at for å få utvikling over tid for et gitt fagfelt og en gitt institusjonstype utenom universitet må disse kurvene multipliseres med kurven for det gitte fagfelt i figur 5. Hensikten med å inkludere disse årstallseffektene for institusjonstype er å korrigere for manglende dekning før omkring 2011 i CRISTIN. For instituttsektoren har den registrerte publiseringssraten (i forhold til universitetene) økt kraftig fram mot omtrent 2011 og deretter flatet ut. For høyskoler og helseforetak har det også vært en økning, men ikke så stor. For disse er også usikkerheten større (videre konfidensintervall), og en skal derfor ikke legge mye vekt på buktningene i kurvene, det er et resultat av måten vi modellerer på.

Figur 7 viser at det er store individuelle forskjeller i publiseringssrate mellom forskerne, selv etter at det er korrigert for de ulike kovariatene. Det betyr at det eksisterer strukturer i data som ikke blir forklart av kovariatene, og det er viktig å ta hensyn til slike underliggende strukturer for å kvantifisere usikkerheten av det vi estimerer.

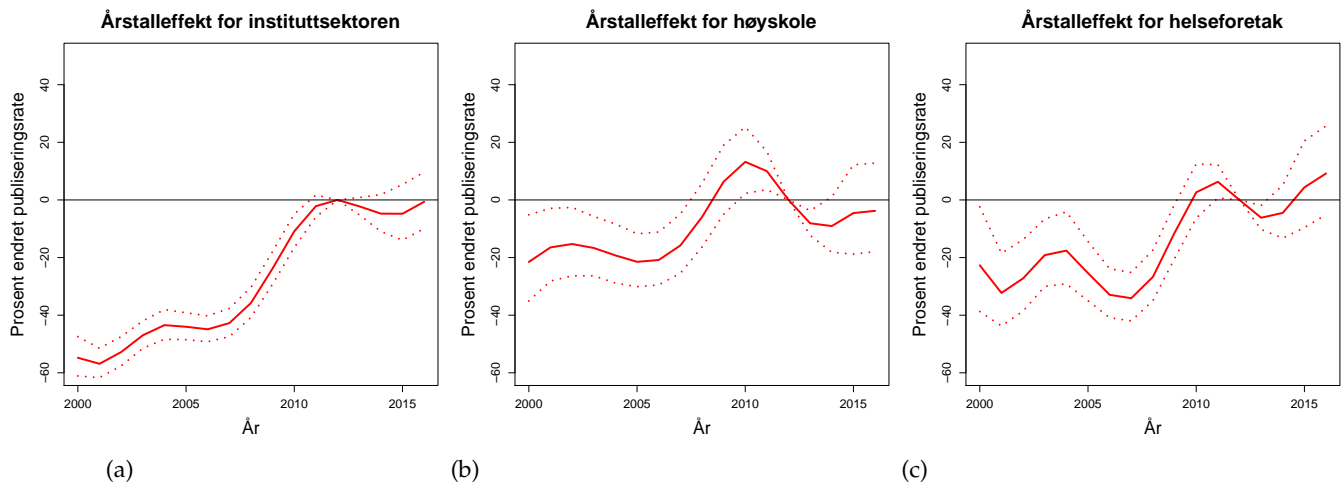


Figur 4. Estimerte effekter på publiseringssrate av a) alder ved publisering (referanseverdi er 50 år), b) alder ved doktordisputas (referanseverdi er 40 år). Her og i øvrige figurer med kurver angir den heltrukne linja estimert forventet effekt og de prikkete linjene et tilhørende 95% konfidensintervall. For forskere som ikke har tatt doktorgrad er det estimert en effekt relativt til en forsker som tok doktorgrad i en alder av 40 år. Dette estimatet er vist som en grønn heltrukken linje med et 95% konfidensintervall vist med grønne, prikkete linjer.

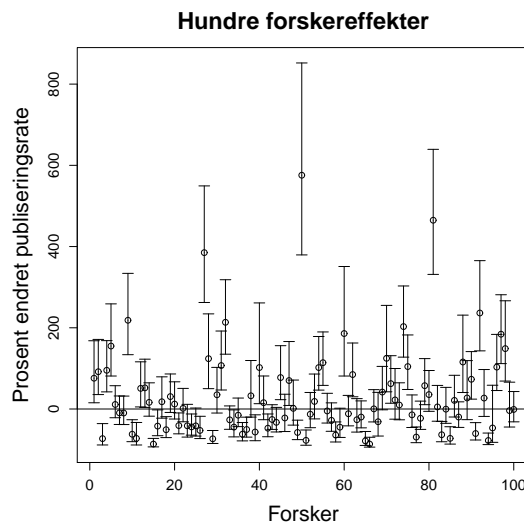


Figur 5. Estimerte årstalleffekter av forskerfagfelt på publiseringssrate. Panel a) realfag, b) humaniora og samfunnsfag og c) medisin og helse. Referanseinstitusjonen er et universitet, slik at dette dermed er årstalleffekten for tre ulike fagfelt i universitetssektoren. Referanseåret er 2011, slik at alle effekter er relativt til publiseringssraten i dette året.

Hensikten med den del av modellen vi har beskrevet så langt har vært å korrigere for viktige faktorer som varierer mellom forskerne og som har betydning for publiseringssraten. Til sammen angir de normal publiseringssrate for en forsker som aldri, eller så langt i karrieren, ikke har fått tildelt FRIPRO-midler. FRIPRO-effekten (figur 8) angir den re-

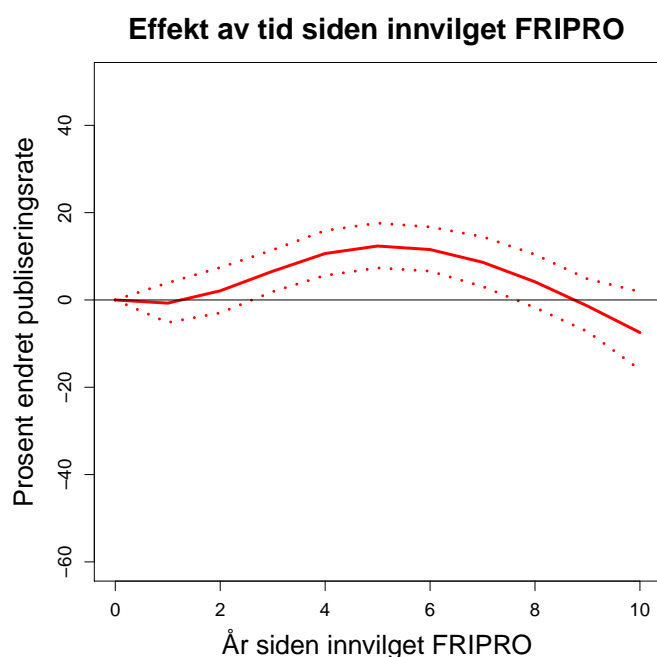


Figur 6. Estimerte årstalleffekter av institusjonstype på publiseringssrate, relativt til universitetssektoren som er referanse-kategorien. Panel a) instituttsektoren, b) høyskole og c) helseforetak. Referanseåret er 2011, slik at alle effekter er relativt til publiseringssraten i dette året.



Figur 7. Estimerte tilfeldige effekter på forskernivå i publikasjonsmodellen, for 100 ulike forskere. Dette skal ta høyde for individuelle forskjeller mellom forskerne som ikke lar seg forklare av kovariatene som er inkludert i modellen.

lative endring i forventet publiseringssrate i årene etter tildeling av et FRIPRO-prosjekt, sammenlignet med om forskeren ikke hadde fått tildelt et slikt prosjekt. Fra og med to år etter tildelingsåret er det en økning i dette tillegget i publiseringssraten med en topp mot omkring 12% økt publiseringssrate fem år etter tildeling. Deretter blir effekten mindre, og åtte år etter tildeling er den estimerte effekten 0. Merk at selv om den estimerte effekten er negativ i år ni og ti er den ikke signifikant forskjellig fra 0, så en rimelig fortolkning er at FRIPRO-effekten er 0 fra og med åtte år etter tildeling. Hvis vi antar at et prosjekt starter året etter utlysning og varer i tre eller fire år, dvs. slutter i år fire eller fem, tyder resultatene på at tildeling av et FRIPRO-prosjekt er assosiert med en økning av publiseringssraten under og i noen år i etterkant av prosjektet, men at det ikke er noen langtidseffekt som



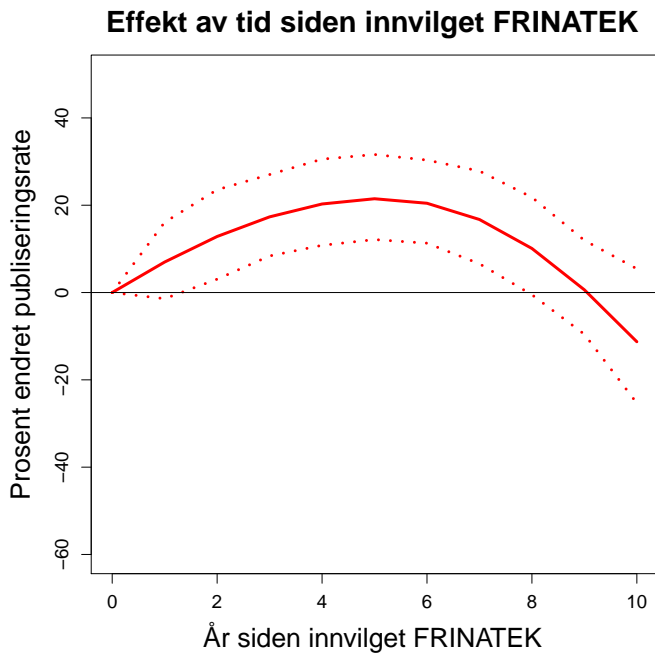
Figur 8. Estimert effekt på publiseringsrate av tildeling av FRIPRO-midler som funksjon av tid siden tildelingsår.

varer videre i karrieren.

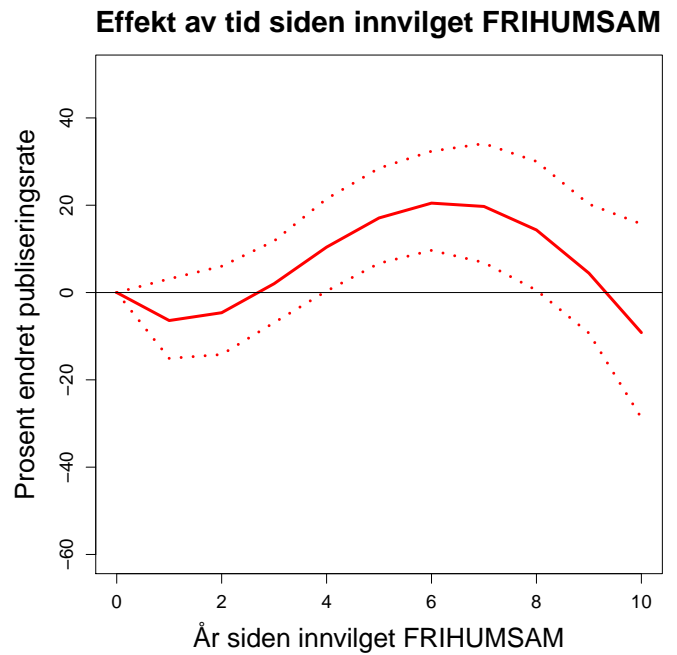
For å kvantifisere hva dette har å si totalt sett har vi for de forskerne som fikk tildelt et FRIPRO-prosjekt beregnet deres forventede antall publikasjoner med et FRIPRO-prosjekt og sammenlignet det med forventede antall publikasjoner for de samme forskerne om de ikke hadde fått et FRIPRO-prosjekt. Beregningen viser at 1275 ekstra artikler (95% konfidensintervall fra 414 til 2111) kan tilskrives tildeling av FRIPRO-prosjektene. Det tilsvarer i snitt 0.27 (0.09, 0.46) flere publikasjoner per år per forsker i perioden etter FRIPRO-tildeling.

Til slutt deler vi FRIPRO-effekten inn i en effekt for henholdsvis FRINATEK, FRIHUMSAM, FRIMEDBIO og STORFORSK. Dette gir en signifikant bedre tilpasning til dataene (p-verdi 0.029), dvs. det er signifikant forskjell mellom effektene av de ulike inndelingene. Figur 9 viser at mesteparten av den felles FRIPRO-effekten kan tilskrives FRINATEK. FRINATEK-prosjekter er klart assosiert med en økt publiseringsrate med en topp fem år etter tildeling med nær 20% økt publiseringsrate, men effekten er 0 igjen ni år etter tildeling. For FRIHUMSAM er det også en signifikant økning i publiseringsraten, men den kommer noe seinere i prosjektet og har noe kortere varighet. For FRIMEDBIO ser det derimot ikke ut til at tildeling av et FRIPRO-prosjekt er assosiert med publiseringsraten, mens for STORFORSK (som kun omfatter 15 tildelinger) er usikkerheten så stor at en ikke kan konkludere.

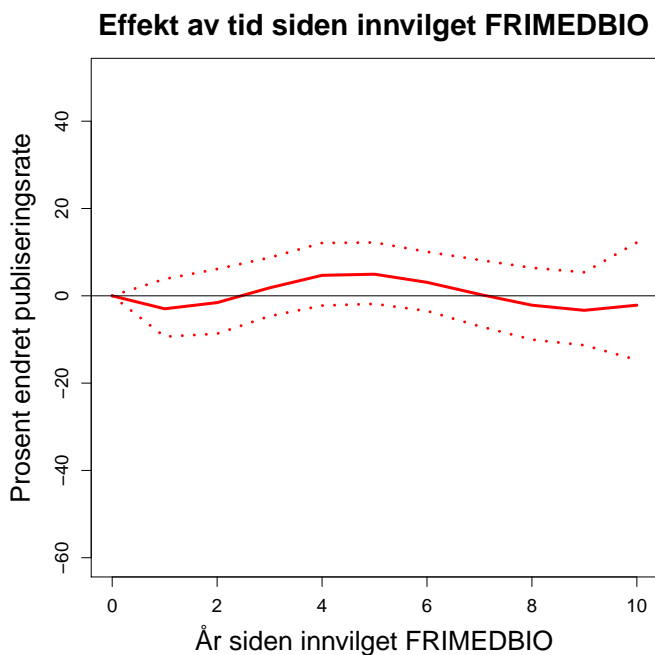
I publikasjonsmodellen referert til her har vi, ved å følge AIC-kriteriet, brukt ni basisfunksjoner for effekt av alder, fire for effekt av alder ved oppnådd doktorgrad, seks for årstallseffekten for fagfelt, sju for årstallseffekten av institusjonstype og fire for effekten



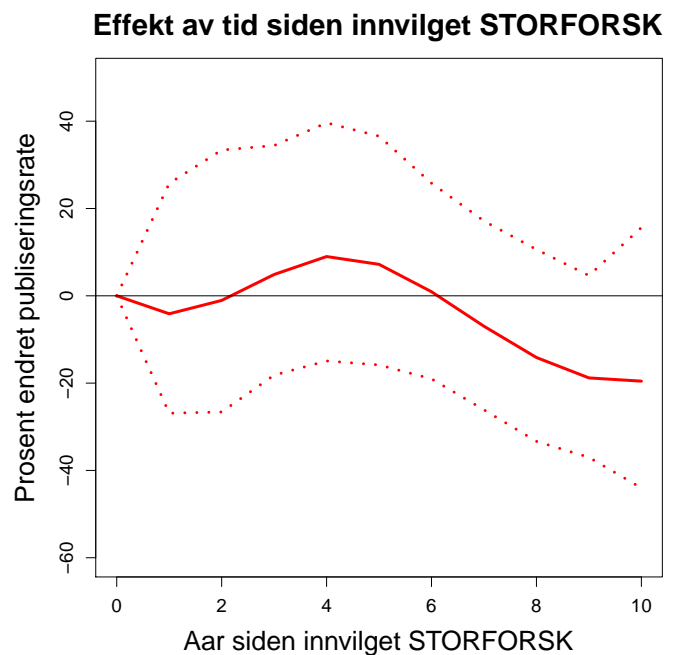
(a)



(b)



(c)



(d)

Figur 9. Estimert effekt på publiseringssrate av tildeling av FRIPRO-midler som funksjon av tid siden tildelingsår, fordelt på de ulike typer FRIPRO-prosjekter, a) FRINATEK, b) FRIHUMSAM, c) FRIMEDBIO og d) STORFORSK. For STORFORSK er det ekstra stor usikkerhet fordi det er svært få som har fått tildelt slike midler.

av tid etter FRIPRO-tildeling (også ved inndeling i de enkelte fagkomiteer).

Tabell 9 viser estimerte regresjonsparametere av den enklere, alternative publiseringss-

analysen beskrevet i seksjon 4.3. Figur 10 viser estimert effekt av alder ved tenkt start av FRIPRO-prosjekt med den enklere publiseringsmodellen. Fra utskriften i tabell 9 ser vi at FRIPRO-parameteren ble estimert til 0.19 med 95% konfidensintervall gitt ved (0.091, 0.29). Dette er på logaritmisk skala, og på original skala betyr det et en forsker som har fått tildelt et FRIPRO-prosjekt publiserer $100 * (\exp(0.19) - 1) \approx 21\%$ (10-34%) flere artikler per år i etterkant av prosjektet enn en sammenlignbar forsker som ikke har fått et slikt prosjekt, hvor "i etterkant" er definert som seks og sju år etter tildelingsåret. Dette estimatet er høyere i den mer detaljerte analysen beskrevet over, men konfidensintervallene fra de to analysene overlapper hverandre, så vi kan si at den enklere analysen til en viss grad støtter opp under resultatene vi har vist så langt.

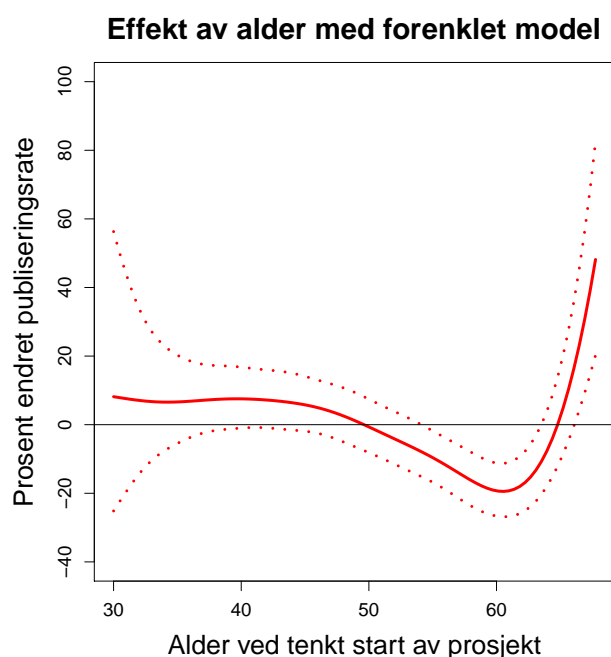
Forenklet publikasjonsmodell		
Parameter	Estimat med 95% K.I.	P-verdi
Konstantledd	1.08 (0.95, 1.21)	$< 10^{-5}$
Tildelt FRIPRO	0.19 (0.091, 0.29)	0.00015
Mann	0.02 (-0.07, 0.11)	0.64
$\log(Z + 1)$	0.60 (0.55, 0.65)	$< 10^{-5}$
Utenfor forskerpersonalregisteret tid $t + 6$	-0.89 (-1.04, -0.75)	$< 10^{-5}$
Utenfor forskerpersonalregisteret tid $t - 1$	0.26 (0.04, 0.47)	0.019
Medisin og helse F på tid $t + 6$	-0.03 (-0.18, 0.24)	0.80
Humaniora og samfunnsvitenskap F på tid $t + 6$	-0.35 (-0.59, -0.11)	0.0040
Administrasjon F på tid $t + 6$	-0.87 (-1.4, -0.32)	0.0018
Medisin og helse F på tid $t - 1$	-0.01 (-0.22, 0.20)	0.94
Humaniora og samfunnsvitenskap F på tid $t - 1$	-0.01 (-0.25, 0.22)	0.91
Administrasjon F på tid $t - 1$	0.34 (-0.46, 1.14)	0.40

Tabell 9. Estimerte regresjonsparametere i den forenklete publikasjonsmodellen med 95% konfidensintervall. Effekten av å få tildelt FRIPRO er markert med fet skrift. P-verdien er regnet ut ved normaltilnærming.

5.2 Siteringsanalysen

Siteringsmodellen beskriver hvordan årlig antall siteringer per artikkel avhenger av ulike kovariater. Tabell 10 viser maksimum likelihood-estimatene for parameterne i denne modellen. Effekt av kjønn og alder ved doktorgrad er ikke signifikante (høye p-verdier), mens alle andre kovariater har en tydelig signifikant effekt. Verdien på standardavviket σ_a til artikleffekten tilsier at det er store forskjeller i hvilken grad de ulike artiklene blir sitert utover det de øvrige kovariater forklarer. Standardavviket σ_f til forskereffekten er noe mindre enn for artikleffekten.

Referanseforskeren vil i 2011 kunne forvente 3.3 siteringer per artikkel ($\exp(\beta_0) \cdot \exp(0.5\sigma_f^2) \cdot \exp(0.5\sigma_a^2) = \exp(0.560) \cdot \exp(0.5 \cdot 0.569^2) \cdot \exp(0.5 \cdot 0.973^2) = 3.3$). Merk at størrelsesordenen stemmer godt overens med gjennomsnittlig siteringer per artikkel gitt i tabell 5. Øvrige resultater fortolkes relativt til denne verdien, og vi går gjennom effekten av de ulike kovariater en for en, og avslutter igjen med å presentere resultatene for effekten av å få tildelt et FRIPRO-prosjekt.

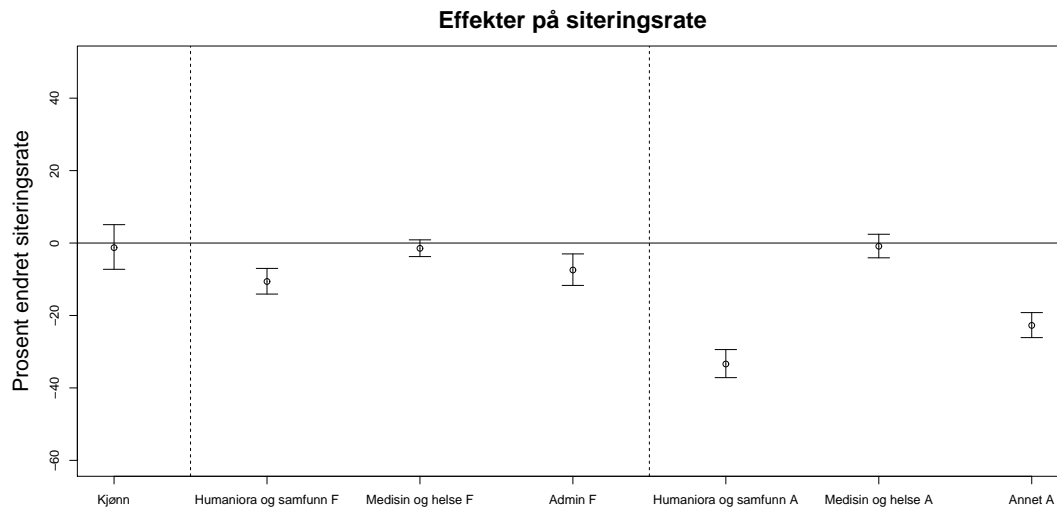


Figur 10. Effekt av alder ved tenkt start av FRIPRO-prosjekt i den forenklete publikasjonsanalysen.

Kovariater i siteringsmodellen., se ligning (4)		
Parameter	Estimat med 95% K.I.	L.R. signifikansnivå
Konstantledd, β_0	0.560 (0.470,0.650)	ikke relevant
Kjønn	-0.008 (-0.070, 0.054)	0.802
Fagfelt forsker	Se figur 11	$< 10^{-5}$
Fagfelt artikkel	Se figur 11	$< 10^{-5}$
Tid siden publisasjon	Se figur 12a	$< 10^{-5}$
Alder	Se figur 12b	$< 10^{-5}$
Doktorgrad og eventuell alder ved avlagt doktorgrad	Se figur 12c	0.206
Årseffekt fagfelt	Se figur 13	$< 10^{-5}$
Tid siden innvilget FRIPRO	Se figur 15	0.006
Hyperparametre i siteringsmodellen, se ligning (3) og (4)		
σ_f	0.569 (0.545, 0.594)	
σ_a	0.973 (0.967, 0.980)	
ν	9.47 (9.31, 9.62)	

Tabell 10. Estimerte parametre med 95% konfidensintervall og tilhørende p-verdier for siteringsmodellen. Merk at dette gjelder på den logaritmiske skalaen i modell (2), mens alle figurer vises på original skala som prosentvise effekter.

Figur 11 viser estimert prosentvis effekt av de kategoriske kovariatene på original skala. Siteringsraten per artikkel skrevet av en mann er nærmest identisk med en artikkel med en kvinnelig forfatter, som i alle slike uttalelser her korrigert for øvrige variable. Vi minner om at referanseåret er 2011 og for referansen er realfag både for forskerens og artikkelens fagfelt. Siteringsraten var i dette året så godt som identisk for realfag og



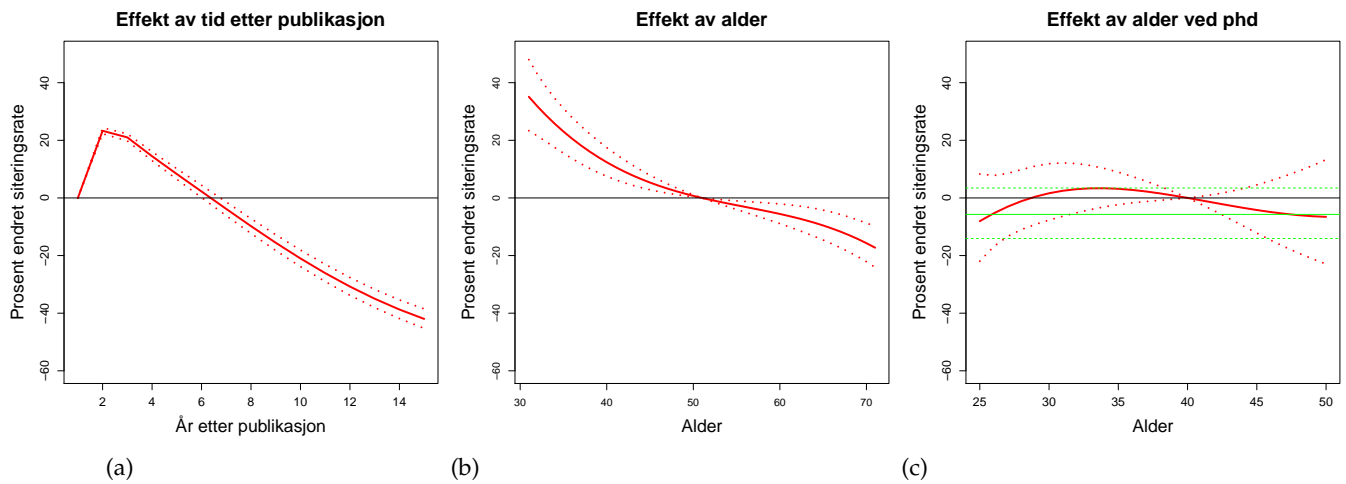
Figur 11. Estimerte effekter av kategoriske variable på siteringsrate. Disse fortolkes i forhold til referansekategoriene som for de ulike grupper av variable som er i) kjønn: kvinne, ii) forskerfagfelt: realfag, iii) artikkelfagfelt: realfag.

medisin og helse, for begge typer fagfeltinndeling. Imidlertid var siteringsraten for en forsker klassifisert til fagfeltet humaniora og samfunnsfag omkring 10% lavere enn for en forsker innen realfag, for en artikkel klassifisert i en felles kategori. For artikkelfagfelt var siteringsraten for en artikkel innen humaniora og samfunnsfag omkring 35% lavere enn for en realfagsartikkel. Dette siste gjelder også for en realfagsforsker som publiserte en artikkel klassifisert som humaniora og samfunnsfag.

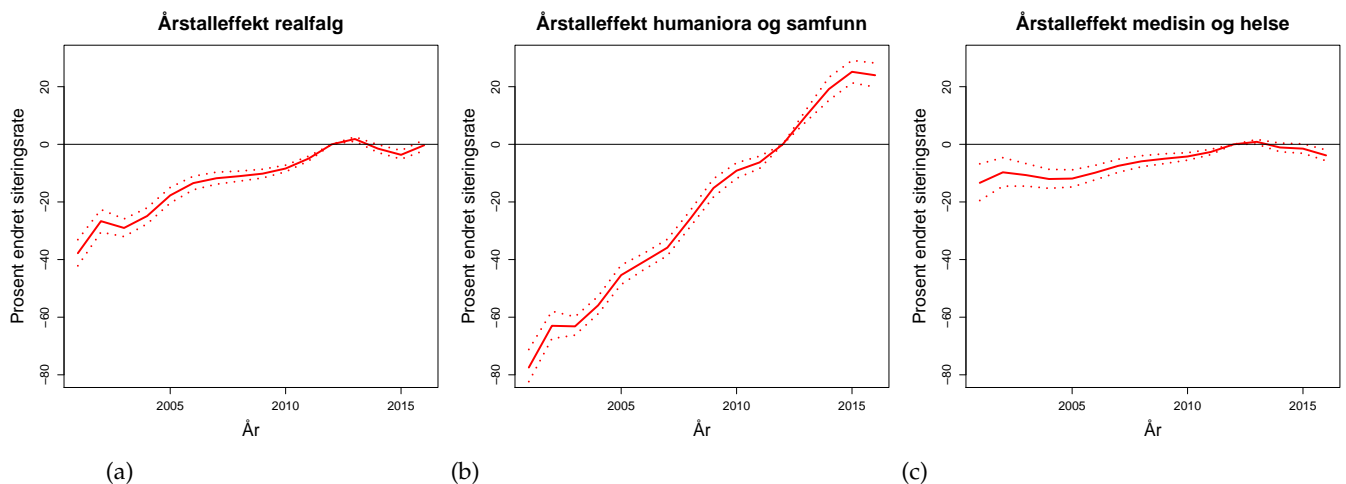
Figur 12a viser hvordan siteringsraten for en artikkel varierer over tid etter publisering relativt til det første året etter publisering, som er satt som referanse. Vi har sett bort fra siteringer i selve publiseringsåret. Antall siteringer er på topp to år etter publisering og etter seks år er antall siteringer på nivå med hva det var det første året etter publisering. Artikler skrevet av yngre forskere får i forventning en god del flere siteringer enn artikler skrevet av eldre forskere (figur 12b). Derimot ser det ikke ut som alder ved avlagt doktorgrad har noen betydning (figur 12c).

Siteringsraten som funksjon av publiseringsår har økt over tid for alle tre fagfelt, men minst for medisin og helse og desidert mest for humaniora og samfunnsfag (13). Vi minner om at siteringsraten er basert på artikler registrert i WoS. Resultatene gjenspeiler nok derfor i stor grad at dekningsgraden for humaniora og samfunnsfag har vært dårlig i WoS, og trolig har bedret seg betraktelig fra 2000 til 2016, hvor den forventede siteringsraten i samme periode er sjudoblet. Videre er det rimelig å anta at en del av den økende siteringsraten for alle fagfelt skyldes en generell økning i publiseringsgrad som igjen gir flere siteringer.

Som nevnt tidligere er variasjonen i de tilfeldige artikkel-effektene større enn i de tilfeldige forskereffektene ($\sigma_a > \sigma_f$), hvilket også gjenspeiles i figur 14. Det betyr at variasjonen i siteringer som ikke blir forklart av kovariatene, kan i større grad bli beskrevet av en artikkel-effekt enn en forskereffekt. Det kan tolkes som at modellen estimerer større variasjon



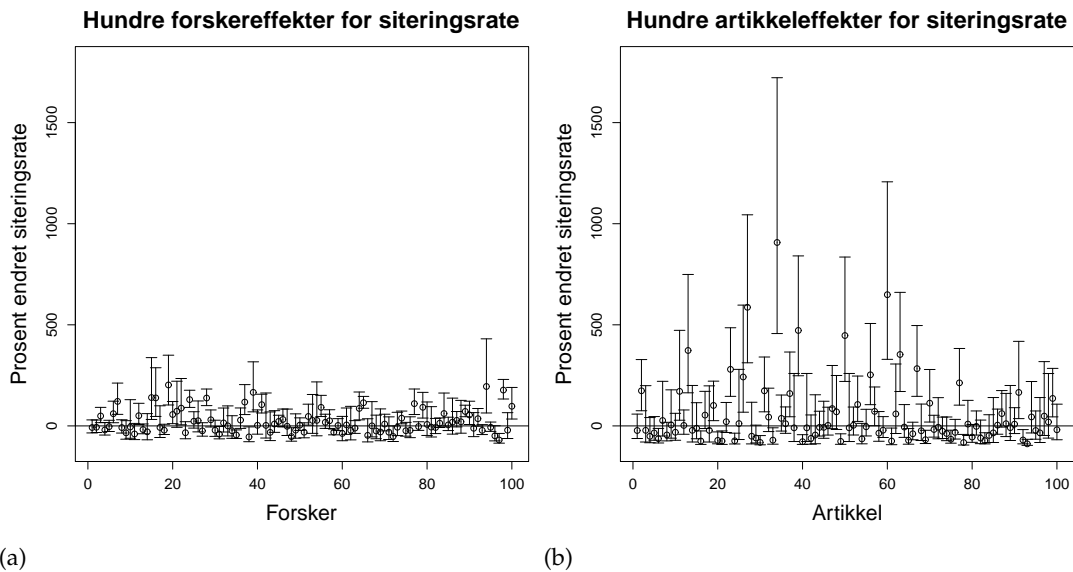
Figur 12. Estimerte effekter på siteringsrate av a) tid etter publikasjon (referanseår er første år etter publiseringsåret), b) alder ved publikasjon (referanseverdi er 50 år), c) alder ved doktordisputas (referanseverdi er 40 år).



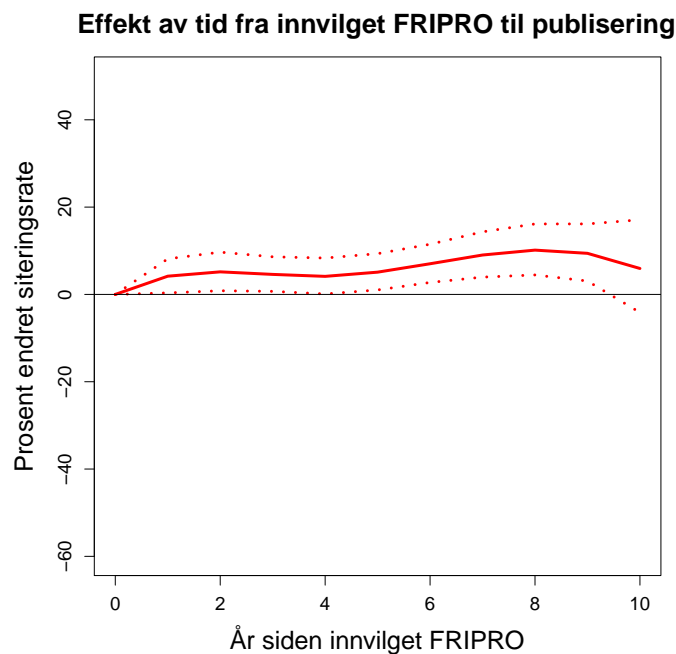
Figur 13. Estimerte årstalleffekter av forskerfagfelt på siteringsrate som funksjon av publiseringsår. Panel a) realfag, b) humaniora og samfunnsfag og c) medisin og helse. Referanseåret er 2011, slik at alle effekter er relativt til publiseringsraten i dette året.

innad i en forskers artikkelportefølge enn mellom gjennomsnittlig siteringer til hver enkelt forsker. Det er viktig å ta hensyn til slike underliggende strukturer for å kvantifisere usikkerheten av det vi estimerer.

Det gjenstår å presentere resultatene for effekten av å få tildelt et FRIPRO-prosjekt. Figur 15 viser at det er en signifikant økt siteringsrate for artikler publisert etter tildeling av et FRIPRO-prosjekt. Allerede for en artikkel publisert ett år etter tildeling er det en forventet økning på omtrent 5%, og denne er økt til omtrent 10% for artikler publisert åtte år etter tildelingsåret. Arbeidet til en artikkel publisert året etter tildeling vil imidlertid vanligvis være utført i forkant av prosjektet, slik at det er vanskelig å fortolke den del av kurven som effekt av en FRIPRO-tildeling.



Figur 14. Estimerte tilfeldige effekter på a) forskernivå og b) artikkelnivå i siteringsmodellen, for 100 ulike forskere og 100 ulike artikler. Dette skal ta høyde for individuelle forskjeller mellom forskerne og mellom artiklene som ikke lar seg forklare av kovariatene som er inkludert i modellen.



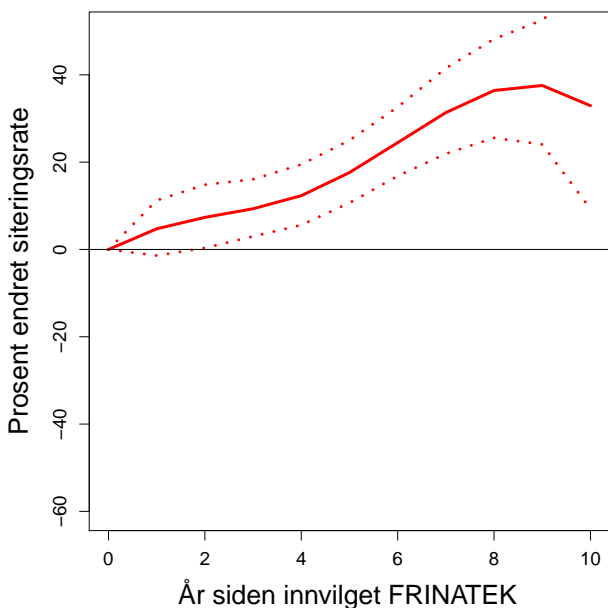
Figur 15. Estimert effekt på siteringsrate av tildeling av FRIPRO-midler som funksjon av tid fra tildelingsår til publikasjonsår for en artikkel som er publisert i etterkant av FRIPRO-tildeling.

Når vi deler FRIPRO-effekten inn i de ulike fagkomiteene blir bildet atskillig mer nyanisert (figur 16). For forskere som har fått tildelt et FRINATEK-prosjekt øker publiseringsraten jevnlig og er nær 40% høyere for en artikkel publisert åtte år etter tildeling enn en artikkel som er publisert i tildelingsåret. Derimot er tildeling av et FRIHUMSAM-prosjekt assosiert med en synkende siteringsrate. Dette kan synes som et merkelig resultat, og det

kan skyldes at denne effekten er blandet sammen med den svært kraftige generelle økningen i siteringsraten innen humaniora og samfunnsfag, og at usikkerheten som vises i figuren er underestimert. Når det gjelder FRIMEDBIO er tildeling av et slikt prosjekt assosiert med en økning i siteringsraten på omkring 5%. Når det gjelder STORFORSK er usikkerheten så stor at vi ikke kan trekke noen konklusjoner. Dette er som forventet, men STORFORSK er inkludert for kompletthetens skyld. Inndeling av FRIPRO-effekten i de fire underkategoriene ga forøvrig en signifikant bedre tilpasning til dataene (p-verdi $< 10^{-5}$).

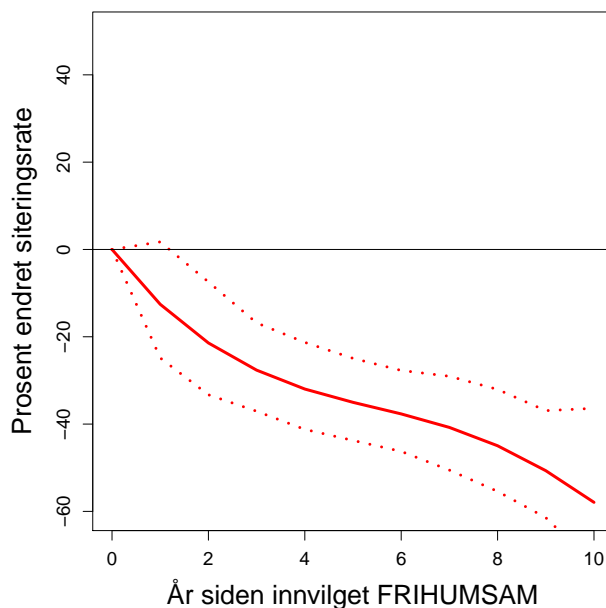
Ved bruk av AIC-kriteriet, endte vi for siteringsmodellen opp med å bruke fire basisfunksjoner for effekt av tid siden innvilget FRIPRO, fire basisfunksjoner for alder, fire basisfunksjoner for effekt av alder ved oppnådd doktorgrad, fem basisfunksjoner for effekt av tid siden publikasjon, ti basisfunksjoner for årstallseffekten og fire basisfunksjoner for effekt av tid siden innvilget FRIPRO (også ved inndeling i fagkomiteer).

Effekt av tid fra innvilget FRINATEK til publisering



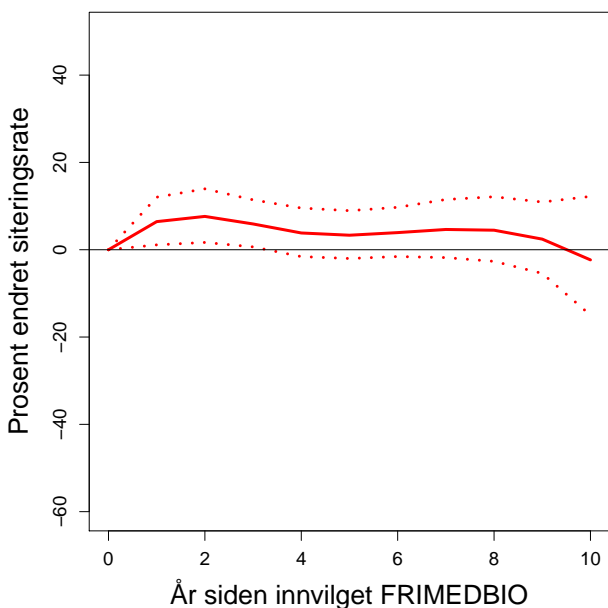
(a)

Effekt av tid fra innvilget FRIHUMSAM til publisering



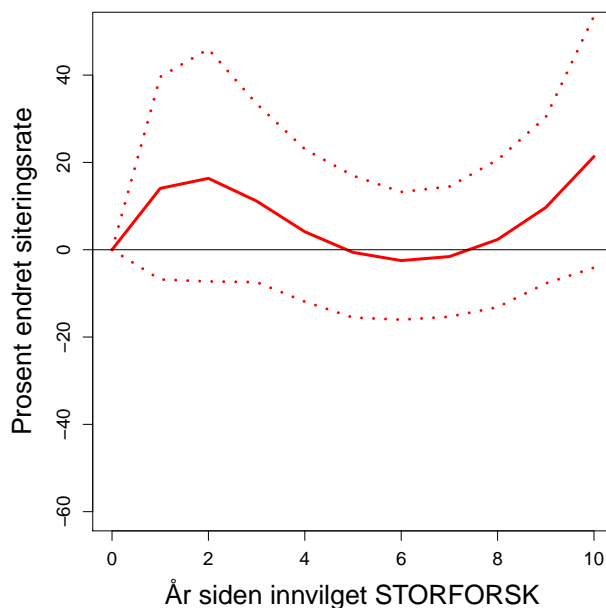
(b)

Effekt av tid fra innvilget FRIMEDBIO til publisering



(c)

Effekt av tid fra innvilget STORFORSK til publisering



(d)

Figur 16. Estimert effekt på siteringsrate av tildeling av FRIPRO-midler som funksjon av tid fra tildelingsår til publikasjonsår for en artikkel som er publisert i etterkant av FRIPRO-tildeling, fordelt på de ulike typer FRIPRO-prosjekter, a) FRINATEK, b) FRIHUMSAM, c) FRIMEDBIO og d) STORFORSK. For STORFORSK er det ekstra stor usikkerhet fordi det er svært få som har fått tildelt slike midler. Når det gjelder FRIHUMSAM må vi videre fortolke resultatet i lys av at det har vært en sju-dobling av den generelle siteringsraten innen humaniora og samfunnsfag i løpet av studieperioden, så det er en fare for at disse to effektene i noen grad er konfundert eller blandet sammen.

6 Diskusjon

I analysene vi har utført har vi så langt vi har kunnet korrigert for faktorer som kan tenkes å påvirke en forskers publiseringsrate og siteringsraten per artikkel, for på den måten å isolere effekten av å få en FRIPRO-tildeling fra de andre faktorene. Imidlertid er det en del momenter en bør ha i bakhodet når en fortolker resultatene.

For det første er det slik at observert assosiasjon ikke nødvendigvis impliserer en kausal sammenheng. Det er flere mulige forklaringer til at publikasjonsmodellen finner en positiv korrelasjonsstruktur mellom publiseringsraten og tid siden innvilget FRIPRO-prosjekt. En mulig forklaring, som vil være en kausal begrunnelse, er at forskeren på grunn av de bevilgede midlene økte sin vitenskapelig produktivitet, og dermed økte publiseringsraten. En annen begrunnelse, som ikke henviser til en kausal sammenheng, er at de som fikk innvilget et FRIPRO-prosjekt typisk ville være mer produktive uavhengig om de fikk FRIPRO-støtte. En slik selektiv økning av produktivitet, uavhengig av om forskeren fikk FRIPRO-støtte, kan for eksempel komme av at FRIPRO typisk tildeles forskere som har en høyere forventet fremtidig økning av produktivitet av grunner ikke inkludert i modellen. Hver søknad som behandles i FRIPRO gis en hovedkarakter mellom 1 og 7 basert på en totalbedømmelse, som blant annet inkluderer tidligere publiseringshistorie. I årene 2011-2016 har 32% av søknadene fått karakter 4 eller lavere, 37% har fått karakter 5, 24% karakter 6 og 6% karakter 7. Det er kun de søknadene med karakter 6 eller 7 som har blitt tildelt FRIPRO-midler.

At en artikkel har høy kvalitet betyr ikke nødvendigvis at den blir mye sitert. En viktig forklaringsvariabel for siteringsraten må være hvor mange som har lest og forstått artikkelen. Når en forsker får tildelt FRIPRO-prosjekt, har prosjektlederen mulighet til å danne seg et nettverk rundt forskningen sin. Dannelsen av et slikt nettverk kan medføre at flere personer leser og forstår artikler skrevet av prosjektlederen. En slik forklaring kan begrunne at siteringsraten øker med en gang forskeren får tildelt FRIPRO, se figur 15, da det er rimelig at artikler publisert under innvilgelsesåret kan være relatert til FRIPRO-prosjektet.

Vi gjentar at den noe overraskende negative effekten av tildeling av et FRIHUMSAM-prosjekt med hensyn på siteringsraten kan ha årsaker som ikke er relatert til tildeling av et slikt prosjekt. Som vi har påpekt, har den generelle siteringsraten for forskere innen humaniora og samfunnsfag økt kraftig i løpet av studieperioden, langt kraftigere enn den estimerte nedgangen i siteringsrate etter tildeling av et FRIHUMSAM-prosjekt. En mulig forklaring av den kraftige negative effekten av tildeling av et FRIHUMSAM-prosjekt kan være at de to effektene er blandet sammen. En blanding mellom de to effektene kan, for eksempel, forekomme hvis forskere som får avslått FRIHUMSAM-søknad i større grad samarbeider med forskere i fagfelt med høyere dekningsgrad eller med høyere siteringsrate. Hvis de to effektene er blandet, er det ikke overraskende at vi får et stort utslag for FRIHUMSAM-effekten da årstalleffekten er veldig stor for humaniora og samfunn.

I analysen har vi latt til og med 10 år etter innvilget prosjekt være en øvre grense for

hvor FRIPRO-kovariatene påvirker publikasjons- og siteringsraten, det vil si at effekten av FRIPRO er satt til null etter 10 år. Dette valget ble begrunnet med at det er vanskelig å tolke usikkerheten i områder med lite data når den er modellert som en kontinuerlig funksjon. Vi utførte to alternative analyser med ulik håndtering av observasjoner mer enn 10 år etter innvilget prosjekt. i) Slike data ble utelatt fra analysen. ii) Ingen øvre grense for FRIPRO-effekten. De to alternative analysene endret resultatene for FRIPRO-effekten opp til 10 år i liten grad. Gjentas analysen om noen år vil datagrunnlaget være bedre da det er høyere dekningsgrad i datakildene i nyere tid, og vi vil ha både fler og lengre tidsserier per forsker. Dermed vil FRIPRO-effekten også kunne estimeres for mer enn 10 år etter tildeling.

For siteringsanalysen har vi basert oss på publikasjoner i kategorien "Article" registrert i WoS. Det betyr blant annet at monografier ikke kommer med, og samtidig vet vi at WoS har hatt dårlig dekning for tidsskrifter innen humaniora og samfunnsfag. Vi vektet videre publikasjoner likt i alle analyser, det vil si at vi ikke skilte mellom publikasjoner der prosjektlederen var førsteforfatter eller medforfatter, og utnyttet ikke antallet forfattere til hver enkelt publikasjon.

Til sist er det viktig å huske på at det eksisterer alternative finansieringskilder. Det er dermed tilleggseffekten utover alternativ finansiering vi har estimert. Da analysen kun omhandler prosjektlederne, og det eksisterer andre finansieringskilder, vil trolig den totale effekten av et FRIPRO-prosjekt være større enn det vi har estimert. Vi ønsker dermed å påpeke at resultatene i analysen ikke kan direkte brukes til å måle nytten av FRIPRO.

Referanser

Devore, J. L. og Berk, K. N. (2007). *Modern mathematical statistics with applications*. Cengage Learning.

Hastie, T. J. og Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. og Bell, B. (2015). TMB: automatic differentiation and Laplace approximation. *arXiv preprint arXiv:1509.00660*.

Wood, S. (2017). *Package 'mgcv'*. R package version 1.8-23. Tilgjengelig fra: <http://CRAN.R-project.org/package=mgcv>.