Chain Ladder Metoden og Mack's Modell sammenlignet med Poissonbaserte modeller

av

Monica Sagosen

Oppgave for graden

Master i Statistikk
Finansteori og Forsikringsmatematikk

Universitetet i Bergen

Matematisk Institutt

Juni 2010

Takk

Jeg vil gjerne få takke min veileder Trygve S. Nilsen for hjelp med å finne en interessant oppgave, og god hjelp underveis i skriveprosessen.
Videre vil jeg også takke min forlovede Christian Holmefjord for moralsk støtte og oppmuntrende ord gjennom alle mine år ved universitetet.
Mine medstudenter vil jeg takke for lange lunchpauser og gode diskusjoner.
Til slutt vil jeg også takke familie og venner for støtte gjennom hele studietiden.

Innhold

1
Introduksjon

1.1 Innledning

Hensikten med denne oppgaven er hovedsakelig å se på reservering til fremtidige krav. Vi vil først og fremst se på Chain Ladder metoden og Mack sin modell. Dette vil bli gjort i henholdsvis kapittel 2 og kapittel 3, men før vi kommer til dette punktet vil vi i kapittel 1 se litt på det generelle hendelsesforløpet til en skade.

Etter at vi har studert Mack sin modell vil vi i kapittel 4 forsøke å simulere data etter en modell som passer med Mack sine tre antakelser, den sammensatte Poisson modellen. Til slutt vil vi i kapittel 5 tilpasse en overdispersert Poisson regresjon til dataene, før vi i kapittel 6 prøver å konkludere med hvilken metode som er å foretrekke, basert på de funn vi har gjort.

Oppgavens tre første kapittel er i hovedsak basert på kompendiumet *Lecture notes on estimating outstanding claims in general insurance* som er skrevet av Walter Neuhaus [Neuh 06].

Noe teori, hovedsakelig i kapittel 5, samt eksempelet som er brukt gjennom hele oppgaven, er tatt fra artikkelen *Stochastic claims reserving in general insurance* som er skrevet av P.D. England og R.J. Verrall [Engl 02].

1.2 Hendelsesforløpet

De fleste av oss vil en eller annen gang i løpet av livet komme i en situasjon der et forsikringsselskap er involvert. Hver eneste dag skjer det tusenvis av uhell eller ulykker, og vi skal nå gå gjennom det hendelsesforløpet som da blir satt i gang.

Først og fremst er man avhengig av å være forsikret mot det aktuelle tapet, man må ha tegnet en forsikringskontrakt. En generell skadeforsikringskontrakt løper som oftes over et forhåndsspesifisert tidsrom, vanligvis et år. Denne kontraken forteller personen som har tegnet den, heretter kalt forsikringstaker, hvilke rettigheter han har dersom det skulle inntreffe en skade i det tidsrommet kontrakten gjelder. Det er imidlertid også en del vilkår som følger med, der forsikringsselskapet speseifiserer tilfeller der de ikke vil dekke en eventuell skade. Dersom en slik skade inntreffer risikerer forsikringstaker at han ikke får noe dekket, men må ta hele tapet på egen kappe. Ofte vil slike skader uansett bli meldt til forsikringsselskapet, og skaden må da gjennom det samme behandligsforløpet.

En forsikringskontrakt, eller polise, spesifiserer hva eller hvem som er dekket, og hva de er dekket mot. Innen skadeforsikring er det stort sett ting som forsikres, man forsikrer huset, båten, bilen og hunden. Innen livsforsikring er det personer som forsikres, man kan forsikre seg mot død, sykdom, uførhet og så videre. I polisen vil det være spesifisert hva som er forsikret, dette være ting eller person. Dersom det er personer som er forsikret refereres disse til som de forsikrede. Det vil altså ikke alltid være samme person som er forsikringstaker og den forsikrede, vi kan f.eks. tenke oss tilfeller der en familiefar forsikrer hele familien mot ulykke. Han selv vil da være forsikringstaker, mens det enkelte familiemedlemmet er den forsikrede.

Dersom det skjer en skade må denne rapporteres inn til forsikringsselskapet. Man må da oppgi datoen skaden skjedde, og det er denne datoen som benyttes når forsikringsselskapet sjekker om skaden kan knyttes til en polise. Det er altså viktig at man hadde forsikring når skaden skjedde, det er ikke nok at man har forsikring når skaden blir rapportert. Dersom det er usikkert hva som er skadedatoen, som man f.eks. kan tenke seg er tilfellet dersom det har vært en liten vannlekkasje i et hus, som senere har ført til råte, blir det kalt inn eksperter som prøver å anslå skadedatoen. I slike tilfeller må forsikringsselskapet som forsikret huset på skadedatoen dekke skaden. Det er altså ikke uvanlig med forsinkelse mellom skadedatoen og datoen når skaden blir rapportert inn, kalt anmeldt dato. Denne forsinkelsen kan variere mellom et par timer til flere år. Man kan tenke seg flere grunner til at det blir en forsinkelse, som at

- Skaden oppdages ikke før det er gått en tid. Dette kan som nevnt ovenfor f.eks. skje ved vannskader i hus, men vi kan også tenke oss at ulike personskader kan ha lang forsinkelse, da det kan ta flere år før skaden virkelig får betydning for den forsikrede.

- Dersom den forsikrede selv har vært påvirket av skaden, som i en ulykke, kan det ta tid før han er bra nok til å få rapportert inn skaden.

- I forbindelse med helligdager som f.eks. jul vil det naturlig nok bli en forsinkelse, da alt er stengt.

- Dersom skaden skjer på hytten er det ikke sikkert den blir oppdaget før det er gått en stund.

Når skaden er blitt anmeldt, blir den registrert i systemet, med skadedato og anmeldt dato. Nå må forsikringsselskapet ta stilling til om skaden er noe som de vil dekke, eller om den faller inn under de tilfellene som er spesifisert i villkårene at ikke dekkes. Alle skader, eller krav, som forsikringsselskapet holder på med refereres til som "åpne". På mindre skader, der man gjerne har små utbetalinger, kan kravet bli avgjort ganske raskt. Et typisk eksempel på dette vil være reiseforsikring, som blandt annet vil dekke ekstra utgifter til kjøp av nødvendige ting dersom bagasjen blir borte. Dersom det er en litt større skade vil det nok ta litt tid før saken er avsluttet. Forsikringsselskapet må da gjøre en vurdering av skadens omfang og forventet kostnad, og muligens også gjøre inspeksjoner på skadestedet. Er det personskader involvert må mest sannsynlig en lege konsulteres, og alt slikt tar tid. I denne perioden betales det stort sett ikke ut noen store beløp, men muligens noen delutbetalinger som kan dekke takstutgifter, legekonsultasjoner og andre småutgifter.

De store utbetalingen vil komme når skaden er blitt evaluert. Dersom et hus har brent ned vil gjenoppbyggingen komme i gang, er det personskade vil den medisinske behandlingen og rehabiliteringen starte. I denne perioden er den totale kostnaden for skaden ikke kjent, og må vurderes fortløpende. Kravet vil bli sett på som "åpent" frem til det ikke forventes at det vil komme flere utbetalinger. Hvor lang tid dette tar er svært ulikt, snakker vi om et hus som må bygges opp igjen vil kravet være åpent frem til huset står ferdig. Er det personskader vi snakker om kan det gå svært lang tid, da det ofte tar mange år å bli totalt rehabilitert. Har man for eksempel hatt en uføreforsikring og man blir ufør, kan utbetalingene pågå resten av livet.

Når man ikke forventer at det skal komme flere utbetalinger blir kravet "lukket". Dersom det senere skulle vise seg at den forsikrede ikke var fornøyd med kompensasjonen han mottok, eller det kommer nye opplysninger om saken, kan kravet åpnes igjen, og må da gå gjennom de samme stegene en gang til. Man kan også tenke seg at den forsikrede ikke kommer til enighet med forsikringsselskapet, og i slike tilfeller kan kravet bli tatt for retten.

I det øyeblikket forsikringsselskapet får rapportert en skade, skal det sette av så mye penger som det tror det totale kravet kommer til å bli på. Dette beløpet kan de sette på grunnlag av tidligere erfaring, og man kan skille mellom to estimater; estimatet på det totale kravet, som er et estimat på hva kravets totale utbetaling kommer til å bli, og estimatet på det gjenstående kravet, som er et estimat på hva de fremtidige utbetalingene kommer til å bli. Naturlig nok er kravets totale estimat summen av det gjenstående estimatet og tidligere utbetalinger. Det man ønsker er at det totale estimatet for kravet skal holde seg noenlunde konstant over tid, mens det utestående estimatet skal gå mot null.

1.3 Reservering til fremtidige krav

Hvert eneste år blir det anmeldt mange skader som har hatt lang forsinkelse, altså skader som ikke har inntruffet i inneværende år. I Norge er alle forsikringsselskap forpliktet til å sette av beløp til å dekke slike skader i det året skaden faktisk skjedde. Forsikringsselskapene må altså hvert år reservere penger til krav som de ikke vet om, men som vil komme i fremtiden. Dette kalles IBNR reservering, der IBNR står for "Incurred But Not Reported".

Estimeringen av IBNR-reserver er en av de viktigste jobbene for en aktuar i et forsikringsselskap, da disse estimatene vil påvirke hele selskapets lønnsomhet. Dersom det reserveres for mye penger kan det se ut som om det går dårlig for selskapet, og prisene på forsikring settes muligens opp, noe som ikke er populært hos kundene. Om det reserveres for lite kan det se ut som om det går veldig bra, og prisene settes ned. Dersom det da kommer en eller flere store uforutsette skader fra tidligere skadeår kan det få alvorlige konsekvenser for selskapet.

Det er altså svært viktig at selskapene reserverer riktige beløp. Det eksisterer mange metoder for å beregne reservene, og en av dem er Chain Ladder metoden som vi skal se nærmere på i de neste kapittelene.

2 Chain Ladder Metoden

2.1 Generelle benevnelser

En forsikringsperiode går vanligvis over et år, men kan også gå over kortere tidsperioder som f.eks. kvartaler eller måneder. Til tross for dette vil standard notasjon for de diskrete tidsperiodene i denne oppgaven være "år".

Vi befinner oss nå ved slutten av år J. Året da skaden skjedde benevner vi med j, og dersom vi lar det første året vi har data fra benevnes med år 1, vil vi ha at $j = 1, \cdots, J$. Antall år vi har observert utviklingen til en skade avhenger av hvor lang tid det har gått siden skaden inntraff. Dersom skaden skjedde i år 1 vil vi ha $J - 1$ år med utvikling, mens en skade som inntraff i år J vil ikke ha noen kjent utvikling. Det året skaden inntreffer kaller vi for utviklingsår 0, da det ikke er noen kjent utvikling i dette året. Antall utviklingsår for en skade som inntraff i år j vil altså være $e = J - j$, der $e = 0, \cdots, D$. Benevnelsen D er for det maksimale antall år med utvikling som kan observeres, og vi vil da ha at $D = J - 1$. Indirekte har vi da antatt at når en skade har utviklet seg i D år, så vil det ikke komme noen flere utbetalinger, og skaden vil være lukket. Dette vil si at vi har ett år som er fullt utviklet, for skader som inntraff i skadeår 1 vil alle utviklingsårene fra 0 til D være kjent.

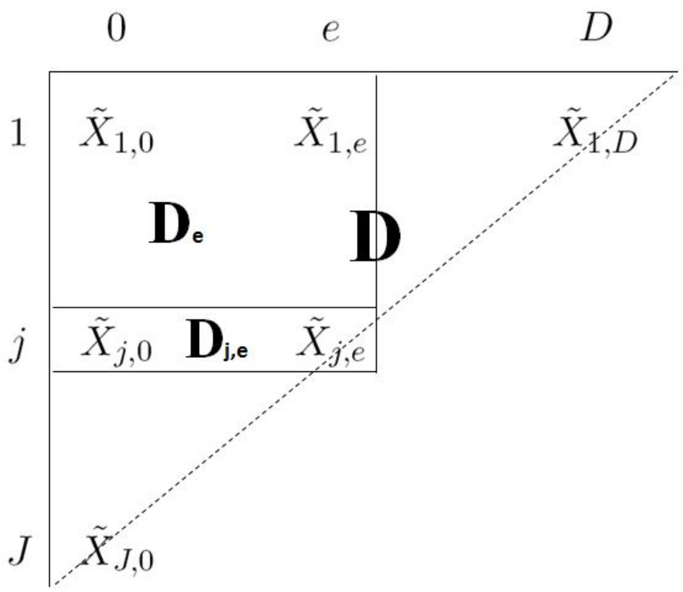
I denne oppgaven vil vi ikke se på enkeltskader, men på de totale kravene for de ulike skadeårene. Benevnelsen $X_{j,e}$ vil representere den inkrementelle endringen for det totale kravet, som skjer i løpet av utviklingsår e, for skader som inntraff i skadeår j. Dette tallet kan altså være både positiv eller negativ, men i de fleste tilfeller positivt. I de tilfellene der vi får en negativ inkrementell endring kan vi tenke oss at noen skader har blitt lukket, og de totale utbetalingene ble mindre enn forventet. Da vil disse skadene generere et negativt tilskudd til det totale kravet, og dersom dette tilskuddet er større enn økningen, vil vi få en negativ inkrementell endring.

Det vil også være hensiktsmessig å ha et uttrykk for den kumulative statusen for hvert skadeår og utviklingsår, og vi bruker benevnelsen $\tilde{X}_{j,e}$. Dette vil altså være de akkumulerte totale kravene for skader som inntraff i år j, frem til utviklingsår e. Med andre ord: summen av de inkrementelle endringene frem til utviklingsår e. For enkelhets skyld definerer vi også $\tilde{X}_j = \tilde{X}_{j,D}$.

2.2 Chain Ladder metoden

Chain Ladder metoden går ut på at de fremtidige kravene som vil komme, på skader som allerede har skjedd men ikke er anmeldt, kan estimeres ved å se på utviklingen til de tidligere skadeårene. Det gjøres da en antakelse om at skadeårene vil utvikle seg på samme måte. Dette er en ganske annerkjent metode, den vil naturlig nok ikke gi helt riktige svar, men dersom man har mange skadeår og utviklingsår å basere estimatet på, er det kanskje en bra tilnærming.

Som nevnt tidligere antar vi at vi befinner oss ved slutten av år J, og vi har observert $j = 1, \cdots, J$ skadeår. Utviklingen for disse skadeårene vil være $e = 0, \cdots, D$, og dette kan illustreres ved den velkjente trekanten, der vi har skadeår på den loddrette aksen og utviklingsår på den horisontale aksen:



Her vil de data som er kjent betegnes med $\mathbf{D} = \{\tilde{X}_{j,e} : j = 1, \cdots, J, e = 0, \cdots, J - j\}$. Vi ønsker også å ha en betegnelse for de kjente data for skadeår j opp til utviklingsår e, og disse benevnes med $\mathbf{D}_{j,e} = \{\tilde{X}_{j,e'} : e' = 0, \cdots, e\}$. Vi har fremdeles at $D = J - 1$.

Vi ønsker nå å fylle ut nedre halvdel av denne matrisen med de estimerte verdiene for de fremtidige akkumulerte kravene. For å få til dette er vi avhengig av å vite hvor mye de akkumulerte kravene endrer seg fra et år til det neste. Ved å basere oss på de kjente observasjonene ønsker vi altså å finne et forholdstall, som vil representere økningen i det totale akkumulerte kravet. Vi vil få et forholdstall for hvert utviklingsår, kalt δ_e^*, og for forsinkelse $e = 1, \cdots, D$ kan disse empiriske utviklingsfaktorene beregnes ut fra følgende likning:

$$\delta_e^* = \frac{\sum_{j=1}^{J-e} \tilde{X}_{j,e}}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}} \tag{2.1}$$

I artikkelen til Neuhaus står det at denne likningen kun gjelder fra $e = 2$, men som vi straks skal se er det ingen grunn til at den ikke skal kunne gjelde for $e = 1$ også.

La oss nå se litt nærmere på hva denne likningen egentlig forteller oss. Antar at vi nå befinner oss i år 10, altså $J = 10$, og at vi er interessert i å estimere utviklingsfaktoren for år 7, $e = 7$. Denne vil da være gitt ved

$$\delta_7^* = \frac{\sum_{j=1}^{3} \tilde{X}_{j,e}}{\sum_{j=1}^{3} \tilde{X}_{j,e-1}} = \frac{\tilde{X}_{1,7} + \tilde{X}_{2,7} + \tilde{X}_{3,7}}{\tilde{X}_{1,6} + \tilde{X}_{2,6} + \tilde{X}_{3,6}}$$

Vi tar altså summen av alle observasjonene vi har for utviklingsår 7, og deler dette på summen av observasjonene for utviklingsår 6, for de tilsvarende skadeårene. Legg altså merke til at begge summene går over like mange observasjoner. Dette vil gi oss et forholdstall mellom totale akkumulerte krav frem til utviklingsår 7, i forhold til hva de totale akkumulerte kravene, for samme skadeår, var i utviklingsår 6. Dette tallet representerer gjennomsnittlig endring for det totale kravet i utviklingsår 7. Dersom denne er lik 1 kan man konkludere med at det i snitt ikke er noen endring, er den større enn 1 har det totale akkumulerte kravet i gjennomsnitt økt, og er den mindre enn 1 har det totale akkumulerte kravet minket i snitt. I de fleste tilfeller vil vi ha at $\delta_e^* > 1$, men ettersom e nærmer seg D vil den gå mot 1.

La oss nå se hva som skjer dersom vi prøve å finne utviklingsfaktoren for $e = 1$, δ_1^*. Dersom vi igjen bruker likning 2.1 vil vi få at

$$\delta_1^* = \frac{\sum_{j=1}^{9} \tilde{X}_{j,e}}{\sum_{j=1}^{9} \tilde{X}_{j,e-1}} = \frac{\tilde{X}_{1,1} + \tilde{X}_{2,1} + \cdots + \tilde{X}_{9,1}}{\tilde{X}_{1,0} + \tilde{X}_{2,0} + \cdots + \tilde{X}_{9,0}}$$

Som vi ser er dette uttrykket like meningsfylt som det vi hadde for δ_7^*, det vil representere den gjennomsnittlige endringen som skjer fra utviklingsår 0 til 1. I resten av denne oppgaven vil vi derfor la e gå fra 1 til D i likning 2.1.

Videre ønsker vi også et estimat på hva de totale akkumulerte kravene for hvert skadeår kommer til å bli, altså status ved slutten av utviklingsår D. For skadeår $j = 2, \cdots, J$ er dette estimatet gitt ved

$$\bar{X}_j = \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D} \delta_e^* \tag{2.2}$$

La oss studere denne likningen litt grundigere. Som ovenfor antar vi at vi befinner oss ved slutten av år 10, noe som fører til at det maksimale antall utviklingsår som kan observeres er 9, altså $D = 9$. Dersom vi ønsker et estimat for det totale akkumulerte kravet for skadeår 3, blir likningen

$$\bar{X}_3 = \tilde{X}_{3,7} \cdot \prod_{e=8}^{9} \delta_e^* = \tilde{X}_{3,7} \cdot \delta_8^* \cdot \delta_9^*$$

For å finne et estimat på det totale akkumulerte kravet tar vi altså den siste kjente observasjonen, som i dette tilfellet er verdien vi har for utviklingsår 7, og multipliserer denne med de estimerte utviklingsfaktorene for de resterende utviklingsårene, δ_8^* og δ_9^*. Disse utviklingsfaktorene baserer seg, som vi har sett tidligere, på de kjente observasjonene for de aktuelle utviklingsårene. Det er her antakelsen om at skadeårene vil oppføre seg likt kommer inn i bildet, ved å bruke utviklingsfaktorene for utviklingsår 8 og 9 antar man indirekte at skadeår 3 vil oppføre seg på samme måte som skadeår 1 og 2.

Ved hjelp av disse to likningene kan vi altså estimere de totale akkumulerte kravene for alle skadeårene. For å kunne vurdere om dette er en god modell må vi imidlertid også vite noe om det forventede kvadratiske avviket til disse estimatene. I neste kapittel vil vi se nærmere på hvordan Mack gjør noen antakelser for å finne et uttrykk for akkurat dette.

3 Mack sin modell

For å kunne vurdere om Chain Ladder metoden gir gode estimat på de forventede totale akkumulerte kravene, må vi se litt grundigere på det forventede kvadratiske avviket (MSE). Vi begynner da med det generelle uttrykket for dette, før vi vil utføre en del matematiske utregninger slik at vi får et uttrykk som vi kan regne ut ved hjelp av de data vi har. For å komme til denne endelige likningen må det imidlertid gjøres noen antakelser, og det er her Mack's modell kommer inn. Dette vil bli kommentert mer inngående senere, la oss nå begynne med det generelle uttrykket.

3.1 Generelt uttrykk for MSE

For hvert skadeår j har vi nå funnet et uttrykk for forventede totale akkumulerte krav, $\bar{X}_j$, basert på de observerte data $\mathbf{D}$. Det er som sagt av interesse å finne ut hvor mye vi forventer at dette estimatet avviker fra det faktiske kravet, og vi ønsker nå å finne et uttrykk for MSE for prediktoren $\bar{X}_j$:

$$E[(\tilde{X}_j - \bar{X}_j)^2|\mathbf{D}]$$

Her vil $\bar{X}_j$ være en størrelse som kan beregnes ut fra de data vil har tilgang på, så denne vil være kjent og kan betraktes som en konstant. Den andre størrelsen $\tilde{X}_j$ vil være en stokastisk variabel. Dette uttrykket kan vi ikke finne direkte, så vi må omformulere det. Vi bruker da en ganske vanlig fremgangsmåte, der vi legger til og trekker fra samme tall. Etter at det er gjort multipliserer vi ut parantesen, før vi går videre til å vurdere hvert ledd for seg selv.

$$
\begin{aligned}
\mathrm{E}\left\{\left(\tilde{X}_j-\bar{X}_j\right)^2|\mathbf{D}\right\} &= \mathrm{E}\left\{\left(\tilde{X}_j-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)+\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)^2|\mathbf{D}\right\} \\
&= \mathrm{E}\left\{\left(\tilde{X}_j-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)\right)^2|\mathbf{D}\right\} \\
&+ \mathrm{E}\left\{\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)^2|\mathbf{D}\right\} \\
&+ 2\cdot\mathrm{E}\left\{\left(\tilde{X}_j-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)\right)\cdot\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)|\mathbf{D}\right\}
\end{aligned}
\tag{3.1}
$$

Vi vil nå studere leddene på høyre side i denne likningen litt nærmere.
Begynner med det første leddet:

$$\mathrm{E}\left\{\left(\tilde{X}_j-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)\right)^2|\mathbf{D}\right\} = \mathrm{Var}\left\{\tilde{X}_j|\mathbf{D}\right\}$$

Denne overgangen får vi direkte fra definisjonen av betinget varians.
Vi går videre med det andre leddet på høyre side i 3.1:

$$\mathrm{E}\left\{\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)^2|\mathbf{D}\right\}$$

Dersom $\mathbf{D}$ er kjent, så vet vi at også $\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)$ kjent, og kan betraktes som en konstant. Det er også klart at dersom $\mathbf{D}$ er kjent så kan $\bar{X}_j$ sees på som en konstant. Alt som er opphøyd i andre potens er da konstant når det betinges på $\mathbf{D}$, og kan bli tatt utenfor forventningen. Uttrykket kan da skrives som:

$$\mathrm{E}\left\{\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)^2|\mathbf{D}\right\} = \left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right\}^2$$

Videre ser vi nå litt nærmere på kryssleddet fra 3.1:

$$\mathrm{E}\left\{\left(\tilde{X}_j-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)\right)\cdot\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)|\mathbf{D}\right\}$$

Siden alt er betinget på $\mathbf{D}$ kan leddet $\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)$ betraktes som en konstant, og derfor bli tatt utenfor forventningen. Vi står da igjen med den betingede forventningen til differansen mellom $\tilde{X}_j$ og dens forventede verdi, som vil være 0:

$$
\begin{aligned}
&\left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)\cdot\mathrm{E}\left\{\left(\tilde{X}_j-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)\right)|\mathbf{D}\right\} \\
&= \left(\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right)\cdot\left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)\right\} \\
&= 0
\end{aligned}
$$

Dersom vi nå setter sammen det vi har funnet, ser vi at MSE for estimatoren $\bar{X}_j$, likning

3.1, kan skrives som:

$$\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)^2 |\mathbf{D}\right\} = \mathrm{Var}\left\{\tilde{X}_j|\mathbf{D}\right\} + \left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right\}^2 \tag{3.2}$$

3.2 Mack sin modell

For å kunne fortsette utregningen av dette uttrykket gjøres det nå tre modellantakelser. Det er disse antakelsene som danner grunnlaget for Mack sin modell, og de er som følger:

1 Det eksisterer konstanter $\delta_1, \delta_2, \cdots, \delta_D$ slik at

$$\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,e-1}\right) = \delta_e \tilde{X}_{j,e-1} \quad \text{for} \quad e = 1, \cdots, D$$

2 Det eksisterer konstanter $\gamma_1, \gamma_2, \cdots, \gamma_D$ slik at

$$\mathrm{Var}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,e-1}\right) = \gamma_e \tilde{X}_{j,e-1} \quad \text{for} \quad e = 1, \cdots, D$$

3 Historiene $\mathbf{D}_{j,D}$ og $\mathbf{D}_{k,D}$ er uavhengige for $j \neq k$.

I de videre utregningene vil dette bli referert til som antakelse 1, 2 og 3.

Utviklingsfaktorene $\delta_1, \delta_2, \cdots, \delta_D$ er ukjente, men vi bruker $\delta_1^*, \cdots, \delta_D^*$ som estimatorer for disse. Mack hevder videre at disse estimerte utviklingsfaktorene vil være forventningsrette og ukorrelerte, og dette vil vi nå bevise.

Vi har fra tidligere (se likning 2.1) at

$$\delta_e^* = \frac{\sum_{j=1}^{J-e} \tilde{X}_{j,e}}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}}$$

Dersom vi nå tar forventningen av dette, og videre bruker regelen om dobbel forventning, får vi

$$\begin{aligned} \mathrm{E}\left[\delta_e^*\right] &= \mathrm{E}\left[\frac{\sum_{j=1}^{J-e} \tilde{X}_{j,e}}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}}\right] \\ &= \mathrm{E}\left[\mathrm{E}\left(\frac{\sum_{j=1}^{J-e} \tilde{X}_{j,e}}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}} |\mathbf{D}_{e-1}\right)\right] \end{aligned}$$

Her betinger vi på $\mathbf{D}_{e-1}$, altså alt som er observert opp til utviklingsår $e-1$. Vi kan da ta summen i nevneren utenfor den innerste forventningen, siden denne da kan sees på som en konstant.

$$\begin{aligned}
\mathrm{E}\left[\delta_e^*\right] &= \mathrm{E}\left[\frac{1}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\mathrm{E}\left(\sum_{j=1}^{J-e}\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right)\right] \\
&= \mathrm{E}\left[\frac{1}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\sum_{j=1}^{J-e}\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right)\right]
\end{aligned}$$

Her vil det ikke spille noen rolle om det betinges på $\mathbf{D}_{e-1}$ eller om det betinges på $\mathbf{D}_{j,e-1}$, da sistnevnte vil gi like mye informasjon her, siden det kun er skadeår j som til enhver tid er interessant. Videre kan vi bruke antakelse 1, og får at

$$\begin{aligned}
\mathrm{E}\left[\delta_e^*\right] &= \mathrm{E}\left[\frac{1}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\sum_{j=1}^{J-e}\delta_e\tilde{X}_{j,e-1}\right] \\
&= \mathrm{E}\left(\delta_e\right) \\
&= \delta_e
\end{aligned} \tag{3.3}$$

Den siste overgangen kommer også fra antakelse 1, som sier at $\delta_1, \cdots, \delta_D$ er konstanter, og vi har nå bevist at estimatorene for δ_e er forventningsrette.
Ønsker videre å se på påstanden om at de også er ukorrelerte, dvs. $\mathrm{Cov}(\delta_e^*, \delta_{e'}^*) = 0$ for $e \neq e'$. Begynner da med å se litt grundigere på $\mathrm{E}(\delta_e^*\delta_{e'}^*)$, da denne vil behøves til denne utregningen.

$$\mathrm{E}\left[\delta_e^*\delta_{e'}^*\right] = \mathrm{E}\left[\frac{\sum_{j=1}^{J-e}\tilde{X}_{j,e}}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\cdot\frac{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'}}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}\right]$$

Antar at $e < e'$, slik at når vi bruker regelen om dobbel forventning betinger vi på det som vil gi mest informasjon, altså $D_{e'-1}$. Kan da ta hele uttrykket for δ_e^* utenfor den innerste forventningen, samt nevneren fra uttrykket for $\delta_{e'}^*$:

$$\begin{aligned}
\mathrm{E}\left[\delta_e^*\delta_{e'}^*\right] &= \mathrm{E}\left[\mathrm{E}\left(\frac{\sum_{j=1}^{J-e}\tilde{X}_{j,e}}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\cdot\frac{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'}}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}|\mathbf{D}_{e'-1}\right)\right] \\
&= \mathrm{E}\left[\frac{\sum_{j=1}^{J-e}\tilde{X}_{j,e}}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\cdot\frac{1}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}\mathrm{E}\left(\sum_{j=1}^{J-e'}\tilde{X}_{j,e'}|\mathbf{D}_{e'-1}\right)\right]
\end{aligned}$$

I det siste leddet har vi her forventningen til en sum, men vi kan uten problemer bytte om rekkefølgen, slik at vi får summen til forventningene. Det spiller heller ingen rolle om det betinges på $\mathbf{D}_{e'-1}$ eller $\mathbf{D}_{j,e'-1}$ da disse gir like mye nyttig informasjon. Dersom vi i tillegg bruker antakelse 1 også her, får vi at:

$$\begin{aligned}
\mathrm{E}\left[\delta_e^*\delta_{e'}^*\right] &= \mathrm{E}\left[\frac{\sum_{j=1}^{J-e}\tilde{X}_{j,e}}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\cdot\frac{1}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}\sum_{j=1}^{J-e'}\delta_{e'}\tilde{X}_{j,e'-1}\right] \\
&= \mathrm{E}\left[\frac{\sum_{j=1}^{J-e}\tilde{X}_{j,e}}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}\cdot\delta_{e'}\right] \\
&= \delta_{e'}\mathrm{E}\left(\delta_e^*\right) \\
&= \delta_{e'}\delta_e
\end{aligned}$$

Dette viser at $\mathrm{Cov}(\delta_e^*, \delta_{e'}^*) = 0$, og siden dette resonementet kan gjentas for $e > e'$, kan vi konkludere med at Mack sin påstand om at δ_e^* og $\delta_{e'}^*$ er ukorrelerte for $e \neq e'$, er korrekt.

Siden heller ikke konstantene $\gamma_1, \cdots, \gamma_D$ er kjente, trenger vi en estimator for disse. Mack hevder da at estimatoren

$$\gamma_e^* = \frac{1}{J-e-1}\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e^*\right)^2 \tag{3.4}$$

vil være forventningsrett for γ_e, og dette vil vi nå bevise. Vi begynner da med å se på den betingede forventningen

$$\mathrm{E}\left[(J-e-1)\gamma_e^*|\mathbf{D}_{e-1}\right] = \mathrm{E}\left[\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e^*\right)^2|\mathbf{D}_{e-1}\right] \tag{3.5}$$

Ved å bruke samme fremgangsmåte som tidligere, der vi legger til og trekker fra en konstant, samt antakelse 3 om uavhengighet mellom skadeår, kan dette uttrykket skrives som

$$\begin{aligned}
&\mathrm{E}\left[\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e - \delta_e^* + \delta_e\right)^2|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\mathrm{E}\left[\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e\right)^2|\mathbf{D}_{e-1}\right] \\
&+ \sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\mathrm{E}\left[\left(\delta_e^* - \delta_e\right)^2|\mathbf{D}_{e-1}\right] \\
&- 2\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\mathrm{E}\left[\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e\right)\left(\delta_e^* - \delta_e\right)|\mathbf{D}_{e-1}\right]
\end{aligned} \tag{3.6}$$

Vi vil nå studere de tre uttrykkene på høyre side av likningen hver for seg. Før vi begynner vil vi imidlertid se på forventningen til $\tilde{X}_{j,e}/\tilde{X}_{j,e-1}$, da denne vil være nyttig senere.

Denne utregningen blir på samme måte som i likning 3.3, og Mack sin antakelse om forventning blir brukt:

$$
\begin{aligned}
\mathrm{E}\left[\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}\right] &= \mathrm{E}\left[\mathrm{E}\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}|\mathbf{D}_{e-1}\right)\right] \\
&= \mathrm{E}\left[\frac{1}{\tilde{X}_{j,e-1}}\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right)\right] \\
&= \mathrm{E}\left[\frac{1}{\tilde{X}_{j,e-1}}\cdot\delta_e\tilde{X}_{j,e-1}\right] \\
&= \mathrm{E}\left[\delta_e\right] \\
&= \delta_e
\end{aligned}
$$

Vi går nå videre med utregningen av det første uttrykket på høyre side i likning 3.6, og vil i denne utregningen benytte Mack sin antakelse om varians.

$$
\begin{aligned}
&\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\mathrm{E}\left[\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}-\delta_e\right)^2|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\mathrm{E}\left[\left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}-\mathrm{E}(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}})\right)^2|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\mathrm{Var}\left[\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\frac{1}{\tilde{X}^2_{j,e-1}}\mathrm{Var}\left[\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e}\frac{1}{\tilde{X}_{j,e-1}}\gamma_e\tilde{X}_{j,e-1} \\
&= \sum_{j=1}^{J-e}\gamma_e \\
&= (J-e)\gamma_e
\end{aligned}
\tag{3.7}
$$

Vi fortsetter med utregningen av det andre leddet på høyre side i likning 3.6.
I denne utregningen vil vi benytte Mack sin andre og tredje antakelse, samt uttrykket

for δ_e^*, som vi kan finne fra likning 2.1.

$$
\begin{aligned}
&\sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{E}\left[(\delta_e^* - \delta_e)^2 | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{E}\left[\left(\delta_e^* - \mathrm{E}(\delta_e^*)\right)^2 | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{Var}\left[\delta_e^* | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{Var}\left[\frac{\sum_{k=1}^{J-e} \tilde{X}_{k,e}}{\sum_{k=1}^{J-e} \tilde{X}_{k,e-1}} | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \frac{1}{(\sum_{k=1}^{J-e} \tilde{X}_{k,e-1})^2} \mathrm{Var}\left[\sum_{k=1}^{J-e} \tilde{X}_{k,e} | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \frac{1}{(\sum_{k=1}^{J-e} \tilde{X}_{k,e-1})^2} \sum_{k=1}^{J-e} \mathrm{Var}\left[\tilde{X}_{k,e} | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \frac{1}{(\sum_{k=1}^{J-e} \tilde{X}_{k,e-1})^2} \sum_{k=1}^{J-e} \gamma_e \tilde{X}_{k,e-1} \\
&= \gamma_e
\end{aligned}
\tag{3.8}
$$

Til slutt ser vi på det kryssleddet. Vi bruker regelen for kovarians til å skrive dette litt om, noe som gir oss

$$
\begin{aligned}
&\sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{E}\left[(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e)(\delta_e^* - \delta_e) | \mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{Cov}\left[(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e), (\delta_e^* - \delta_e) | \mathbf{D}_{e-1}\right] \\
&+ \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \mathrm{E}\left[(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e) | \mathbf{D}_{e-1}\right] \mathrm{E}\left[(\delta_e^* - \delta_e) | \mathbf{D}_{e-1}\right]
\end{aligned}
$$

Her vil det andre leddet på høyre side bli 0, da vi allerede har vist at δ_e^* er forventningsrett. Dersom vi nå bruker defininsjonen på δ_e^*, samt bruker antakelsen om uavhengighet mellom skadeår fra linje 4 til 5, og antakelsen om varians fra linje 5 til 6, finner vi for det første leddet at

$$\begin{aligned}
&\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}\text{Cov}\left[(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e), (\delta_e^* - \delta_e)|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1}\text{Cov}\left[(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}), \delta_e^*|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1}\text{Cov}\left[\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}}, \frac{\sum_{k=1}^{J-e}\tilde{X}_{k,e}}{\sum_{k=1}^{J-e}\tilde{X}_{k,e-1}}|\mathbf{D}_{e-1}\right] \\
&= \sum_{j=1}^{J-e} \tilde{X}_{j,e-1}\frac{1}{\tilde{X}_{j,e-1}\cdot\sum_{k=1}^{J-e}\tilde{X}_{k,e-1}}\text{Cov}\left[\tilde{X}_{j,e}, \sum_{k=1}^{J-e}\tilde{X}_{k,e}|\mathbf{D}_{e-1}\right] \\
&= \frac{1}{\sum_{k=1}^{J-e}\tilde{X}_{k,e-1}}\sum_{j=1}^{J-e}\text{Var}\left[\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right] \\
&= \frac{1}{\sum_{k=1}^{J-e}\tilde{X}_{k,e-1}}\sum_{j=1}^{J-e}\gamma_e\tilde{X}_{j,e} \\
&= \gamma_e
\end{aligned} \tag{3.9}$$

Dersom vi nå går tilbake til den betingede forventningen vi var intressert i, likning 3.5, og setter inn uttrykkene vi har funnet, får vi at

$$\text{E}\left[(J-e-1)\gamma_e^*|\mathbf{D}_{e-1}\right] = 3.7 + 3.8 - 2\cdot 3.9 = \gamma_e + (J-e)\gamma_e - 2\gamma_e = (J-e-1)\gamma_e$$

Dette impliserer at $\text{E}\left[\gamma_e^*|\mathbf{D}_{e-1}\right] = \gamma_e$.
Opprinnelig var det forventningen for γ_e^* vi var interessert i, og med de utregningene som nå er gjort kan vi konkludere med at

$$\text{E}\left[\gamma_e^*\right] = \text{E}\left\{\text{E}\left[\gamma_e^*|\mathbf{D}_{e-1}\right]\right\} = \text{E}\left[\gamma_e\right] = \gamma_e,$$

og beviset er fullført.
Vi har nå bevist at Mack sine påstander om at estimatorene δ_e^* og γ_e^* er forventningsrette, samt at δ_e^* og $\delta_{e'}^*$ er ukorrelerte for $e \neq e'$, er korrekte.
Nå fortsetter vi med utregningen av MSE for estimatoren $\bar{X}_j$, likning 3.2, under Mack sine tre antakelser. Likningen så slik ut:

$$\text{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)^2|\mathbf{D}\right\} = \text{Var}\left\{\tilde{X}_j|\mathbf{D}\right\} + \left(\text{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right)^2$$

Vi begynner med å studere det første leddet:

$$\text{Var}\left(\tilde{X}_j|\mathbf{D}\right) = \text{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) \tag{3.10}$$

Denne likheten følger av definisjonen på $\tilde{X}_j$, samt antakelse 3; det spiller ingen rolle om det betinges på hele det kjente området $\mathbf{D}$, eller om vi kun betinger på det området som har betydning for det aktuelle skadeåret, $\mathbf{D}_{j,J-j}$, i og med at vi har antatt uavhengighet mellom skadeårene. Videre får vi, ved bruk av regelen om betinget varians, at

$$\begin{aligned} \mathrm{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) &= \mathrm{E}\left\{\mathrm{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,D-1}\right)|\mathbf{D}_{j,J-j}\right\} \\ &+ \mathrm{Var}\left\{\mathrm{E}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,D-1}\right)|\mathbf{D}_{j,J-j}\right\} \end{aligned} \tag{3.11}$$

Ved bruk av antakelse 1 og 2 ser vi at

$$\mathrm{E}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,D-1}\right) = \delta_D\tilde{X}_{j,D-1}, \text{ og } \mathrm{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,D-1}\right) = \gamma_D\tilde{X}_{j,D-1}$$

Dersom dette settes inn i i likning 3.11, kan denne uttrykkes som

$$\mathrm{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) = \mathrm{E}\left\{\gamma_D\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right\} + \mathrm{Var}\left\{\delta_D\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right\}$$

Siden γ_D og δ_D er konstanter, kan disse bli tatt utenfor henholdsvis forventningsuttrykket og variansuttrykket, og vi har da

$$\mathrm{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) = \gamma_D\mathrm{E}\left\{\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right\} + \delta_D^2\mathrm{Var}\left\{\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right\} \tag{3.12}$$

Før vi kommer videre må vi gjøre noen små utregninger. Begynner med første del på høyre side av likning 3.12, og prøver å finne et generelt uttrykk for denne.

Ser på antakelse 1 hos Mack. Denne sier at $\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,e-1}\right) = \delta_e\tilde{X}_{j,e-1}$ for $e = 1,\cdots,D$. Dette forteller oss at forventet kumulativ status ved slutten av utviklingsår e, kan finnes ved å betinge på de tidligere observerte data, $\mathbf{D}_{j,e-1}$. Dersom vi står ved tid J og ser på skadeår j kan vi betinge på det som er kjent til nå, altså $\mathbf{D}_{j,J-j}$. Ser på uttrykket $\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,J-j}\right)$ for $e > J-j$. Setter inn for enkelte verdier av e, og ser om vi finner et mønster. Begynner med $e = J-j+1$:

Ved bruk av antakelsen om forventning kan vi da skrive

$$\mathrm{E}\left(\tilde{X}_{j,J-j+1}|\mathbf{D}_{j,J-j}\right) = \delta_{J-j+1}\tilde{X}_{j,J-j}$$

Dersom vi lar $e = J-j+2$ får vi at

$$\begin{aligned} \mathrm{E}\left(\tilde{X}_{j,J-j+2}|\mathbf{D}_{j,J-j}\right) &= \mathrm{E}\left\{\mathrm{E}\left(\tilde{X}_{j,J-j+2}|\tilde{X}_{j,J-j+1},\mathbf{D}_{j,J-j}\right)|\mathbf{D}_{j,J-j}\right\} \\ &= \mathrm{E}\left\{\mathrm{E}\left(\tilde{X}_{j,J-j+2}|\mathbf{D}_{j,J-j+1}\right)|\mathbf{D}_{j,J-j}\right\} \\ &= \mathrm{E}\left(\delta_{J-j+2}\tilde{X}_{j,J-j+1}|\mathbf{D}_{j,J-j}\right) \\ &= \delta_{J-j+2}\mathrm{E}\left(\tilde{X}_{j,J-j+1}|\mathbf{D}_{j,J-j}\right) \\ &= \delta_{J-j+2}\cdot\delta_{J-j+1}\tilde{X}_{j,J-j} \end{aligned}$$

Også her har vi brukt Mack sin første antakelse, samt regelen om repetert betinget forventning. Vi begynner nå å ane et mønster, men ser også på tilfellet når $e = J - j + 3$ for å se om det stemmer også da:

$$\begin{aligned} \mathrm{E}\left(\tilde{X}_{j,J-j+3}|\mathbf{D}_{j,J-j}\right) &= \mathrm{E}\left\{\mathrm{E}\left(\tilde{X}_{j,J-j+3}|\mathbf{D}_{j,J-j+2}\right)|\mathbf{D}_{j,J-j}\right\} \\ &= \delta_{J-j+3}\mathrm{E}\left(\tilde{X}_{j,J-j+2}|\mathbf{D}_{j,J-j}\right) \\ &= \delta_{J-j+3}\cdot\delta_{J-j+2}\cdot\delta_{J-j+1}\tilde{X}_{j,J-j} \end{aligned}$$

Vi ser nå et tydelig mønster, og konkluderer med at

$$\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,J-j}\right) = \tilde{X}_{j,J-j}\prod_{i=J-j+1}^{e}\delta_i \;\text{ for } e = J-j+1,\cdots,D \tag{3.13}$$

Dette forteller oss at dersom man ønsker å finne forventede akkumulerte totale krav for skadeår j opp til utviklingsår e, der $e > J - j$, så tar man siste kjente verdi, og multipliserer med utviklingsfaktorene for de fremtidige årene. Ved å bruke likning 3.13 med $e = D - 1$, kan vi nå skrive likning 3.12 som

$$\begin{aligned} \mathrm{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) = \gamma_D\tilde{X}_{j,J-j}&\prod_{e=J-j+1}^{D-1}\delta_e \\ &+\delta_D^2\mathrm{Var}\left(\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right) \end{aligned} \tag{3.14}$$

For å finne et uttrykk for variansen i det andre leddet på høyre side, $\mathrm{Var}\left(\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right)$, bruker vi likning 3.14, men setter inn $D - 1$ istedenfor D:

$$\begin{aligned} \mathrm{Var}\left(\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right) = \gamma_{D-1}\tilde{X}_{j,J-j}&\prod_{e=j-j+1}^{D-2}\delta_e \\ &+\delta_{D-1}^2\mathrm{Var}\left(\tilde{X}_{j,D-2}|\mathbf{D}_{j,J-j}\right) \end{aligned} \tag{3.15}$$

Ønsker videre å finne et uttrykk for $\mathrm{Var}\left(\tilde{X}_{j,D-2}|\mathbf{D}_{j,J-j}\right)$. Bruker da likning 3.15, men setter inn $D - 1$ istedenfor D. Dette gir følgende uttrykk:

$$\begin{aligned} \mathrm{Var}\left(\tilde{X}_{j,D-2}|\mathbf{D}_{j,J-j}\right) = \gamma_{D-2}\tilde{X}_{j,J-j}&\prod_{e=j-j+1}^{D-3}\delta_e \\ &+\delta_{D-2}^2\mathrm{Var}\left(\tilde{X}_{j,D-3}|\mathbf{D}_{j,J-j}\right) \end{aligned}$$

Det ser ut som om vi har kommet frem til et mønster som vil fortsette helt til vi står igjen med $\mathrm{Var}\left(\tilde{X}_{j,J-j+1}|\mathbf{D}_{j,J-j}\right)$. Denne kan vi imidlertid finne et uttrykk for ved hjelp av antakelse 2:

$$\mathrm{Var}\left(\tilde{X}_{j,J-j+1}|\mathbf{D}_{j,J-j}\right) = \gamma_{J-j+1}\tilde{X}_{j,J-j}$$

Dersom vi nå går tilbake til likning 3.14 og setter inn uttrykkene vi har funnet, får vi følgende:

$$
\begin{aligned}
\text{Var}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) &= \gamma_D \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-1} \delta_e \\
&+ \delta_D^2 \text{Var}\left(\tilde{X}_{j,D-1}|\mathbf{D}_{j,J-j}\right) \\
&= \gamma_D \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-1} \delta_e \\
&+ \delta_D^2 \cdot \gamma_{D-1} \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-2} \delta_e \\
&+ (\delta_D)^2 (\delta_{D-1})^2 \text{Var}\left(\tilde{X}_{j,D-2}|\mathbf{D}_{j,J-j}\right) \\
&= \gamma_D \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-1} \delta_e \\
&+ \delta_D^2 \cdot \gamma_{D-1} \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-2} \delta_e \\
&+ (\delta_D)^2 (\delta_{D-1})^2 \gamma_{D-2} \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-3} \delta_e \\
&+ (\delta_D)^2 (\delta_{D-1})^2 (\delta_{D-2})^2 \text{Var}\left(\tilde{X}_{j,D-3}|\mathbf{D}_{j,J-j}\right)
\end{aligned}
$$

Dette uttrykket vil fortsette å utvide seg helt til vi kommer frem til det siste leddet, som vil se slik ut:

$$
\begin{aligned}
&(\delta_D)^2 (\delta_{D-1})^2 (\delta_{D-2})^2 \cdots (\delta_{J-j+2})^2 \text{Var}\left(\tilde{X}_{j,J-j+1}|\mathbf{D}_{j,J-j}\right) \\
&= (\delta_D)^2 (\delta_{D-1})^2 (\delta_{D-2})^2 \cdots (\delta_{J-j+2})^2 \cdot \gamma_{J-j+1} \tilde{X}_{j,J-j}
\end{aligned}
$$

Det endelige uttrykket for Var $\left(\tilde{X}_j|\mathbf{D}\right)$, likning 3.10, blir da:

$$\begin{aligned}
\operatorname{Var}\left(\tilde{X}_j|\mathbf{D}\right) &= \gamma_D \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-1} \delta_e \\
&+ (\delta_D)^2 \cdot \gamma_{D-1} \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-2} \delta_e \\
&+ (\delta_D)^2 (\delta_{D-1})^2 \gamma_{D-2} \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{D-3} \delta_e \\
&+ \cdots \\
&+ (\delta_D)^2 (\delta_{D-1})^2 (\delta_{D-2})^2 \cdots (\delta_{J-j+3})^2 \cdot \gamma_{J-j+2} \tilde{X}_{j,J-j} \prod_{e=J-j+1}^{J-j+1} \delta_e \\
&+ (\delta_D)^2 (\delta_{D-1})^2 (\delta_{D-2})^2 \cdots (\delta_{J-j+2})^2 \cdot \gamma_{J-j+1} \tilde{X}_{j,J-j}
\end{aligned} \tag{3.16}$$

Vi har nå funnet et uttrykk for det første leddet i likning 3.2. Som vi ser er ikke dette uttrykket på samme komprimerte form som det Neuhaus oppgir i sin likning 6.15. Vi vil derfor studere hans uttrykk litt grundigere, for å sjekke at de er like. Uttrykket Neuhaus oppgir er:

$$
\begin{aligned}
\operatorname{Var}\left(\tilde{X}_j|\mathbf{D}\right) &= \tilde{X}_{j,J-j} \sum_{e=J-j+1}^{D} \left(\prod_{e'=J-j+1}^{e-1} \delta_{e'} \right) \gamma_e \left(\prod_{e'=e+1}^{D} \delta_{e'}^2 \right) \\
&= \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{J-j} \delta_{e'} \right) \gamma_{J-j+1} \left(\prod_{e'=J-j+2}^{D} \delta_{e'}^2 \right) \\
&+ \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{J-j+1} \delta_{e'} \right) \gamma_{J-j+2} \left(\prod_{e'=J-j+3}^{D} \delta_{e'}^2 \right) \\
&+ \cdots \\
&+ \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{D-2} \delta_{e'} \right) \gamma_{D-1} \left(\prod_{e'=D}^{D} \delta_{e'}^2 \right) \\
&+ \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{D-1} \delta_{e'} \right) \gamma_{D} \left(\prod_{e'=D+1}^{D} \delta_{e'}^2 \right) \\
&= \tilde{X}_{j,J-j} \gamma_{J-j+1} \cdot (\delta_{J-j+2})^2 \cdots (\delta_D)^2 \\
&+ \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{J-j+1} \delta_{e'} \right) \gamma_{J-j+2} \cdot (\delta_{J-j+3})^2 \cdots (\delta_D)^2 \\
&+ \cdots \\
&+ \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{D-2} \delta_{e'} \right) \gamma_{D-1} \cdot (\delta_D)^2 \\
&+ \tilde{X}_{j,J-j} \left(\prod_{e'=J-j+1}^{D-1} \delta_{e'} \right) \gamma_{D}
\end{aligned} \tag{3.17}
$$

Som vi ser er dette uttrykket identisk med det som vi kom frem til (likning 3.16). I de videre utregningene vil vi bruke Neuhaus sitt mer komprimerte uttrykk.

Uttrykket inneholder ukjente størrelser. Disse kan vi imidlertid estimere ved å substituere de ukjente parametrene med sine empiriske størrelser. Likning 3.17 kan da estimeres ved

$$
\operatorname{Var}^*\left(\tilde{X}_j|\mathbf{D}\right) = \tilde{X}_{j,J-j} \sum_{e=J-j+1}^{D} \left(\prod_{e'=J-j+1}^{e-1} \delta^*_{e'} \right) \gamma^*_e \left(\prod_{e'=e+1}^{D} (\delta^*_{e'})^2 \right) \tag{3.18}
$$

Dette uttrykket ønsker vi å forenkle litt. Ut fra likning 2.2 ser vi at dersom man setter $D = e - 1$ får man

$$
\bar{X}_{j,e-1} = \tilde{X}_{j,J-j} \prod_{e'=J-j+1}^{e-1} \delta^*_{e'}
$$

Dersom man omskriver dette uttrykket slikt at produkt-leddet står alene på venstre side, ser vi at det første produkt-leddet i likning 3.18 kan omskrives til

$$\prod_{e'=J-j+1}^{e-1} \delta_{e'}^{*} = \frac{\bar{X}_{j,e-1}}{\tilde{X}_{j,J-j}}$$

Det andre produkt-leddet i likning 3.18 kan også omformuleres ved hjelp av likning 2.2. Dette gjøres ved å sette $J - j = e - 1$:

$$\begin{aligned}\bar{X}_j &= \tilde{X}_{j,e-1} \prod_{e'=e}^{D} \delta_{e'}^{*} \\ &= \tilde{X}_{j,e-1}\delta_e^{*} \prod_{e'=e+1}^{D} \delta_{e'}^{*}\end{aligned}$$

Dersom vi også her omformulerer likningen ovenfor slik at vi får produkt-leddet alene på venstre side ser vi at

$$\prod_{e'=e+1}^{D} \delta_{e'}^{*} = \frac{\bar{X}_j}{\tilde{X}_{j,e-1}\delta_e^{*}}$$

Vi ser fra likning 3.18 at $\delta_{e'}^{*}$ skal være opphøyet i andre potens, og siden det er et produkt vi ser på kan dette gjøres uten problemer. Likning 3.18 kan da omskrives slik at

$$\mathrm{Var}^{*}\left(\tilde{X}_j|\mathbf{D}\right) = \sum_{e=J-j+1}^{D} \tilde{X}_{j,e-1}\gamma_e^{*}\left(\frac{\bar{X}_j^2}{\tilde{X}_{j,e-1}^2}\right)\frac{1}{\delta_e^{*2}} \tag{3.19}$$

Som vi ser inneholder summen faktoren $\tilde{X}_{j,e-1}$. Denne vil være ukjent for $e > J - j$, siden vi ikke har observasjoner lenger enn til utviklingsår $J - j$. Dette problemet løses ved å erstatte $\tilde{X}_{j,e-1}$ med den estimerte størrelsen $\bar{X}_{j,e-1}$, som vil gi mening for alle e vi summerer over i 3.19.

$$\begin{aligned}\mathrm{Var}^{*}\left(\tilde{X}_j|\mathbf{D}\right) &= \sum_{e=J-j+1}^{D} \bar{X}_{j,e-1}\gamma_e^{*}\left(\frac{\bar{X}_j^2}{\bar{X}_{j,e-1}^2}\right)\frac{1}{\delta_e^{*2}} \\ &= \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma_e^{*}}{\delta_e^{*2}}\left(\frac{1}{\bar{X}_{j,e-1}}\right)\end{aligned} \tag{3.20}$$

Vi har nå kommet frem til en tilnærming for det første leddet i likningen for MSE, likning 3.2. Dersom vi setter dette inn, vil likningen se slik ut:

$$\begin{aligned}\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)^2|\mathbf{D}\right\} &\approx \mathrm{Var}^{*}\left\{\tilde{X}_j|\mathbf{D}\right\} + \left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right\}^2 \\ &= \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma_e^{*}}{\delta_e^{*2}}\left(\frac{1}{\bar{X}_{j,e-1}}\right) + \left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right\}^2\end{aligned} \tag{3.21}$$

Nå ønsker vi å gå videre med å studere det andre leddet på høyre side litt nærmere. Ved bruk av antakelse 3 kan leddet $\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)$ omskrives til:

$$\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) = \mathrm{E}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right)$$

Dersom vi ser tilbake til likning 3.13, ser vi imidlertid at vi allerede har funnet et uttrykk for denne. Vi fant da at

$$\mathrm{E}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,J-j}\right) = \tilde{X}_{j,J-j}\prod_{i=J-j+1}^{e}\delta_i \quad \text{for} \quad e = J-j+1,\cdots,D$$

Dersom vi bruker denne formelen på vårt uttrykk, ved å la $e = D$, får vi at

$$\mathrm{E}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right) = \tilde{X}_{j,J-j}\prod_{e=J-j+1}^{D}\delta_e$$

Tidligere, se likning 2.2, har vi også sett at $\bar{X}_j = \tilde{X}_{j,J-j}\prod_{e=J-j+1}^{D}\delta_e^*$. Dersom vi setter dette, samt uttrykket for $\mathrm{E}\left(\tilde{X}_{j,D}|\mathbf{D}_{j,J-j}\right)$, inn i det andre leddet i likning 3.21, får vi at:

$$\begin{aligned}\left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right\}^2 &= \left\{\tilde{X}_{j,J-j}\prod_{e=J-j+1}^{D}\delta_e - \tilde{X}_{j,J-j}\prod_{e=J-j+1}^{D}\delta_e^*\right\}^2 \\ &= \tilde{X}_{j,J-j}^2\left\{\prod_{e=J-j+1}^{D}\delta_e - \prod_{e=J-j+1}^{D}\delta_e^*\right\}^2 \\ &= \tilde{X}_{j,J-j}^2\left\{\sum_{e=J-j+1}^{D}\left(\prod_{e'=J-j+1}^{e-1}\delta_{e'}^*\right)\left(\delta_e-\delta_e^*\right)\left(\prod_{e'=e+1}^{D}\delta_{e'}\right)\right\}^2\end{aligned} \tag{3.22}$$

For å tydeliggjøre den siste overgangen, kan vi se på følgende eksempel:

$$\begin{aligned}\prod_{i=1}^{n}a_i - \prod_{i=1}^{n}b_i &= \prod_{i=1}^{n}a_i - b_1\prod_{i=2}^{n}a_i + b_1\prod_{i=2}^{n}a_i - \prod_{i=1}^{n}b_i \\ &= (a_1-b_1)\prod_{i=2}^{n}a_i + b_1\left(\prod_{i=2}^{n}a_i - \prod_{i=2}^{n}b_i\right) \\ &= (a_1-b_1)\prod_{i=2}^{n}a_i + b_1\left[(a_2-b_2)\prod_{i=3}^{n}a_i + b_2\left(\prod_{i=3}^{n}a_i - \prod_{i=3}^{n}b_i\right)\right] \\ &= ... \\ &= \sum_{i=1}^{n}(a_i-b_i)\left(\prod_{j=i+1}^{n}a_j\right)\left(\prod_{j=1}^{i-1}b_j\right)\end{aligned}$$

Dersom vi nå lar indeksen gå fra $e = J - j + 1$ til D, samt setter $a = \delta$ og $b = \delta^*$, så ser vi at

$$\prod_{e=J-j+1}^{D} \delta_e - \prod_{e=J-j+1}^{D} \delta_e^* = \sum_{e=J-j+1}^{D} \left(\prod_{e'=e+1}^{D} \delta_{e'} \right) (\delta_e - \delta_e^*) \left(\prod_{e'=J-j+1}^{e-1} \delta_{e'}^* \right) = \sum_{e=J-j+1}^{D} S_e,$$

og dermed er den siste overgangen i likning 3.22 vist. Denne likningen kan nå skrives som:

$$\left\{ \mathrm{E}\left(\tilde{X}_j | \mathbf{D} \right) - \bar{X}_j \right\}^2 = \tilde{X}_{j,J-j}^2 \left\{ \sum_{e=J-j+1}^{D} S_e \right\}^2$$

Også dette uttrykket inneholder ukjente størrelser. Vi ser imidlertid at dersom vi benytter de empiriske størrelsene her vil vi få null. Dette er nok litt optimistisk, så Mack foreslår derfor å skrive ut uttrykket til

$$\left\{ \mathrm{E}\left(\tilde{X}_j | \mathbf{D} \right) - \bar{X}_j \right\}^2 = \tilde{X}_{j,J-j}^2 \left\{ \sum_{e=J-j+1}^{D} S_e^2 + 2 \sum_{e=J-j+1}^{D} \sum_{e'=e+1}^{D} S_e S_{e'} \right\} \tag{3.23}$$

Han gjør videre tilnærmingene $S_e^2 \approx \mathrm{E}\left(S_e^2 | \mathbf{D}_{e-1} \right)$ og $S_e S_{e'} \approx \mathrm{E}\left(S_e S_{e'} | \mathbf{D}_{e'-1} \right)$, der vi antar $e < e'$. La oss nå se litt nærmere på den første tilnærmingen, til S_e^2:

$$\mathrm{E}\left(S_e^2 | \mathbf{D}_{e-1} \right) = \mathrm{E} \left\{ \left(\prod_{e'=J-j+1}^{e-1} \delta_{e'}^* \right) (\delta_e - \delta_e^*) \left(\prod_{e'=e+1}^{D} \delta_{e'} \right) | \mathbf{D}_{e-1} \right\}^2 \tag{3.24}$$

Ut fra antakelse 1 er δ_e en konstant, og kan sees på som kjent. Det andre produkt-leddet kan da bli tatt utenfor forventningen. I det første produkt-leddet har vi δ_e^*. Om denne er kjent kommer an på hva det betinges på. Generelt har vi (se likning 2.1) at

$$\delta_e^* = \frac{\sum_{j=1}^{J-e} \tilde{X}_{j,e}}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}}$$

Den største e' verdien i vårt uttrykk er $e - 1$. Dersom vi putter dette inn i likningen får vi at

$$\delta_{e-1}^* = \frac{\sum_{j=1}^{J-(e-1)} \tilde{X}_{j,e-1}}{\sum_{j=1}^{J-(e-1)} \tilde{X}_{j,e-2}}$$

Siden det betinges på $\mathbf{D}_{e-1}$ antar vi at alt opp til utviklingsår $e - 1$ er kjent, uavhengig av skadeår. Da vil både $\tilde{X}_{j,e-1}$ og $\tilde{X}_{j,e-2}$ være kjent for $j = 1, 2, \cdots, J - (e-1)$, og dermed vil også begge summene være kjent. Vi kan dermed konkludere med at δ_{e-1}^* er kjent, noe som medfører at $\delta_{e'}^*$ også er kjent for $J - j + 1 < e' < e - 1$. Også det første produkt-leddet kan da bli tatt utenfor forventningen. Den eneste stokastiske delen av likning 3.24 er da $(\delta_e - \delta_e^*)$, som vi nå vil se litt nærmere på forventningen til:

$$\begin{aligned}
\mathrm{E}\left\{(\delta_e - \delta_e^*)^2|\mathbf{D}_{e-1}\right\} &= \mathrm{E}\left\{(\delta_e^* - \delta_e)^2|\mathbf{D}_{e-1}\right\} \\
&= \mathrm{E}\left\{(\delta_e^* - \mathrm{E}(\delta_e^*))^2|\mathbf{D}_{e-1}\right\} \\
&= \mathrm{Var}\left(\delta_e^*|\mathbf{D}_{e-1}\right) \\
&= \mathrm{Var}\left\{\frac{\sum_{j=1}^{J-e}\tilde{X}_{j,e}}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}|\mathbf{D}_{e-1}\right\}
\end{aligned}$$

Her har vi brukt det faktum at δ_e^* er en forventningsrett estimator for δ_e, samt den generelle formelen for varians. Når vi i siste linje her betinger på $\mathbf{D}_{e-1}$ vil nevneren være kjent, og kan sees på som en konstant. Vi trekker derfor denne utenfor variansen, og får at

$$\mathrm{E}\left\{(\delta_e - \delta_e^*)^2|\mathbf{D}_{e-1}\right\} = \frac{1}{\left(\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\right)^2}\mathrm{Var}\left(\sum_{j=1}^{J-e}\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right)$$

Ved antakelse 3, som sier at vi har uavhengighet mellom $\tilde{X}_{j,e}$ og $\tilde{X}_{k,e}$ for $j \neq k$, så vil variansen til summen være lik summen av variansene, og vi får at

$$\mathrm{E}\left\{(\delta_e - \delta_e^*)^2|\mathbf{D}_{e-1}\right\} = \frac{1}{\left(\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\right)^2}\sum_{j=1}^{J-e}\mathrm{Var}\left(\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right) \tag{3.25}$$

Siden vi er interessert i $\mathrm{Var}\left(\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right)$ trenger vi kun ta hensyn til dataene for skadeår j, og det vil derfor gi like mye informasjon å betinge på $\mathbf{D}_{j,e-1}$ som å betinge på $\mathbf{D}_{e-1}$. Fortsetter utregningen av likning 3.25, og bruker også antakelse 2 igjen:

$$\begin{aligned}
\mathrm{E}\left\{(\delta_e - \delta_e^*)^2|\mathbf{D}_{e-1}\right\} &= \frac{1}{\left(\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\right)^2}\sum_{j=1}^{J-e}\mathrm{Var}\left(\tilde{X}_{j,e}|\mathbf{D}_{j,e-1}\right) \\
&= \frac{1}{\left(\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}\right)^2}\sum_{j=1}^{J-e}\gamma_e\tilde{X}_{j,e-1} \qquad (3.26)\\
&= \frac{\gamma_e}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}
\end{aligned}$$

Vi går nå tilbake til likning 3.24, og setter inn det vi har funnet.
Likningen ser da slik ut:

$$
\begin{aligned}
\mathrm{E}\left(S_e^2|\mathbf{D}_{e-1}\right) &= \left(\prod_{e'=J-j+1}^{e-1} \delta_{e'}^*\right)^2 \left(\prod_{e'=e+1}^{D} \delta_{e'}\right)^2 \mathrm{E}\left\{(\delta_e - \delta_e^*)^2|\mathbf{D}_{e-1}\right\} \\
&= \left(\prod_{e'=J-j+1}^{e-1} \delta_{e'}^*\right)^2 \left(\prod_{e'=e+1}^{D} \delta_{e'}\right)^2 \frac{\gamma_e}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}}
\end{aligned}
$$

Denne likningen tilsvarer Neuhaus sin likning 6.19, men vi ser her at vår øvre indeks på summen er ulik Neuhaus sin, da vår går til $J-e$ mens hans går til e. Neuhaus sin artikkel er imidlertid basert på en artikkel av Mack [Mack 93], og i denne artikkelen er det brukt samme indeks som vi har. I denne oppgaven vil vi derfor fortsette å bruke $J-e$ som øvre indeks.

Videre må vi se på $\mathrm{E}\left(S_e S_{e'}|\mathbf{D}_{e'-1}\right)$. Antar at $e < e'$, noe som gir at uttrykket for $S_e S_{e'}$ da vil være:

$$
S_e S_{e'} = \left(\prod_{i=J-j+1}^{e-1} \delta_i^*\right)(\delta_e - \delta_e^*)\left(\prod_{i=e+1}^{D} \delta_i\right) \cdot \left(\prod_{i=J-j+1}^{e'-1} \delta_i^*\right)(\delta_{e'} - \delta_{e'}^*)\left(\prod_{i=e'+1}^{D} \delta_i\right)
$$

Ved samme argumentasjon som vi nettopp gjorde, kan vi se at siden vi betinger på $D_{e'-1}$ er det kun $(\delta_e - \delta_e^*)(\delta_{e'} - \delta_{e'}^*)$ som er stokastisk, og som det vil være nødvendig å ta forventning av.

$$
\mathrm{E}\left\{(\delta_e - \delta_e^*)\left(\delta_{e'} - \delta_{e'}^*\right)|\mathbf{D}_{e'-1}\right\}
$$

Siden vi her betinger på $\mathbf{D}_{e'-1}$, og vi samtidig har antatt at $e < e' \Leftrightarrow e \leq e'-1$ vil δ_e^* være kjent for alle aktuelle e, og kan dermed bli tatt utenfor forventningen. Det kan også δ_e, som ved antakelse 1 er antatt å være en konstant.

$$
\begin{aligned}
\mathrm{E}\left\{(\delta_e - \delta_e^*)\left(\delta_{e'} - \delta_{e'}^*\right)|\mathbf{D}_{e'-1}\right\} &= (\delta_e - \delta_e^*)\,\mathrm{E}\left\{\left(\delta_{e'} - \delta_{e'}^*\right)|\mathbf{D}_{e'-1}\right\} \\
&= (\delta_e - \delta_e^*)\left\{\mathrm{E}\left(\delta_{e'}|\mathbf{D}_{e'-1}\right) - \mathrm{E}\left(\delta_{e'}^*|\mathbf{D}_{e'-1}\right)\right\} \qquad (3.27) \\
&= (\delta_e - \delta_e^*)\left\{\delta_{e'} - \mathrm{E}\left(\delta_{e'}^*|\mathbf{D}_{e'-1}\right)\right\}
\end{aligned}
$$

Vi står nå igjen med kun en betinget forventning. Ser litt nærmere på denne, og i den følgende utregningen blir både antakelse 1 og antakelse 3 fra Mack brukt.

$$\begin{aligned}
\mathrm{E}\left(\delta_{e'}^{*}|\mathbf{D}_{e'-1}\right) &= \mathrm{E}\left\{\frac{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'}}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}|\mathbf{D}_{e'-1}\right\} \\
&= \frac{1}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}\mathrm{E}\left\{\sum_{j=1}^{J-e'}\left(\tilde{X}_{j,e'}|\mathbf{D}_{e'-1}\right)\right\} \\
&= \frac{1}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}\left\{\sum_{j=1}^{J-e'}\mathrm{E}\left(\tilde{X}_{j,e'}|\mathbf{D}_{j,e'-1}\right)\right\} \\
&= \frac{1}{\sum_{j=1}^{J-e'}\tilde{X}_{j,e'-1}}\left\{\sum_{j=1}^{J-e'}\delta_{e'}\tilde{X}_{j,e'-1}\right\} \\
&= \delta_{e'}
\end{aligned}$$

Dersom vi nå setter dette tilbake i likning 3.27 får vi:

$$\begin{aligned}
\mathrm{E}\left\{\left(\delta_e-\delta_e^{*}\right)\left(\delta_{e'}-\delta_{e'}^{*}\right)|\mathbf{D}_{e'-1}\right\} &= \left(\delta_e-\delta_e^{*}\right)\left\{\delta_{e'}-\mathrm{E}\left(\delta_{e'}^{*}|\mathbf{D}_{e'-1}\right)\right\} \\
&= \left(\delta_e-\delta_e^{*}\right)\left(\delta_{e'}-\delta_{e'}\right) \\
&= 0
\end{aligned}$$

Dette forteller oss at $\mathrm{E}\left(S_eS_{e'}|\mathbf{D}_{e'-1}\right)$ er null for $e < e'$, og siden samme prosedyre kan utføres dersom $e > e'$ (ved å bytte navn på variablene) vil forventningen også bli null i dette tilfellet.

Går nå tilbake til likning 3.23, og setter inn det vi har funnet:

$$\begin{aligned}
&\left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right\}^2 \\
&\approx \tilde{X}_{j,J-j}^2\left\{\sum_{e=J-j+1}^{D}\mathrm{E}\left(S_e^2|\mathbf{D}_{e-1}\right)+2\sum_{e=J-j+1}^{D}\sum_{e'=e+1}^{D}\mathrm{E}\left(S_eS_{e'}|\mathbf{D}_{e'-1}\right)\right\} \\
&= \tilde{X}_{j,J-j}^2\sum_{e=j-j+1}^{D}\left(\prod_{e'=J-j+1}^{e-1}\delta_{e'}^{*}\right)^2\left(\prod_{e'=e+1}^{D}\delta_{e'}\right)^2\frac{\gamma_e}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}
\end{aligned}$$

Dersom vi nå erstatter de ukjente størrelsene med de empiriske, vil vi få

$$\begin{aligned}
\left\{\left[\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right]^2\right\}^{*} &= \tilde{X}_{j,J-j}^2\sum_{e=j-j+1}^{D}\left(\prod_{e'=J-j+1}^{e-1}\delta_{e'}^{*}\right)^2\left(\prod_{e'=e+1}^{D}\delta_{e'}^{*}\right)^2\frac{\gamma_e^{*}}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}} \\
&= \tilde{X}_{j,J-j}^2\left(\prod_{e'=J-j+1}^{D}\delta_{e'}^{*}\right)^2\sum_{e=j-j+1}^{D}\frac{1}{(\delta_e^{*})^2}\frac{\gamma_e^{*}}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}
\end{aligned} \tag{3.28}$$

Vi ønsker ikke å ha noe produkt-ledd i likningen vår. Ser at dersom vi omformulerer likning 2.2 kan produktet skrives som

$$\left(\prod_{e'=J-j+1}^{D} \delta_{e'}^{*} \right)^2 = \frac{\bar{X}_j^2}{\tilde{X}_{j,J-j}^2}$$

Setter dette inn og får:

$$\left\{ \left[\mathrm{E}\left(\tilde{X}_j | \mathbf{D} \right) - \bar{X}_j \right]^2 \right\}^* = \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma_e^*}{(\delta_e^*)^2} \frac{1}{\sum_{q=1}^{J-e} \tilde{X}_{q,e-1}} \tag{3.29}$$

Vi har nå funnet et uttrykk for andre ledd i likning 3.2, som så slik ut:

$$\mathrm{E}\left\{ \left(\tilde{X}_j - \bar{X}_j \right)^2 | \mathbf{D} \right\} = \mathrm{Var}\left(\tilde{X}_j | \mathbf{D} \right) + \left\{ \mathrm{E}\left(\tilde{X}_j | \mathbf{D} \right) - \bar{X}_j \right\}^2$$

Tidligere har vi funnet et uttrykk for det første leddet, og dersom vi nå setter inn dette, sammen med uttrykket for det andre leddet som vi nettopp har funnet, vil vi få at

$$\begin{aligned} \mathrm{E}\left\{ \left(\tilde{X}_j - \bar{X}_j \right)^2 | \mathbf{D} \right\} &\approx \mathrm{E}^* \left\{ \left(\tilde{X}_j - \bar{X}_j \right)^2 | \mathbf{D} \right\} \\ &= \mathrm{Var}^* \left(\tilde{X}_j | \mathbf{D} \right) + \left\{ \left[\mathrm{E}\left(\tilde{X}_j | \mathbf{D} \right) - \bar{X}_j \right]^2 \right\}^* \\ &= 3.20 + 3.29 \\ &= \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma_e^*}{(\delta_e^*)^2} \left\{ \frac{1}{\bar{X}_{j,e-1}} + \frac{1}{\sum_{q=1}^{J-e} \tilde{X}_{q,e-1}} \right\} \end{aligned} \tag{3.30}$$

Denne likningen gir oss et estimat for det forventet kvadratisk avvik for estimatoren $\bar{X}_j$.

Vi er også interessert i et estimat av det forventede kvadratiske avviket til summen over alle skadeårene $j = 1, \cdots, J$, altså $\sum_{j=1}^{J} \tilde{X}_j$. En naturlig estimator for denne vil være $\sum_{j=1}^{J} \bar{X}_j$. Det er imidlertid ingen grunn til å ta med det første skadeåret $j = 1$ i dette estimatet, da vi tidligere har antatt at all utvikling er kjent for dette året. Vi vil derfor la summen her gå fra $j = 2$, noe som også stemmer overens med det Mack gjør i sin artikkel [Mack 93]. Forventet kvadratisk avvik gitt de observerte data blir da

$$\mathrm{E}\left\{ \left(\sum_{j=2}^{J} \tilde{X}_j - \sum_{j=2}^{J} \bar{X}_j \right)^2 | \mathbf{D} \right\} = \mathrm{E}\left\{ \left(\sum_{j=2}^{J} (\tilde{X}_j - \bar{X}_j) \right)^2 | \mathbf{D} \right\} \tag{3.31}$$

Vi vil nå gå igjennom en tilsvarende utregning som vi gjorde da vi så på det enkle skadeåret j. Målet er å finne et enklere uttrykk for den totale MSE, og vi begynner da med å se nærmere på summen i 3.31:

$$\begin{aligned}
\left(\sum_{j=2}^{J}(\tilde{X}_j - \bar{X}_j)\right)^2 &= \left\{(\tilde{X}_2 - \bar{X}_2) + \cdots + (\tilde{X}_J - \bar{X}_J)\right\}\left\{(\tilde{X}_2 - \bar{X}_2) + \cdots + (\tilde{X}_J - \bar{X}_J)\right\} \\
&= (\tilde{X}_2 - \bar{X}_2)^2 + (\tilde{X}_2 - \bar{X}_2)(\tilde{X}_3 - \bar{X}_3) + \cdots + (\tilde{X}_2 - \bar{X}_2)(\tilde{X}_J - \bar{X}_J) \\
&+ (\tilde{X}_3 - \bar{X}_3)(\tilde{X}_2 - \bar{X}_2) + (\tilde{X}_3 - \bar{X}_3)^2 + \cdots + (\tilde{X}_3 - \bar{X}_3)(\tilde{X}_J - \bar{X}_J) \\
&+ \cdots \\
&+ (\tilde{X}_J - \bar{X}_J)(\tilde{X}_2 - \bar{X}_2) + \cdots + (\tilde{X}_J - \bar{X}_J)^2 \\
&= \sum_{j=2}^{J}\left(\tilde{X}_j - \bar{X}_j\right)^2 + 2\sum_{j=2}^{J-1}\sum_{k=j+1}^{J}\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right)
\end{aligned}$$

Setter dette tilbake i likning 3.31 og får

$$\begin{aligned}
&\mathrm{E}\left\{\left(\sum_{j=2}^{J}(\tilde{X}_j - \bar{X}_j)\right)^2 |\mathbf{D}\right\} \\
&= \sum_{j=2}^{J}\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)^2 |\mathbf{D}\right\} + 2\sum_{j=2}^{J-1}\sum_{k=j+1}^{J}\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right)|\mathbf{D}\right\}
\end{aligned} \tag{3.32}$$

Dette uttrykket tilsvarer Neuhaus sin likning 6.22. Som vi ser er imidlertid øvre grense for den andre summen ulik, her går den til $J-1$, mens den hos Neuhaus går til J. Dersom man ser litt nøyere på denne dobbeltsummen kan man imidlertid se at det ikke er noen grunn til at summen skal gå til J, da den påfølgende summen da vil gå fra $J+1$ til J, noe som per definisjon vil gi 0. Vi vil derfor fortsette å bruke $J-1$ som øvre grense for summen. Den første forventningen på høyre side finner vi direkte fra likning 3.30. Den andre forventningen må vi imidlertid se litt nøyere på. I den følgende utregningen brukes antakelse 3 om at det er uavhengighet mellom ulike skadeår, og som vi ser ut fra summerings-indeksene har vi også at $j \neq k$ for alle j og alle k.

$$\begin{aligned}
\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right)|\mathbf{D}\right\} &= \mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)|\mathbf{D}\right\}\mathrm{E}\left\{\left(\tilde{X}_k - \bar{X}_k\right)|\mathbf{D}\right\} \\
&= \left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right\}\left\{\mathrm{E}\left(\tilde{X}_k|\mathbf{D}\right) - \bar{X}_k\right\}
\end{aligned} \tag{3.33}$$

Dersom vi går tilbake til likning 3.22, og tar kvadratroten av denne, får vi at

$$\left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right) - \bar{X}_j\right\} = \tilde{X}_{j,J-j}\left\{\sum_{e=J-j+1}^{D}\left(\prod_{e'=J-j+1}^{e-1}\delta^*_{e'}\right)\left(\delta_e - \delta^*_e\right)\left(\prod_{e'=e+1}^{D}\delta_{e'}\right)\right\}$$

Benytter dette på 3.33, og får

$$
\begin{aligned}
&\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right)|\mathbf{D}\right\} \\
&= \tilde{X}_{j,J-j}\tilde{X}_{k,J-k}\left\{\sum_{e=J-j+1}^{D}\left(\prod_{f=J-j+1}^{e-1}\delta_f^*\right)(\delta_e - \delta_e^*)\left(\prod_{f=e+1}^{D}\delta_f\right)\right\} \\
&\cdot\left\{\sum_{g=J-k+1}^{D}\left(\prod_{h=J-k+1}^{g-1}\delta_h^*\right)(\delta_g - \delta_g^*)\left(\prod_{h=g+1}^{D}\delta_h\right)\right\}
\end{aligned} \tag{3.34}
$$

I dette uttrykket har vi noen ukjente størrelser. Disse kan vi imidlertid ikke bare bytte ut med de tilhørende empiriske størrelsene, da dette vil gi null, noe som er et urealistisk resultat. Vi velger derfor å bruke den betingede forventningen som et estimat for det ukjente uttrykket. Som vi har sett tidligere i oppgaven er det kun i de tilfellene der $e = g$ vi vil få noe bidrag, de andre leddene vil bli 0. Siden vi har at $j < k$ vil e-summen gå over færrest ledd, og det er disse leddene som vil være felles med deler av g-summen. Vi kan dermed skrive likning 3.34 som

$$
\begin{aligned}
&\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right)|\mathbf{D}\right\} \\
&= \tilde{X}_{j,J-j}\tilde{X}_{k,J-k} \\
&\cdot \mathrm{E}\left\{\sum_{e=J-j+1}^{D}\left(\prod_{f=J-j+1}^{e-1}\delta_f^*\right)(\delta_e - \delta_e^*)\left(\prod_{f=e+1}^{D}\delta_f\right)\left(\prod_{h=J-k+1}^{e-1}\delta_h^*\right)(\delta_e - \delta_e^*)\left(\prod_{h=e+1}^{D}\delta_h\right)|\mathbf{D}_{e-1}\right\}
\end{aligned} \tag{3.35}
$$

Ved samme resonnement som tidligere kan vi konkludere med at det eneste stokastiske i denne likningen er $(\delta_e - \delta_e^*)^2$, og det er kun dette uttrykket vi trenger ta forventningen av. Denne utregningen har vi imidlertid gjort tidligere, og fra likning 3.26 har vi at

$$
\mathrm{E}\left\{(\delta_e - \delta_e^*)^2|\mathbf{D}_{e-1}\right\} = \frac{\gamma_e}{\sum_{j=1}^{J-e}\tilde{X}_{j,e-1}}
$$

Vi setter nå dette tilbake i likning 3.35, som da ser slik ut:

$$
\begin{aligned}
&\mathrm{E}\left\{\left(\tilde{X}_j - \bar{X}_j\right)\left(\tilde{X}_k - \bar{X}_k\right)|\mathbf{D}\right\} \\
&= \tilde{X}_{j,J-j}\tilde{X}_{k,J-k} \\
&\cdot\left\{\sum_{e=J-j+1}^{D}\left(\prod_{f=J-j+1}^{e-1}\delta_f^*\right)\left(\prod_{f=e+1}^{D}\delta_f\right)\left(\prod_{h=J-k+1}^{e-1}\delta_h^*\right)\left(\prod_{h=e+1}^{D}\delta_h\right)\frac{\gamma_e}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}\right\}
\end{aligned} \tag{3.36}
$$

Dersom vi nå erstatter de ukjente størrelsene med de empiriske, samt trekker sammen produkt-leddene, kan likning 3.36 skrives som

$$
\begin{aligned}
&\mathrm{E}^*\left\{\left(\tilde{X}_j-\bar{X}_j\right)\left(\tilde{X}_k-\bar{X}_k\right)|\mathbf{D}\right\} \\
&=\tilde{X}_{j,J-j}\tilde{X}_{k,J-k}\cdot\left\{\sum_{e=J-j+1}^{D}\left(\prod_{f=J-j+1}^{D}\delta_f^*\right)\frac{1}{\delta_e^*}\left(\prod_{h=J-k+1}^{D}\delta_h^*\right)\frac{1}{\delta_e^*}\frac{\gamma_e^*}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}\right\}
\end{aligned} \tag{3.37}
$$

Ved bruk av likning 2.2 ser vi at de to produkt leddene kan skrives om på følgende måte:

$$
\prod_{f=J-j+1}^{D}\delta_f^*=\frac{\bar{X}_j}{\tilde{X}_{j,J-j}}
$$

og

$$
\prod_{h=J-k+1}^{D}\delta_h^*=\frac{\bar{X}_k}{\tilde{X}_{k,J-k}}
$$

noe som fører til at likning 3.37 kan forenkles til

$$
\begin{aligned}
&\mathrm{E}^*\left\{\left(\tilde{X}_j-\bar{X}_j\right)\left(\tilde{X}_k-\bar{X}_k\right)|\mathbf{D}\right\} \\
&=\bar{X}_j\bar{X}_k\sum_{e=J-j+1}^{D}\frac{\gamma_e^*}{(\delta_e^*)^2}\frac{1}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}
\end{aligned} \tag{3.38}
$$

Dersom man nå går tilbake til uttrykket for det totale forventede kvadratiske avviket, likning 3.32 og setter inn det vi har funnet, får vi:

$$
\begin{aligned}
&\mathrm{E}\left\{\left(\sum_{j=2}^{J}(\tilde{X}_j-\bar{X}_j)\right)^2|\mathbf{D}\right\} \\
&\approx\sum_{j=2}^{J}\mathrm{E}^*\left\{\left(\tilde{X}_j-\bar{X}_j\right)^2|\mathbf{D}\right\}+2\sum_{j=2}^{J-1}\sum_{k=j+1}^{J}\mathrm{E}^*\left\{\left(\tilde{X}_j-\bar{X}_j\right)\left(\tilde{X}_k-\bar{X}_k\right)|\mathbf{D}\right\} \\
&=\sum_{j=2}^{J}3.30+2\sum_{j=2}^{J-1}\sum_{k=j+1}^{J}3.38 \\
&=\sum_{j=2}^{J}\bar{X}_j^2\sum_{e=J-j+1}^{D}\frac{\gamma_e^*}{(\delta_e^*)^2}\left\{\frac{1}{\bar{X}_{j,e-1}}+\frac{1}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}\right\} \\
&+2\sum_{j=2}^{J-1}\bar{X}_j\left(\sum_{k=j+1}^{J}\bar{X}_k\right)\sum_{e=J-j+1}^{D}\frac{\gamma_e^*}{(\delta_e^*)^2}\frac{1}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}
\end{aligned} \tag{3.39}
$$

Som vi ser er ikke dette uttrykket identisk med Neuhaus sine likninger 6.21 og 6.24. Det er et par ulikheter som alle er blitt kommentert tidligere, men som likevel oppsummeres her:

- De ytterste summene starter hos oss på $j = 2$, mens hos Neuhaus starter disse på $j = 1$. Grunnen til at vi ikke har tatt med $j = 1$ er at vi tidligere har antatt at vi kjenner alle observasjoner for første år, og derfor ikke trenger å estimere dette.

- I det andre leddet går ytterste sum hos oss til $J - 1$, mens den hos Neuhaus går til J. Grunnen til at det ikke er nødvendig å la den gå til J er at man da vil få en sum som går fra $k = J + 1$ til J, noe som per definisjon er 0.

- I begge leddene er det en sum som går fra $q = 1$ til $J - e$ hos oss, mens hos Neuhaus går denne summen til e. Dette antas å være en skrivefeil, da artikkelen av Mack [Mack 93] som Neuhaus baserer seg på har lik øvre indeks som det vi har.

- I det andre leddet til Neuhaus har han en k-sum, men ingen k-indeks inni summen. Dette antas også å være en skrivefeil.

3.3 Program i R

Vi ønsket å lage et program i statistikk-programmet R som vil regne ut likning 3.39 for oss, og dette programmet er i sin helhet gjengitt i appendiks A. Da vi begynte å studere denne likningen nærmere la vi imidlertid merke til at det noen steder brukes en estimert verdi, til tross for at vi har en observert verdi. Dette forekommer i det første leddet i likningen, og det er $1/\bar{X}_{j,e-1}$ som er aktuelt. La oss se på et eksempel:

Dersom vi ser på et tilfelle med 10 skadeår, altså $J = 10$, vil dette medføre at vi har 9 utviklingsår å basere estimeringen på, $D = 9$. Hvis vi nå ønsker å se nærmere på det andre skadeåret, $j = 2$, vil vi få at den andre summen i det første leddet går fra $e = 9$ til 9. Det aktuelle leddet vil da være $1/\bar{X}_{j,e-1} = 1/\bar{X}_{2,8}$. Dette er imidlertid ikke en verdi som skal estimeres, da vi har observert utviklingsår 8 for skader som skjedde i år 2. Dette vil forekomme for den første e-verdien for hver j. Det vi da har gjort er å bruke observasjonen som vi har, til tross for at det er et estimat som ifølge formelen skal brukes.

Da vi laget programmet i R antok vi at input vil være en øvrediagonal matrise bestående av de kjente kravene, og denne vil være på formen

$$\begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,J-1} & X_{1,J} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,J-1} & \mathrm{NA} \\ \vdots & & \iddots & & \vdots \\ X_{J,1} & \mathrm{NA} & \cdots & & \mathrm{NA} \end{bmatrix}$$

Grunnen til at matrisen ikke vil gå fra $e = 0$ er at det er problematisk å ha 0 som indeks i R. Utregningen av denne likningen i R vil altså ikke være helt identisk med den formelen vi har i likning 3.39, da en del indekser må endres for å ta hensyn til dette. Her vil $X_{j,e}$ for $j = 1, \cdots, 10, e = 1, \cdots, J - j + 1$ være de observerte kravene. Ut fra dette kan alle de ukjente størrelsene estimeres.

3.4 Resultat

Vi ønsker nå å teste programmet vårt, for å se om det gir korrekte svar. Vi har da tatt utgangspunk i den øvrediagonale matrisen som står på side 22 i England og Verrall sin artikkel, og har så gjort de nødvendige utregningene for å kunne sammenligne våre svar med de som står på side 35 i samme artikkel. Matrisen som ble brukt som input var:

Incremental Claims Data

	Development Period									
Origin Period	1	2	3	4	5	6	7	8	9	10
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

Programmet vårt estimerer så δ_e^*, γ_e^*, de resterende kravene for hvert skadeår, samt prediksjonsfeilen. De resultatene vi fikk er oppsummert i følgende tabeller:

Oversikt over estimert δ_e^* og γ_e^*:

Tabell 3.1: Resultat

Utviklings-år	Estimert δ^*	Estimert γ^*
1	2.999	27883.16
2	1.624	1108.57
3	1.271	691.43
4	1.172	61.23
5	1.113	119.44
6	1.042	40.82
7	1.033	1.34
8	1.017	7.88
9	1.009	–

Oversikt over estimerte reserver med tilhørende prediksjonsfeil:

Tabell 3.2: Resultat

Skade-år	Estimert reserve	Prediksjons-feil
1	–	–
2	154	206
3	617	623
4	1,636	747
5	2,747	1,469
6	3,649	2,002
7	5,435	2,209
8	10,907	5,358
9	10,650	6,333
10	16,339	24,566
overall	52,134	26,909

Disse resultatene er identiske med dem som står oppgitt på side 35 i artikkelen til England og Verrall, og vi antar derfor at programmet vårt gjør korrekte utregninger.

4 Simulering av data

I utregningen av MSE for det totale kravestimatet, som vi så på i forrige kapittel, ble det underveis i utregningen gjort en del tilnærmelser. Blandt annet ble det gjort to tilnærmelser i likning 3.23 da vi så på

$$\left\{\mathrm{E}\left(\tilde{X}_j|\mathbf{D}\right)-\bar{X}_j\right\}^2=\tilde{X}^2_{j,J-j}\left\{\sum_{e=J-j+1}^{D}S_e^2+2\sum_{e=J-j+1}^{D}\sum_{e'=e+1}^{D}S_eS_{e'}\right\}$$

Her fortsatte vi videre med tilnærmelsene $S_e^2\approx\mathrm{E}\left(S_e^2|\mathbf{D}_{e-1}\right)$ og $S_eS_{e'}\approx\mathrm{E}\left(S_eS_{e'}|\mathbf{D}_{e'-1}\right)$.

Bortsett fra disse to tilnærmelsene brukte vi ved flere anledninger de empiriske størrelsene som estimatorer for de ukjente størrelsene, se blandt annet likningene 3.18, 3.28 og 3.37. Også i likning 3.20 ble det gjort en tilnærmelse da $\tilde{X}_{j,e-1}$ ble erstattet med estimatet $\bar{X}_{j,e-1}$ for $e>J-j$.

Alle slike tilnærmelser vil påvirke resultatet. I hvor stor eller liten grad de får innvirkning avhenger naturlig nok av hvor gode tilnærmelsene som gjøres er. Vi ønsker nå å se om vi kan simulere de data som vi tidligere har estimert, og om vi vil få meningsfylte tall til tross for de tilnærmelsene som tidligere har blitt gjort.

Neuhaus mener at Mack gjør tre veldig spesifikke antakelser, uten å sjekke at det faktisk eksisterer noen modeller der disse vil være oppfylt. Han sier videre at det må være en grunnleggende sannsynlighetsmodell, fra hvilke man kan utlede de tre antakelsene til Mack. Til slutt påpeker han at en modell som vil passe inn med Mack sine antakelser er følgende:

Den inkrementelle utviklingen i utviklingsår e, dersom man betinger på utviklingen opp til utviklingsår $e-1$, vil være generert av en tilfeldig sammensatt Poisson variabel. Frekvensparameteren vil være proporsjonal med $\tilde{X}_{j,e-1}$, og den vil ha en alvorlighetsfordeling H_e som er uavhengig av $\tilde{X}_{j,e-1}$, men som avhenger av utviklingsåret e.

I dette kapittelet ønsker vi å generere data etter denne modellen, samtidig som vi sørger for at Mack sine antakelser er oppfylt. Men før vi setter i gang med dette ser vi litt på teorien bak den sammensatte Poisson prosessen.

4.1 Generelt om den sammensatte Poisson prosessen

En Poisson prosess med intensitet, eller rate, $\lambda > 0$ er en stokastisk prosess $\{N(t); t \geq 0\}$ der følgende tre punkt må være oppfylt [Tayl 98]:

(i) for vilkårlige tidspunkt $t_0 = 0 < t_1 < t_2 < \cdots < t_n$ er prosess-inkrementene $N(t_1) - N(t_0), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$ uavhengige tilfeldige variable.

(ii) for $s \geq 0$ og $t > 0$ har den tilfeldige variabelen $N(s+t) - N(s)$ følgende Poisson fordeling:

$$\mathrm{P}\left\{N\left(s+t\right) - N(s) = k\right\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \text{ for } k = 0, 1, \ldots$$

(iii) $N(0) = 0$

Dersom man slår sammen punkt (ii) og (iii) ser man at $N(t)$ er Poisson fordelt med parameter λt:

$$\mathrm{P}\left\{N\left(t\right) = k\right\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \text{ for } k = 0, 1, \ldots$$

Dermed er forventing og varians gitt ved:

$$\mathrm{E}\left[N(t)\right] = \mathrm{Var}\left[N(t)\right] = \lambda t$$

Vi definerer

$$N(t) = \text{antall krav som er kommet opp til tid } t.$$

Vi antar videre at kravene U_i for $i = 1, 2, \cdots, N(t)$ er stokastiske variable, og at disse er uavhengige og identisk fordelte. Det antas også at U og N er uavhengige.

Siden vi nå er interessert i å betrakte de totale akkumulerte kravene, må U og N kombineres. Vi innfører da prosessen X, som vil være totale akkumulerte krav opp til tid t:

$$X(t) = \sum_{i=1}^{N(t)} U_i$$

Dette vil være en sammensatt Poisson prosess. Vi kan videre uttrykke forventning og varians ved hjelp av N og U:

$$\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\{\mathrm{E}[X|N]\} \\
&= \mathrm{E}\left\{\mathrm{E}\left[\sum_{i=1}^{N} U_i|N\right]\right\} \\
&= \mathrm{E}\{N \cdot \mathrm{E}[U_i]\} \\
&= \mathrm{E}[N]\,\mathrm{E}[U_i]
\end{aligned}$$

Det forventede totale kravet er altså produktet mellom det forventede antall krav og forventet størrelse på hvert krav. Videre kan vi se på variansen til X:

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\{\mathrm{Var}[X|N]\} + \mathrm{Var}\{\mathrm{E}[X|N]\} \\
&= \mathrm{E}\left\{\mathrm{Var}\left[\sum_{i=1}^{N} U_i|N\right]\right\} + \mathrm{Var}\left\{\mathrm{E}\left[\sum_{i=1}^{N} U_i|N\right]\right\} \\
&= \mathrm{E}\{N \cdot \mathrm{Var}[U_i]\} + \mathrm{Var}\{N \cdot \mathrm{E}[U_i]\} \\
&= \mathrm{E}[N]\,\mathrm{Var}[U_i] + \{\mathrm{E}[U_i]\}^2\,\mathrm{Var}[N] \\
&= \mathrm{E}[N]\,\mathrm{E}\left[U_i^2\right]
\end{aligned}$$

Den siste overgangen følger av antakelsen om at $N(t)$ er Poissonfordelt, og dermed har lik forventning og varians.

4.2 Sammensatt Poisson i vårt tilfelle

Vi ønsker nå å studere hvordan Neuhaus sin foreslåtte modell henger sammen med teorien vi nettopp har sett på. Begynner med å studere Mack sin antakelse om forventning:

$$\begin{aligned}
\mathrm{E}\left[\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right] &= \delta_e \tilde{X}_{j,e-1} \\
\mathrm{E}\left[X_{j,e} + \tilde{X}_{j,e-1}|\mathbf{D}_{e-1}\right] &= \delta_e \tilde{X}_{j,e-1} \\
\mathrm{E}\left[X_{j,e}|\mathbf{D}_{e-1}\right] &= (\delta_e - 1)\tilde{X}_{j,e-1}
\end{aligned}$$

Dersom vi ser dette i sammenheng med den generelle teorien fra forrige delkapittel, ser vi at det er naturlig med

$$\mathrm{E}\left[N_{j,e}|\mathbf{D}_{e-1}\right]\mathrm{E}\left[U_{j,e,i}|\mathbf{D}_{e-1}\right] = (\delta_e - 1)\tilde{X}_{j,e-1}$$

Her vil $\mathrm{E}\left[N_{j,e}|\mathbf{D}_{e-1}\right]$ være forventet antall krav som kommer i skadeår j og utviklingsår e, gitt den historie som er kjent, mens $\mathrm{E}\left[U_{j,e,i}|\mathbf{D}_{e-1}\right]$ vil være forventet størrelse på det i'te kravet som kommer da. Vi har imidlertid antatt at U'ene er uavhengige, i motsetning til N, og det er derfor ikke nødvendig å betinge på den kjente historien.

Ut fra Neuhaus sitt eksempel på en passende modell ser vi at han også påpeker at frekvens-parameteren til Poisson variabelen skal være proporsjonal med $\tilde{X}_{j,e-1}$. Det vil altså være naturlig å sette

$$\mathrm{E}\left[N_{j,e}|\mathbf{D}_{e-1}\right] \propto \tilde{X}_{j,e-1} \Rightarrow \mathrm{E}\left[N_{j,e}|\mathbf{D}_{e-1}\right] = c_e \cdot \tilde{X}_{j,e-1}$$

Her vil c_e være en konstant som avhenger av utviklingsåret e. Dette medfører at

$$\mathrm{E}\left[U_{j,e,i}\right] = \frac{1}{c_e}(\delta_e - 1) \tag{4.1}$$

Som vi ser stemmer ikke dette helt overens med det vi finner fra Neuhaus sin første del av likning 6.26, han har ikke med konstant-faktoren $1/c_e$, hans likning så slik ut:

$$(\delta_e - 1) = \int_{-\infty}^{\infty} u \cdot dH_e(u)$$

Her vil da H_e være den kumulative fordelingen til U. Denne kan være kontinuerlig, diskret, eller muligens en kombinasjon av disse to.

Dersom vi nå går videre til Mack sin andre antakelse, ser vi at

$$\begin{aligned} \mathrm{Var}\left[\tilde{X}_{j,e}|\mathbf{D}_{e-1}\right] &= \gamma_e \tilde{X}_{j,e-1} \\ \mathrm{Var}\left[X_{j,e} + \tilde{X}_{j,e-1}|\mathbf{D}_{e-1}\right] &= \gamma_k \tilde{X}_{j,e-1} \\ \mathrm{Var}\left[X_{j,e}|\mathbf{D}_{e-1}\right] &= \gamma_e \tilde{X}_{j,e-1} \end{aligned}$$

Dersom vi ser dette i sammenheng med den generelle teorien fra forrige delkapittel, samt lar $\mathrm{E}\left[N_{j,e}|\mathbf{D}_{e-1}\right] = c_e \cdot \tilde{X}_{j,e-1}$ også her, ser vi at

$$\mathrm{E}\left[U_{j,e,i}^2\right] = \frac{1}{c_e}\gamma_e \tag{4.2}$$

Dersom vi igjen går tilbake til likning 6.26 hos Neuhaus, ser vi at også i dette leddet har vi en konstant-faktor som han ikke har, hans likning så slik ut:

$$\gamma_e = \int_{-\infty}^{\infty} u^2 \cdot dH_e(u)$$

Vi har tidligere antatt at $U_{j,e,i}$ og $N_{j,e}$ er uavhengige, men det er ikke sikkert dette alltid vil være tilfelle. Vi kan f.eks. tenke oss at dersom det en vinterdag er glatt på veiene vil det inntreffe flere skader enn på en varm sommerdag. Det er imidlertid ikke sikkert at disse skadene er så store, det er muligens mange småskader. Da vil vi ha en sammenheng mellom det faktum at vi fikk mange skader, men de var små. Vi antar likevel at vi har uavhengighet i de videre utregningene.

Hvis vi nå ser litt nøyere på den første likningen i Neuhaus sin likning 6.26, har vi:

$$(\delta_e - 1) = \mathrm{E}\left(U_{e,i}\right) = \int_{-\infty}^{\infty} u \cdot dH_e(u) \tag{4.3}$$

Fra tidligere husker vi at δ_e er utviklingsfaktoren for utviklingsår e, som forteller oss hvor mye de akkumulerte kravene øker fra utviklingsår $(e-1)$ til utviklingsår e. Det virker naturlig at økningen i de akkumulerte kravene for et skadeår j vil avta etter som årene går, og δ_e vil dermed også avta ettersom e øker. Venstre side av likning 4.3 vil dermed avta med økende e, noe som medfører at også høyre side av likningen må avta. Høyre side av likningen er imidlertid forventet størrelse på kravene, og vi ser da at størrelsen på kravene må minke etter som årene går. Dette høres litt rart ut, man skulle tro at det var de skadene med store krav som var aktuelle mange år etter skaden faktisk skjedde, se også diskusjon i kapittel 4.3.

Dersom vi inkluderer konstanten som var med i vår likning, har vi:

$$\frac{1}{c_e}(\delta_e - 1) = \mathrm{E}\left(U_{e,i}\right) \tag{4.4}$$

I dette tilfellet kan det være c_e som sørger for at venstre side av likning 4.4 totalt blir større med økende e, til tross for at δ_e minker etter som årene går. Vi må altså ha en c_e som blir mindre når e blir større, da dette vil medføre at $1/c_e$ blir større.

Dette ser også ut til å passe med likningen

$$\mathrm{E}\left[N_{j,e}|\mathbf{D}_{e-1}\right] = c_e \cdot \tilde{X}_{j,e-1}$$

Her vil det være naturlig å anta at $N_{j,e}$ har en avtagende trend med økende e, det vil bli færre og færre nye rapporterte skader jo flere år som går siden skadetidspunktet. Vi ser også at $\tilde{X}_{j,e-1}$ vil øke med økende e, da dette er akkumulerte krav. Igjen vil det altså være c_e som sørger for at høyre side totalt vil minke når e blir stor, dersom denne blir mindre og mindre med tiden.

4.3 Simulering av data

Nå vil vi prøve å simulere data etter den sammensatte Poisson modellen. Vi vil da anta at den øvre trekanten av matrisen er kjent, for så å simulere den nedre halvdelen. Igjen vil vi bruke eksempelet til England og Verrall, som vi også så på i forrige kapittel.

For simuleringen av de ukjente kravene vil vi anta at antall krav som kommer inn er Poisson fordelt med parameter $\lambda_e = c_e\tilde{X}_{j,e-1}$. Videre antas det at størrelsene på kravene er gammafordelt med parametre α_e og β_e. Disse to parametrene kan beregnes dersom man kjenner δ_e og γ_e, og disse kan estimeres ved hjelp av den kjente trekanten. Ved hjelp av likning 4.1 og 4.2, som avhenger av δ_e og γ_e, vil vi nå finne uttrykk for α_e og β_e.

Dersom vi har en gammafordeling vil tettheten være på formen

$$h(u|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} u^{\alpha-1} e^{-u/\beta}$$

Da vil forventning og varians være gitt ved $\mathrm{E}(U) = \alpha\beta$ og $\mathrm{Var}(U) = \alpha\beta^2$. Sammen med våre to likninger gir dette oss at

$$\mathrm{E}(U_{e,i}) = \alpha_e\beta_e = \frac{1}{c_e}(\delta_e - 1)$$

og

$$\mathrm{E}(U_{e,i}^2) = \alpha_e\beta_e^2 + (\alpha_e\beta_e)^2 = \frac{1}{c_e}\gamma_e$$

Etter litt regning finner man da at

$$\alpha_e = \frac{(\delta-1)^2}{c_e\gamma - (\delta-1)^2}$$

og

$$\beta_e = \frac{c_e\gamma - (\delta-1)^2}{c_e(\delta-1)}$$

Som vi vet er det et krav i gammafordelingen at α og β skal være positive, og dette vil være oppfylt for både α_e og β_e så lenge

$$c_e\gamma_e > (\delta_e - 1)^2.$$

Konstant leddet c_e vet vi desverre ikke så mye om. Som tidligere nevnt er ikke denne med i Neuhaus sin modell, men som vi har diskutert mener vi at denne må være med for at vi skal få meningsfylte resultat.

Valg av c_e-vektoren er ganske viktig, da denne påvirker flere deler av simuleringen. Som vi nettopp har sett er c_e med i simuleringen av antall skader, $\lambda_e = c_e\tilde{X}_{j,e-1}$, og den kommer også inn i simuleringen av størrelsen på skadene, gjennom α_e og β_e.

Matrisen vi tar utgangspunkt i består av inkrementelle endringer i kravene for hvert skadeår og utviklingsår. Vi vet ingenting om hvor mange skader de enkelte endringene består av, og heller ikke noe om kravstørrelsene. Det er altså ikke mulig for oss å regne "bakover", og på den måten estimere c_e-vektoren.

Den c_e-vektoren som er blitt brukt i denne oppgaven, ble funnet ved eksperimentering med ulike alternativ. Vi satt opp tre kriterier som vi mener er viktige, og ved hvert forsøk gjorde vi så en vurdering av resultatet. De kriteriene vi satt opp var:

- Det virker rimelig at den gjennomsnittlige kravstørrelsen går opp etter som årene går. De fleste småskader vil bli anmeldt og lukket i løpet av de første årene, mens det kan ta lenger tid med de store. Vi kan ihvertfall tenke oss to grunner til at store skader tar lengre tid, enten kan vi tenke oss at vi ser på en yrkesskade, der det kan gå flere år før skadens omfang blir klart, og skaden ble rapportert, eller det kan være en stor skade der forsikringstaker og forsikringsselskap ikke blir enige, og saken havner i retten. I begge disse tilfellene kan det gå flere år mellom skadedato og det tidspunktet når kravet er oppgjort.

  Vi valgte derfor å legge vekt på forholdet mellom kravstørrelsene og antall anmeldte krav.

- Når nedre halvdel av matrisen er simulert, ønsker vi å undersøke om simuleringen har fulgt de estimerte δ_e^*. Vi beregner derfor en $\hat{\delta}_e$, som er basert på hele matrisen. Hver $\hat{\delta}_e$ vil da være basert på like mange verdier,

$$\hat{\delta}_e = \frac{\sum_{j=1}^{J} \tilde{X}_{j,e}}{\sum_{j=1}^{J} \tilde{X}_{j,e-1}} \quad \text{for} \quad e = 1, \cdots, D$$

  Vi valgte å ta denne med i vurderingen også, og så da på hvor mye den avviker fra δ_e^*, samt hvor stort standardavviket til differansen var.

- Til slutt ønsket vi å se på hvordan de simulerte reservene ble i forhold til de estimerte, sammen med prediksjonsfeilen for begge alternativene.

De resultatene vi kom frem til ved de ulike alternativene for c_e-vektor, er inkludert i appendiks B. Vi bestemte oss for å bruke vektoren
$c_e = (0.76,\ 0.21,\ 0.055,\ 0.03,\ 0.015,\ 0.0045,\ 0.003,\ 0.001,\ 0.00055)$, da det var denne vektoren som etter vårt skjønn kom best ut, basert på de tre overstående kriteriene.

Vi har nå alt vi trenger for å simulere de ukjente kravene. Denne simuleringen har vi valgt å gjøre 100 ganger, for så å se på gjennomsnittet over alle forsøkene. Hele programmet som utfører denne simuleringen er gjengitt i appendiks C, så vi vil ikke kommentere hvert enkelt steg som er gjort her, men vil trekke frem hovedpoengene og resultatene.

4.4 Resultat

Vi tok som sagt utgangspunkt i eksempelet fra artikkelen til England og Verrall, der vi begynte med å estimere δ_e^* og γ_e^*. Videre gjorde vi samme beregninger som i kapittel 3, der vi til slutt fant prediksjonsfeilen for hvert skadeår, samt den totale prediksjonsfeilen for hele modellen. Dette var ved bruk av Mack sin modell.

Incremental Claims Data

	Development Period									
Origin Period	1	2	3	4	5	6	7	8	9	10
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

Videre fortsatte vi med å simulere nedre del av matrisen. Denne simuleringen ble gjort 100 ganger, og de aktuelle verdiene ble tatt vare på ved hver simulering. Det som ble sett på som aktuelt var:

- De simulerte reservene for hvert skadeår.
- De simulerte prediksjonsfeilene for hvert skadeår.
- Den simulerte prediksjonsfeilen for hele modellen.

I tillegg til dette ønsket vi også å studere δ_e^* litt mer grundig. Etter at nedre halvdel av matrisen var simulert valgte vi derfor å beregne $\hat{\delta}_e$, basert på hele matrisen, ikke kun øvre halvdel. Dette ble gjort for hver av de 100 simuleringene, og vi vil videre se på gjennomsnittet av disse, $\bar{\hat{\delta}}_e$. Vi vil også se på gjennomsnittlig differanse mellom $\bar{\hat{\delta}}_e$ og δ_e^*, før vi til slutt ser på standardavviket til denne differansene. Oppsummert ble resultatet slik:

Tabell 4.1: Resultat

Utviklings-år	δ_e^*	$\bar{\hat{\delta}}_e$	Gjennomsnittlig differanse	Standardavvik differansen
1	2.999	3.001	0.002	0.2762
2	1.624	1.623	-0.001	0.0559
3	1.271	1.270	-0.001	0.0384
4	1.172	1.172	0.000	0.0140
5	1.113	1.112	-0.001	0.0163
6	1.042	1.043	0.001	0.0117
7	1.033	1.033	0.000	0.0023
8	1.017	1.017	0.000	0.0053
9	1.009	1.009	0.000	0.0025

Som vi ser ble differansen mellom $\bar{\delta_e}$ og δ_e^* veldig liten. Også standardavviket er lite, så det ser ikke ut som om de tilnærmelser som har blitt gjort har påvirket δ_e^* i nevneverdig grad.

Ved første øyekast kan det virke litt underlig at standardavviket går nedover når utviklingsårene går oppover. I utviklingsår 1 er det for eksempel kun 2 observasjoner mer ved beregningen av $\hat{\delta}_1$, i forhold til estimeringen av δ_1^*. I utviklingsår 9 er det til sammenligning 18 flere observasjoner. Grunnen til at standardavviket likevel minker er at δ_e^* også minker. I utviklingsår 9 har vi at δ_9^* er 1.009, altså en forventet økning på 0.9%. Når det er så små tall vi ser på, vil ikke standardavviket bli så stort, til tross for at det er flere observasjoner.

Vi vil nå fortsette med å se på hvordan de simulerte reservene ble i forhold til de estimerte. Denne tabellen forteller oss hva de estimerte reservene for hvert skadeår ble, et gjennomsnitt av de simulerte reservene, samt en gjennomsnittlig differanse mellom simulert og estimert. Videre ser vi på prediksjonsfeilen, som blir kommentert mer under tabellen. Resultatene ble som følger:

Tabell 4.2: Differanse mellom simulerte og estimerte gjenstående reserver

Skade-år	Estimerte reserver	Simulerte reserver	Differanse	Pred. feil simulering	Pred. feil estimering
1	–	–	–	–	–
2	154	178	24	151	206
3	617	659	42	421	623
4	1,636	1,603	−33	540	747
5	2,747	2,618	−129	1,199	1,469
6	3,649	3,754	105	1,841	2,002
7	5,435	5,400	−35	2,229	2,209
8	10,907	10,601	−306	4,356	5,358
9	10,650	10,786	136	6,352	6,333
10	16,339	16,225	−114	18,649	24,566
Overall	52,134	51,824	−310	20,211	26,909

Den simulerte reserven avviker ikke så mye fra den estimerte, og den har også mindre prediksjonsfeil enn den estimerte. Det ser altså ut som om modellen som Neuhaus foreslår passer ganske bra, til tross for de tilnærmelser som er blitt gjort i utregningen.

Man kan imidlertid lure på om den modellen som Neuhaus foreslår, som passer med Mack sine tre antakelser, er i tråd med det man vil se i virkeligheten. Ut fra modellen antas det at det betingede forventede antall skader skal være proporsjonalt med totale akkumulerte krav. Dette vil medføre at dersom man i et skadeår har en del store skader, noe som vil gi et høyt akkumulert beløp, så kan man i neste utviklingsår forvente å få mange skader. Dette høres litt rart ut, burde virkelig antall nye skader avhenge av akkumulerte krav frem til siste kjente år? Eller kan man tenke seg at dersom man har mange store skader ett år, så blir det færre det neste, at man totalt sett kommer til å ende opp med samme utfall, det varierer bare litt hvilket år skaden blir anmeldt. Man kan isåfall tenke seg at man burde ha en mer “glattet” modell, der det ikke blir lagt fullt så mye vekt på slike unormalt store skadebeløp.

5 Overdispersert Poisson Modell

I forrige kapittel så vi på den sammensatte Poisson modellen. Der antok vi, som vanlig, at forventning og varians er lik. Dette er imidlertid ikke alltid tilfelle i det virkelige liv, det vil ofte være større varians enn forventning. Vi kan da ha en overdispersert Poisson modell, som vi nå skal se litt nærmere på.

5.1 Generelt om GLM og overdispersert Poisson modell

Forskjellen mellom en overdispensert Poisson modell og den vanlige Poisson modellen er at variansen ikke lenger er lik forventningen, men er proporsjonal med den. Overdispersjonen representeres med en faktor $\phi > 1$, som kan estimeres som en del av tilpassningsprosessen. Veldig enkelt kan vi da si at

$$\mathrm{E}[X] = m \;\text{ og }\; \mathrm{Var}[X] = \phi m$$

Her vil X være en responsvariabel, som det blir gjort uavhengige observasjoner av ved bestemte verdier for forklaringsvariablene $y_i, \cdots, y_p$. Denne påvirkningen kan kun skje gjennom en enkel lineær funksjon, den lineære prediktoren η:

$$\eta = \beta_1 y_1 + \cdots + \beta_p y_p$$

Forventningen m vil være en glatt invertibel funksjon av den lineære prediktoren,

$$m = h(\eta) \quad \text{og} \quad \eta = h^{-1}(m) = g(m)$$

der g kalles link funksjonen mellom forventningen og den lineære prediktoren.
Dersom tettheten til den overdisperserte variabelen X er på kanonisk form kan den, ifølge Venables og Ripley [Vena 99], skrives som:

$$f(X_i; \theta_i) = \exp\left\{A_i\left[X_i\theta_i - \gamma(\theta_i)\right]/\phi + \tau(X_i, \phi/A_i)\right\} \tag{5.1}$$

Her vil ϕ være dispersjonsparameteren, A_i er en kjent vekt, og parameteren θ_i avhenger av den lineære prediktoren. Dersom ϕ er ukjent blir den behandlet som en nuisance parameter. I vårt tilfelle vil vi anta at alle observasjonene har lik vekting, og vi vil da sette $A_i = 1$.

For å finne ut hvordan dispersjonsparameteren ϕ påvirker estimeringen av parametrene, må vi undersøke hvordan den påvirker forventningen og variansen til X. Vi begynner da med å finne score funksjonen, $U(\theta; X)$:

$$
\begin{aligned}
f(X;\theta) &= \exp\left\{\left[X\theta - \gamma(\theta)\right]/\phi + \tau(X,\phi)\right\} \\
l(\theta;X) &= \log f(X;\theta) = \left[X\theta - \gamma(\theta)\right]/\phi + \tau(X,\phi) \\
U(\theta;X) &= \tfrac{dl(\theta;X)}{d\theta} = \left[X - \gamma'(\theta)\right]/\phi
\end{aligned}
$$

Vi vet at når vi ser på en score funksjon, så er forventningen til scoren null, og variansen kan skrives som $\mathrm{Var}(U) = -\mathrm{E}(U')$. Vi vil nå bruke disse to tingene til å finne forventning og varians til X. Begynner med å se på forventningen til U:

$$
\begin{aligned}
\mathrm{E}(U) &= \mathrm{E}\left[(X - \gamma'(\theta))/\phi\right] \\
&= \left[\mathrm{E}(X) - \gamma'(\theta)\right]/\phi \\
&= 0 \\
\\
\mathrm{E}(X) &= \gamma'(\theta)
\end{aligned}
$$

Som nevnt tidligere skal forventningen være en funksjon av den lineære prediktoren, $\mathrm{E}(X) = m = h(\eta)$. Siden dispersjonsparameteren ikke har noen innvirkning på forventningen til X, har den heller ingen innvirkning på estimeringen av parametrene $\beta_1, \cdots, \beta_p$. Vi fortsetter med utregningen av variansen til X:

$$
\begin{aligned}
\mathrm{Var}(U) &= -\mathrm{E}(U') \\
\mathrm{Var}\left[(X - \gamma'(\theta))/\phi\right] &= -\mathrm{E}\left[-\gamma''(\theta)/\phi\right] \\
(1/\phi)^2\mathrm{Var}(X) &= (1/\phi)\gamma''(\theta) \\
\mathrm{Var}(X) &= \phi\gamma''(\theta)
\end{aligned}
$$

Som vi ser kan dispersjonsparameteren ha innvirkning på variansen til estimatene for parametrene. La oss se på det vanlige tilfellet med en Poisson fordelt variabel X, der vi har lik forventning og varians, μ. Fordelingen til X vil da være gitt ved:

$$
f(X_i;\theta_i) = \exp\left\{X\log\mu - \mu - \log(X!)\right\}
$$

Dersom dette skal være lik uttrykket til Venables og Ripley, likning 5.1, må vi ha at

$$\begin{aligned} \theta_i &= \log \mu \\ \gamma(\theta_i) &= \mu = e^{\theta_i} \\ \phi &= 1 \end{aligned}$$

Som ventet får vi da at ϕ må være lik 1, noe som gjør at uttrykkene for forventning og varians blir like. I de tilfeller der man får $\phi < 1$ eller $\phi > 1$, har man henholdsvis underdispersjon og overdispersjon. Da vil ϕ påvirke variansen til parametrene, og forventning og varians vil ikke lenger være lik. La oss nå se på hva som vil skje dersom man har overdispersjon, som vi antar er tilfellet med våre data.

5.2 GLM og overdispersert Poisson modell i vårt tilfelle

Vi har som tidligere variabelen $X_{j,e}$, den inkrementelle endringen for kravene i utviklingsår e, for skader som skjedde i skadeår j. Denne antar vi nå at følger en overdispersert Poisson fordeling, med forventning $m_{j,e}$ og varians som er proporsjonal med denne:

$$\mathrm{E}[X_{j,e}] = m_{j,e} \text{ og } \mathrm{Var}[X_{j,e}] = \phi m_{j,e}$$

Summen av de inkrementelle kravene i kolonne e må være positiv, noe som er en begrensning på modellen. Det er likevel tillatt med noen negative inkrementelle krav, så lenge summen blir positiv.

Vi bruker en logaritmisk link funksjonen, og den lineære prediktoren

$$\eta_{j,e} = \log(m_{j,e}) = c + \alpha_j + \beta_e$$

Som vi ser er strukturen fremdeles en slags chain ladder, det er fremdeles en parameter for hver rad j, og en parameter for hver kolonne e.

Overdispersjonen blir tatt hensyn til ved å estimere den ukjente skaleringsparameteren ϕ som en del av tilpasningsprosedyren. Vanligvis vil denne estimeringen basere seg på bruk av deviansen. Et uttrykk for deviansen kan vi finne fra boken "Modern Applied Statistics with S-PLUS" [Vena 99], likning 7.5:

$$D_M = 2\sum_{i=1}^{n} \left[\{X_i\theta(X_i) - \gamma(\theta(X_i))\} - \{X_i\hat{\theta}_i - \gamma(\hat{\theta}_i)\} \right]$$

Dersom vi igjen ser på tilfellet med Poisson prosessen, så ser vi at uttrykkene fra forrige delkapittel nå vil bli

$$\begin{aligned}\theta(X_i) &= \log X_i \\ \gamma(\theta(X_i)) &= X_i \\ \hat{\theta}_i &= \log \hat{\mu}_i \\ \gamma(\hat{\theta}_i) &= \hat{\mu}_i\end{aligned}$$

Vi setter dette inn i den overstående likningen:

$$\begin{aligned}D_M &= 2\sum_{i=1}^{n}\left[\{X_i \log X_i - X_i\} - \{X_i \log \hat{\mu}_i - \hat{\mu}_i\}\right] \\ &= 2\sum_{i=1}^{n}\left[X_i \log(X_i/\hat{\mu}_i) - (X_i - \hat{\mu}_i)\right]\end{aligned} \tag{5.2}$$

Når deviansen er regnet ut, kan man finne en estimator for ϕ, ved å bruke likning 7.6 fra samme boken. Her vil n være antall observasjoner, og $p < n$ er antall regresjons-parametre:

$$\hat{\phi} = D_M/(n-p) \tag{5.3}$$

Som nevnt tidligere i dette kapittelet ønsker vi i vårt tilfelle å tillate at enkelte inkrementelle krav er negative. Dette vil gi oss et problem i utregningen av deviansen. Dersom vi går tilbake til forrige delkapittel og ser på scorefunksjonen, ser vi at maximum likelihood estimatoren for θ_i finnes ved å sette $X_i = \gamma'(\hat{\theta}_i)$:

$$U(\theta; X) = \left[X - \gamma'(\theta)\right]/\phi$$

Vi har også fra forrige delkapittel av dersom man ser på en Poisson prosess, så er $\gamma(\theta_i) = e^{\theta_i}$, noe som medfører at $\gamma'(\theta_i) = e^{\theta_i}$. Dersom vi nå tillater at enkelte X_i kan være negative, ser vi fra maximum likelihood estimatoren at det medfører at $\gamma'(\theta_i) = e^{\theta_i}$ er negativ, noe som ikke er mulig. Vi kan altså ikke bruke denne metoden for å finne en estimator for ϕ.

Vi går nå tilbake til likning 5.2, som så slik ut:

$$D_M = 2\sum_{i=1}^{n}\left[X_i \log(X_i/\hat{\mu}_i) - (X_i - \hat{\mu}_i)\right]$$

Vi gjør så en Taylorutvikling for det første uttrykket som står inni summen, omkring $X_i = \hat{\mu}_i$. Denne gir oss til ledende orden:

$$X_i \log(X_i/\hat{\mu}_i) = (X_i - \hat{\mu}_i) + \frac{1}{2}\frac{(X_i - \hat{\mu}_i)^2}{\hat{\mu}_i} + \cdots$$

Hvis vi nå setter dette inn i likningen, ser vi at

$$
\begin{aligned}
D_M &\approx D_M^* \\
&= 2\sum_{i=1}^{n}\left[(X_i - \hat{\mu}_i) + \frac{1}{2}\frac{(X_i - \hat{\mu}_i)^2}{\hat{\mu}_i} - (X_i - \hat{\mu}_i)\right] \\
&= \sum_{i=1}^{n}\frac{(X_i - \hat{\mu}_i)^2}{\hat{\mu}_i}
\end{aligned}
$$

Dersom vi nå benytter likning 5.3 med denne nye deviansen, har vi funnet en ny estimator for ϕ:

$$
\tilde{\phi} = D_M^*/(n-p) = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(X_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \tag{5.4}
$$

Vi har nå kommet frem til likning 7.8 i boken “Modern Applied Statistics with S-PLUS” [Vena 99]. Det er denne estimatoren for ϕ vi må bruke i vårt tilfelle.
Dersom man bruker statistikk-program som for eksempel R til å tilpasse GLM, vil også dispersjonsparameteren ϕ bli estimert. Vanligvis vil man da bruke en innebygget funksjon i R, *quasipoisson*, som vil bruke den første metoden med deviansen for å estimere ϕ. Her er et utdrag av dette programmet:

```
...
dev.resids = function(y,mu,wt) 2*wt*(y*log(ifelse(y==0, 1, y/mu))
- (y - mu))
aic = function(y,n,mu,wt,dev) NA
initialize = expression({
   if (any y<0)) stop("negative values not allowed for the
   quasiPoisson family")
   n = rep.int(1, nobs)
   mustart = y + 0.1
})
...
```

Som vi ser vil vi i første linje få problemer, dersom vi har en y som er negativ. Dette var imidlertid forventet, da vi tidligere fant ut at vi i vårt tilfelle ikke kan bruke denne metoden med devianse for å estimere ϕ. Heldigvis er det allerede en løsning på dette problemet, David Firth har kommet med følgende kode som skal erstatte koden vi nettopp så på, der den andre metoden for estimering av ϕ brukes:

```
...
dev.resids = function(y,mu,wt) wt*(y-mu)^2/mu
aic = function(y,n,mu,wt,dev) NA
initialize = expression({
   n = rep(1, nobs)
   mustart = rep(mean(y),length(y))
})
...
```

Her ser vi at den første linjen er blitt erstattet med likning 5.4, og i vårt tilfelle har vi at *wt = A = 1*.

Begge disse versjonene av *quasipoisson* er gjengitt i avsnitt 3.5 i De Silvas artikkel "An Introduction to R: Examples for Actuaries" [De S 06], der også fremgangsmåten for beregning av GLM i vårt tilfelle blir gjennomgått. Dette er gjengitt i sin helhet i appendiks D, og vi vil nå se litt mer på teorien bak utregningen av MSEP, som blir brukt i denne artikkelen.

5.3 Mean Squared Error of Prediction

Dersom vi har en tilfeldig variabel X med predikert verdi $\bar{X}$, er MSEP gitt ved likningen:

$$\begin{aligned}\text{MSEP} &= \mathrm{E}\left[(X-\bar{X})^2\right] \\ &= \mathrm{E}\left[(X-\mathrm{E}X-\bar{X}+\mathrm{E}X)^2\right] \\ &= \mathrm{E}\left[(X-\mathrm{E}X)^2\right]+\mathrm{E}\left[(\bar{X}-\mathrm{E}X)^2\right]-2\mathrm{E}\left[(X-\mathrm{E}X)(\bar{X}-\mathrm{E}X)\right]\end{aligned} \tag{5.5}$$

Den første delen av likningen er per definisjon variansen til X. Ser litt nærmere på den andre delen:

$$\begin{aligned}\mathrm{E}\left[(\bar{X}-\mathrm{E}X)^2\right] &= \mathrm{E}\left[(\bar{X}-\mathrm{E}\bar{X}-\mathrm{E}X+\mathrm{E}\bar{X})^2\right] \\ &= \mathrm{E}\left[(\bar{X}-\mathrm{E}\bar{X})^2\right]+\mathrm{E}\left[(\mathrm{E}X-\mathrm{E}\bar{X})^2\right]-2\mathrm{E}\left[(\bar{X}-\mathrm{E}\bar{X})(\mathrm{E}X-\mathrm{E}\bar{X})\right] \\ &= \mathrm{Var}(\bar{X})+(\mathrm{E}X-\mathrm{E}\bar{X})^2-2(\mathrm{E}X-\mathrm{E}\bar{X})\mathrm{E}(\bar{X}-\mathrm{E}\bar{X}) \\ &= \mathrm{Var}(\bar{X})+(\text{“bias”})^2\end{aligned}$$

Kryssleddet forsvinner siden $\mathrm{E}(\bar{X}-\mathrm{E}\bar{X})=0$.

"Biasen", som er differansen mellom forventningen til estimatoren og den sanne verdien til parameteren, antar vi at er liten, og dropper i de videre utregningene. Dersom vi nå fortsetter med å se på det siste leddet fra likning 5.5, samt erstatter $\mathrm{E}X$ med $\mathrm{E}\bar{X}$ i det siste leddet, får vi at:

$$\begin{aligned}\mathrm{E}\left[(X-\mathrm{E}X)(\bar{X}-\mathrm{E}X)\right] &\approx \mathrm{E}\left[(X-\mathrm{E}X)(\bar{X}-\mathrm{E}\bar{X})\right] \\ &= \mathrm{E}\left\{\mathrm{E}\left[(X-\mathrm{E}X)(\bar{X}-\mathrm{E}\bar{X})|\text{historie}\right]\right\} \\ &= \mathrm{E}\left\{(\bar{X}-\mathrm{E}\bar{X})\underbrace{\mathrm{E}\left[(X-\mathrm{E}X)|\text{historie}\right]}_{=0}\right\} \\ &= 0\end{aligned}$$

Dersom vi nå setter tilbake alt vi har funnet i likning 5.5, ser vi at:

$$\text{MSEP}(\bar{X}) \approx \mathrm{Var}(X)+\mathrm{Var}(\bar{X})$$

Vi går nå videre til å se på dette i vårt tilfelle, der $X_{j,e}$ er inkrementell endring for skadene skjedd i år j, i utviklingsår e. Benevnelsen $\bar{X}_{j,e}$ brukes som tidligere om den predikerte verdien for denne kombinasjonen. Utregningen av MSEP vil bli akkurat lik som i det generelle tilfellet, og vi får at

$$\text{MSEP}(\bar{X}_{j,e}) \approx \text{Var}(X_{j,e}) + \text{Var}(\bar{X}_{j,e}) \tag{5.6}$$

Vi har antatt at $X_{j,e}$ har en overdispersert Poisson fordeling, og variansen er dermed gitt ved $\text{Var}(X_{j,e}) = \phi m_{j,e}$. For å finne variansen til $\bar{X}_{j,e}$ bruker vi Delta metoden, der vi gjør en Taylor utvikling av første orden. Generelt vil dette si at dersom V er en tilfeldig variabel med $\text{E}V = \mu$, vil en første ordens tilnærming gi oss

$$\begin{aligned} g(V) &\approx g(\mu) + g'(\mu)(V - \mu) \\ \text{Var}[g(V)] &\approx [g'(\mu)]^2 \text{Var}(V) \end{aligned}$$

I vårt tilfelle vil

$$\begin{aligned} V &= \bar{\eta}_{j,e} \quad \text{og} \quad g(V) = \bar{X}_{j,e} = e^{\bar{\eta}_{i,j}} \\ \mu &= \eta_{i,j} \quad \text{og} \quad g(\mu) = e^{\eta_{j,e}} \end{aligned}$$

Dersom vi setter dette inn i formelen for variansen, får vi at

$$\begin{aligned} \text{Var}(\bar{X}_{j,e}) &\approx \left[\frac{d}{d\eta_{j,e}} e^{\eta_{j,e}}\right]^2 \text{Var}(\bar{\eta}_{j,e}) \\ &= [e^{\eta_{j,e}}]^2\, \text{Var}(\bar{\eta}_{j,e}) \\ &= m_{j,e}^2 \text{Var}(\bar{\eta}_{j,e}) \end{aligned}$$

Vi setter nå det vi har funnet tilbake i likning 5.6, samtidig som vi erstatter $m_{j,e}$ med de estimerte verdiene. Vi vil da få et estimat for $\text{MSEP}(\bar{X}_{j,e})$, men for enkelhets skyld vil vi ikke ha "hatt" på denne videre i oppgaven:

$$\widehat{\text{MSEP}(\bar{X}_{j,e})} = \text{MSEP}(\bar{X}_{j,e}) \approx \phi \bar{m}_{j,e} + \bar{m}_{j,e}^2 \text{Var}(\bar{\eta}_{j,e})$$

Denne forteller oss prediksjonsfeilen til den ene inkrementelle endringen $\bar{X}_{j,e}$. Vi ønsker nå å se på MSEP for det totale kravestimatet. Vi begynner med å innføre Δ, som representerer det området som skal estimeres:

$$\begin{bmatrix} X_{1,0} & X_{1,1} & \cdots & X_{1,D-1} & X_{1,D} \\ X_{2,0} & X_{2,1} & \cdots & X_{2,D-1} & \bar{X}_{1,D} \\ \vdots & & & \Delta & \vdots \\ X_{J,0} & \bar{X}_{J,1} & \cdots & & \bar{X}_{J,D} \end{bmatrix}$$

Videre innfører vi benevnelsen $\bar{X}_{++}$ for forventningen til det totale kravestimatet, summert over både skadeår og utviklingsår. Denne kan da settes opp som en sum:

$$\bar{X}_{++} = \sum_{j,e \in \Delta} \bar{X}_{j,e}$$

For å finne MSEP til denne summen går vi tilbake til likning 5.6, og setter inn $\bar{X}_{++}$:

$$\text{MSEP}[\bar{X}_{++}] = \text{Var}[X_{++}] + \text{Var}[\bar{X}_{++}] \tag{5.7}$$

La oss nå se litt nærmere på de to leddene hver for seg. I begge utregningene vil vi bruke følgende regel for varians til en sum:

$$\text{Var}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j)$$

Vi begynner med å se på det første leddet på høyre side i likning 5.7:

$$\begin{aligned} \text{Var}[X_{++}] &= \text{Var}\left[\sum_{j,e \in \Delta} X_{j,e}\right] \\ &= \sum_{j,e \in \Delta} \text{Var}(X_{j,e}) + \sum_{\substack{j_1,e_1 \in \Delta \\ j_2,e_2 \in \Delta \\ j_1,e_1 \neq j_2,e_2}} \text{Cov}(X_{j_1,e_1}, X_{j_2,e_2}) \\ &= \sum_{j,e \in \Delta} \phi m_{j,e} \end{aligned} \tag{5.8}$$

Vi har antatt at det er uavhengighet mellom de ulike skadeårene, og vil derfor ikke få noen kovarians fra leddene der $j_1 \neq j_2$. Vi har også antatt at $X_{j,e}$ er en Poissonfordelt variabel, og som tidligere nevnt er et av kjennetegnene ved Poissonprosessen at den er hukommelsesløs. Det vil altså heller ikke bli noe kovarians fra leddene der $e_1 \neq e_2$, og hele summen vil bli 0. Går videre med det andre leddet i likning 5.7.

$$
\begin{aligned}
\mathrm{Var}[\bar{X}_{++}] &= \mathrm{Var}\left[\sum_{j,e\in\Delta} \bar{X}_{j,e}\right] \\
&= \mathrm{Var}\left[\sum_{j,e\in\Delta} e^{\bar{\eta}_{j,e}}\right] \\
&\approx \mathrm{Var}\left[\sum_{j,e\in\Delta} \left(e^{\eta_{j,e}} + m_{j,e}(\bar{\eta}_{j,e} - \eta_{j,e})\right)\right] \\
&= \mathrm{Var}\left[\sum_{j,e\in\Delta} m_{j,e}\bar{\eta}_{j,e}\right] \\
&= \sum_{j,e\in\Delta} m_{j,e}^2 \mathrm{Var}\left[\bar{\eta}_{j,e}\right] + \sum_{\substack{j_1,e_1\in\Delta \\ j_2,e_2\in\Delta \\ j_1,e_1\neq j_2,e_2}} m_{j_1,e_1} m_{j_2,e_2} \mathrm{Cov}\left(\bar{\eta}_{j_1,e_1}, \bar{\eta}_{j_2,e_2}\right)
\end{aligned}
\tag{5.9}
$$

I denne utregningen har vi brukt Delta metoden med en Taylor utvikling av første grad, og vi har også brukt det faktum at når vi tar varians til en konstant så får vi 0. Dersom vi nå erstatter $m_{j,e}$ med de estimerte verdiene, samt setter det vi har funnet tilbake i likning 5.7, så har vi nesten kommet frem til likning 7.9 i artikkelen til England og Verrall [Engl 02]. Eneste forskjell er at de har et 2 tall foran den siste summen, men dette antar vi er en skrivefeil, da det ikke brukes noe 2 tall i utregningen av likningen.

$$
\begin{aligned}
\mathrm{MSEP}[\bar{X}_{++}] &= \mathrm{Var}[X_{++}] + \mathrm{Var}[\bar{X}_{++}] \\
&= 5.8 + 5.9 \\
&= \sum_{j,e\in\Delta} \phi\bar{m}_{j,e} + \sum_{j,e\in\Delta} \bar{m}_{j,e}^2 \mathrm{Var}\left[\bar{\eta}_{j,e}\right] + \sum_{\substack{j_1,e_1\in\Delta \\ j_2,e_2\in\Delta \\ j_1,e_1\neq j_2,e_2}} \bar{m}_{j_1,e_1}\bar{m}_{j_2,e_2} \mathrm{Cov}\left(\bar{\eta}_{j_1,e_1}, \bar{\eta}_{j_2,e_2}\right)
\end{aligned}
\tag{5.10}
$$

De Silva [De S 06] har laget et program i R som regner ut prediksjonsfeilen ved bruk av denne likningen, samt den totale forventede reserven. Hele dette programmet er gjengitt i appendiks D.

5.4 Resultat

Dersom vi igjen ser på eksempelet fra artikkelen til England og Verrall, og bruker De Silva sitt program, får vi ut at forventet reserve er på $52,134$, med prediksjonsfeil på $17,613$.

Programmet til De Silva gir kun ut den totale forventede reserven og prediksjonsfeilen, ikke fordelt på skadeår. Siden vi også er interessert i å ha dette fordelt på skadeår, har vi laget en ekstra programmeringsdel, som også er inkludert i appendiks D.

Tabell 5.1: Resultat

Skade-år	GLM reserver	Prediksjons-feil
1	–	–
2	154	538
3	617	1,084
4	1,636	1,719
5	2,747	2,160
6	3,649	2,362
7	5,435	3,025
8	10,907	4,871
9	10,650	5,881
10	16,339	12,572
overall	52,134	17,613

Tabellen ovenfor viser forventede reserver og prediksjonsfeil per skadeår.

6 Konklusjon

Hensikten med denne oppgaven var å studere reservering til fremtidige krav, først og fremst ved bruk av Chain Ladder metoden og Mack sin modell. Vi har sett at Mack gjør tre antakelser for å komme frem til sitt resultat, noe Neuhaus mener ikke kan gjøres uten å sjekke at det faktisk eksisterer noen modeller som oppfyller disse kravene. Neuhaus påpeker videre et modell-alternativ som vil tilfredsstille Mack sine krav, den sammensatte Poisson modellen.

I hele oppgaven har vi brukt eksempelet fra side 22 i artikkelen til England og Verrall som utgangspunkt. Vi har antatt at det øvre triangelet av matrisen er kjent, og har så brukt ulike metoder for å beregne reserver og prediksjonsfeil på den ukjente delen.

Vi laget et program i R som regnet ut reserve og prediksjonsfeil ved bruk av Mack sin modell. Videre tok vi utgangspunkt i modellen som Neuhaus foreslo, den sammensatte Poissonmodellen, og simulerte nedre trekant av matrisen 100 ganger. Vi brukte gjennomsnittlige verdier i den videre regningen. Til slutt brukte vi programmet til De Silva, som ved hjelp av GLM regnet ut reserve og prediksjonsfeil. Oppsummert ble resultatene som følger:

Tabell 6.1: Oppsummering

	Total reserve	Total prediksjonsfeil
Mack's modell	52,134	26,909
Simulering	51,824	20,211
De Silva's program	52,134	17,613

Som vi ser får vi ganske lik reserve ved alle tre fremgangsmåtene. At reservene ved bruk av Mack sin modell og De Silva sitt program er like, er som forventet. Teorien bak programet til De Silva er hentet fra artikkelen til England og Verrall, og i denne artikkelen står det i avsnitt 2.1.5 følgende:

> *The first step to obtaining the prediction error is to formulate an underlying statistical model making assumptions about the data. If the aim is to provide a stochastic model which is analogous to the chain-ladder technique, then an obvious first requirement is that the predicted values should be the same as those of the chain-ladder technique.*

Det er altså ikke uventet at de to reservene er like. La oss imidlertid undersøke om det kun er totalen som er lik, eller også per skadeår. Her er en oppsummeringen fordelt på skadeår:

Tabell 6.2: Oppsummering per skadeår

Skade-år	Estimerte reserver	Simulerte reserver	GLM reserver	Pred. feil estimering	Pred. feil simulering	Pred. feil GLM
1	–	–	–	–	–	–
2	154	178	154	206	151	538
3	617	659	617	623	421	1,084
4	1,636	1,603	1,636	747	540	1,719
5	2,747	2,618	2,747	1,469	1,199	2,160
6	3,649	3,754	3,649	2,002	1,841	2,362
7	5,435	5,400	5,435	2,209	2,229	3,025
8	10,907	10,601	10,907	5,358	4,356	4,871
9	10,650	10,786	10,650	6,333	6,352	5,881
10	16,339	16,225	16,339	24,566	18,649	12,572
Overall	52,134	51,824	52,134	26,909	20,211	17,613

Som vi ser er det ikke bare de totale reservene for Mack sin modell og De Silva sitt program som er identiske, også fordelt på skadeår er de like. Dette stemmer bra overens med ønsket om å lage en stokastisk modell som er analog til Chain Ladder teknikken.

Vi ser at også de simulerte reservene ligger i samme område som de vi fikk ved estimeringen i Mack sin modell og ved bruk av GLM, og dette kan tyde på at de tilnærmelser som Mack gjør underveis ikke får så store konsekvenser for prediksjonen av de totale reservene. Dersom vi ser på prediksjonsfeilen, ser vi imidlertid at Mack sin modell helt klart har størst feil. Dette kan komme av de tilnærmelser som har blitt gjort, de fører muligens til større usikkerhet rundt resultatet.

Det er vanskelig å trekke en endelig konklusjon ut fra de data som vi har her. Dersom målet er å ha minst mulig prediksjonsfeil bør man velge modellen som De Silva har brukt, mens dersom man er litt mer forsiktig og vil være på den sikre siden kan man velge Mack sin modell. I en bedrift kan muligens bedriftens politikk også være med på å bestemme hvilken modell man går for, dersom bedriften generelt har en forsiktig politikk, vil de ikke nødvendigvis velge den modellen som har minst prediksjonsfeil, mens dersom bedriften har en offensiv politikk, er det mulig de kun vil velge modell basert på prediksjonsfeilen.

I bunn og grunn har alle modellene ting som taler både for og imot seg. Til slutt må man bare velge den modellen man har mest tro på, den modellen som har forutsetninger som passer best med det tilfellet man ser på.

Bibliografi

[De S 06] N. De Silva. "An Introduction to R: Examples for Actuaries. Actuarial Toolkit Working Party, version 0.1". 2006. http://toolkit.pbworks.com/R-Examples-For-Actuaries.

[Engl 02] P. England and R. Verrall. "Stochastic claims reserving in general insurance". *British Actuarial Journal*, Vol. 8, No. 3, pp. 443–518, 2002.

[Mack 93] T. Mack. "Distribution-free calculation of the standard error of chain ladder reserve estimates". *Astin Bulletin*, Vol. 23, No. 2, pp. 213–225, 1993.

[Neuh 06] W. Neuhaus. "Lecture notes on estimating outstanding claims in general insurance". 2006. Unpublished lecture notes.

[Tayl 98] H. Taylor and S. Karlin. *An introduction to stochastic modeling*. Academic Press San Diego, 3rd Ed., 1998.

[Vena 99] W. Venables and B. Ripley. *Modern Applied Statistics with S-PLUS*. Springer, 3rd Ed., 1999.

A

Program i R som regner ut prediksjonsfeilen

Vi ønsker å lage et program i R som regner ut likning 3.39:

$$
\begin{aligned}
&\mathrm{E}\left\{\left(\sum_{j=2}^{J}(\tilde{X}_j - \bar{X}_j)\right)^2 |\mathbf{D}\right\} \\
&= \sum_{j=2}^{J}\bar{X}_j^2 \sum_{e=J-j+1}^{D}\frac{\gamma_e^*}{(\delta_e^*)^2}\left\{\frac{1}{\bar{X}_{j,e-1}} + \frac{1}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}\right\} \\
&+ 2\sum_{j=2}^{J-1}\bar{X}_j\left(\sum_{k=j+1}^{J}\bar{X}_k\right)\sum_{e=J-j+1}^{D}\frac{\gamma_e^*}{(\delta_e^*)^2}\frac{1}{\sum_{q=1}^{J-e}\tilde{X}_{q,e-1}}
\end{aligned}
$$

I R er det imidlertid ikke mulig å ha indeks 0 i en matrise, så i det følgende programmet vil vi la e gå fra 1 til J. Indeksene som brukes vil da ikke være helt identiske med det som har stått tidligere i oppgaven, men dette er da for å korrigere for denne endringen.

Input vil være en matrise på formen

$$
\begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,J-1} & X_{1,J} \\
X_{2,1} & X_{2,2} & \cdots & X_{2,J-1} & \mathrm{NA} \\
\vdots & & \iddots & & \vdots \\
X_{J,1} & \mathrm{NA} & \cdots & & \mathrm{NA}
\end{bmatrix}
$$

Denne matrisen kan enten inneholde utbetalinger for hvert utviklingsår, eller akkumulerte tall. Dersom tallene ikke er akkumulerte må de akkumuleres, da det er dette som trengs for de videre beregningene. I dette programmet antas det at input ikke er akkumulert, men dersom det motsatte er tilfelle er det bare å hoppe over det første steget.

Først må vi finne ut hvor stor matrise vi har, da dette vil fortelle oss hvor mange skadeår vi ser på.

```
J=length(X_matrise)
```

Vi vil nå lage en akkumulert matrise, som er den som vil bli brukt i de videre utregningene.

```
X_tilde_matrise = matrix(rep(NA),J,J)
for (i in 1:J){
   if (i==1) {X_tilde_matrise[,i] = X_matrise[,i]}
   else X_tilde_matrise[,i] = X_tilde_matrise[,i-1] + X_matrise[,i]
}
```

Videre vil vi lage en vektor som består av de estimerte utviklingsfaktorene. Formelen for dette er gitt ved likning 2.1:

$$\delta_e^* = \frac{\sum_{j=1}^{J-e} \tilde{X}_{j,e}}{\sum_{j=1}^{J-e} \tilde{X}_{j,e-1}}$$

```
delta = rep(0,J-1)
for (e in 2:J){
   delta[e-1] = sum(X_tilde_matrise[1:(J-e+1),e]) /
                sum(X_tilde_matrise[1:(J-e+1),e-1])
}
```

Vi trenger også en vektor som består av γ_e^*. Dette uttrykket finner vi fra likning 3.4.

$$\gamma_e^* = \frac{1}{J-e-1} \sum_{j=1}^{J-e} \tilde{X}_{j,e-1} \left(\frac{\tilde{X}_{j,e}}{\tilde{X}_{j,e-1}} - \delta_e^* \right)^2$$

Det er ikke mulig å estimere γ_D^*, da denne kun har en observasjon å basere seg på. Det er derfor vanlig å sette denne lik et av de to foregående estimatene. Her har vi valgt å sette $\gamma_D^* = \gamma_{D-2}^*$.

```
gamma = rep(NA,J-1)
for (e in 1:(J-2)){
   gamma[e] = (1/(J-e-1))*sum(X_tilde_matrise[1:(J-e),e] *
           ((X_tilde_matrise[1:(J-e),e+1] / X_tilde_matrise[1:(J-e),e])
           - delta[e])^2)
   }
gamma[J-1] = gamma[J-3]
```

Man trenger også en matrise med estimatene av de totale kravene. Denne vil bestå av det samme øvre triangelet som i den akkumulerte matrisen, mens det nedre triangelet vil bli fylt ut med de estimerte verdiene:

$$\begin{bmatrix} \tilde{X}_{1,1} & \cdots & \cdots & \tilde{X}_{1,J} \\ \vdots & & \iddots & \bar{X}_{2,J} \\ \vdots & \iddots & \iddots & \vdots \\ \tilde{X}_{J,1} & \bar{X}_{J,2} & \cdots & \bar{X}_{J,J} \end{bmatrix}$$

```
totale_krav_matrise = X_tilde_matrise
for (j in 2:J){
   for (e in 1:J){
      if (e>J-j+1){
         totale_krav_matrise[j,e] = X_tilde_matrise[j,J-j+1] *
                                    prod(delta[(J-j+1):(e-1)])
      }
   }
}
```

Vi har nå funnet alle estimatorene som trengs for å regne ut likning 3.39, men for å kunne sammenlige vår utregning med den i artikkelen til England og Verrall ønsker vi også å finne ut hvor mye som må reserveres for hvert skadeår. Lager da en vektor som er J lang, der vi plasserer differansen mellom det totale estimatet, og den siste kjente observasjonen:

```
tot_est_vektor = c(rep(NA),J)
for (j in 1:J){
   tot_est_vektor[j] = totale_krav_matrise[j,J]
                     - totale_krav_matrise[j,J-j+1]
}
```

Dersom vi nå begynner med å se på det første leddet i likning 3.39 skal vi finne et uttrykk for

$$\sum_{j=2}^{J} \bar{X}_j^2 \sum_{e=J-j+1}^{D} \frac{\gamma_e^*}{(\delta_e^*)^2} \left\{ \frac{1}{\bar{X}_{j,e-1}} + \frac{1}{\sum_{q=1}^{J-e} \tilde{X}_{q,e-1}} \right\}$$

Vi starter da med å lage en vektor med lengde J. Videre lager vi en for-løkke, som for hvert skadeår j går igjennom likningen, og plasserer resultetet på plass nummer j i vektoren.

```
forventning1 = c(rep(NA,J))
forventning1[1] = 0
for (j in 2:J){
   temp = 0
   temp2 = 0
   for (e in ((J-j+1):(J-1))){
      temp = gamma[e] / delta[e]^2 * ((1/totale_krav_matrise[j,e])
           +(1 / sum(X_tilde_matrise[1:(J-e),e])))
      temp2 = temp2 + temp
   }
   forventning1[j] = totale_krav_matrise[j,J]^2 * temp2
}
```

Dersom vi nå hadde tatt kvadratroten av hvert ledd i vektoren, ville vi funnet prediksjonsfeilen for hvert skadeår.

Vi kan nå gå videre til det andre leddet i likning 3.39, som er

$$\sum_{j=2}^{J-1} \bar{X}_j \left(\sum_{k=j+1}^{J} \bar{X}_k \right) \sum_{e=J-j+1}^{D} \frac{\gamma_e^*}{(\delta_e^*)^2} \frac{1}{\sum_{q=1}^{J-e} \tilde{X}_{q,e-1}}$$

Denne utregningen gjør vi på samme måte som den forrige, vi begynner med å lage en vektor med lengde J, der vi plasserer de utregnede verdiene for hvert skadeår j:

```
forventning2 = c(rep(NA,J))
forventning2[1] = forventning2[J] = 0
for (j in 2:(J-1)){
   temp = 0
   temp2 = 0
   for (e in ((J-j+1):(J-1))){
      temp = gamma[e]/delta[e]^2 * (1/sum(X_tilde_matrise[1:(J-e),e]))
      temp2 = temp2 + temp
   }
   forventning2[j] = totale_krav_matrise[j,J] *
                     (sum(totale_krav_matrise[((j+1):J),J])) * temp2
}
```

Derom vi nå summerer disse to uttrykkene over de gyldige j-verdiene, har vi funnet det ønskede uttrykket:

```
kvadrert_feilestimat = sum(forventning1[2:J])
                     + 2*sum(forventning2[2:(J-1)])
```

For å finne det totale feilestimatet over alle skadeår tar vi roten av dette, og får at

```
feilestimat = sqrt(kvadrert_feilestimat)
```

B

Valg av c_e-vektor

Som vi nevnte i kapittel 4 har vi ikke så mye å basere valget av c_e-vektor på. I matrisen som vi tar utgangspunkt i er det kun oppgitt inkrementelle krav, og vi vet ingenting om hvor mange skader, og hvor store enkeltkrav, disse består av. Dersom vi hadde visst dette, ville det vært mulig å ta utgangspunkt i de tallene, for så å regne "bakover", og på den måten få et estimat for c_e.

Siden datasettet opprinnelig kommer fra artikkelen *Historical Loss Development Study (1991)*, av Reinsurance Association of America, sendte vi dem en mail der vi spurte om det var mulig å få tak i tallene som ligger bak matrisen. Vi fikk da følgende svar:

> *Unfortunately, that information is not available. The RAA issues a data call to collect the reserve development data from participating companies. That information is generally considered to be proprietary and confidential. Therefore, we use a third party to aggregate the responses. It is a blind study to us. We do not have access to that level of detail.*

Siden vi da kun har den oppgitte matrisen å gå ut fra, er det ikke mulig å estimere c_e. Vi valgte derfor å sette opp tre kriterier, som nevnt i kapittel 4, for så å ta valget basert på resultatene i forhold til disse kriteriene. Oppsummert var de tre kriteriene våre:

- Gjennomsnittlig kravstørrelse skal gå opp.
- Utviklingsfaktorene som er basert på hele matrisen, $\hat{\delta}_e$, skal være ganske like de estimerte utviklingsfaktorene som kun baseres på øvre trekant, δ_e^*.
- De simulerte og estimerte reservene skal være ganske like.

Vi prøvde med flere varianter av c_e-vektor, og 3 av forsøkene er inkludert her. La oss nå se på hvilke resultater vi fikk ved bruk av de ulike vektorene:

Alternativ 1:
$c_e = (0.83,\ 0.21,\ 0.07,\ 0.03,\ 0.007,\ 0.004,\ 0.002,\ 0.0004,\ 0.00025)$
Vi begynner med å se på forholdet mellom størrelsen på kravene og antall krav. Som vanlig er skadeårene nedover, og utviklingsårene bortover:

Tabell B.1: Gjennomsnittlig størrelse på kravene

	0	1	2	3	4	5	6	7	8	9
1										
2										40.4
3									37.9	35.1
4								16.8	45.8	38.5
5							10.9	16.8	48.2	39.0
6						16.0	9.4	17.1	39.6	34.9
7					5.9	15.6	10.4	15.7	38.6	38.5
8				4.3	5.6	15.6	12.9	17.1	39.9	38.2
9			2.9	3.8	5.6	13.8	11.0	16.2	46.2	31.5
10		2.9	2.7	4.2	6.0	14.0	9.3	16.4	37.2	37.5

Som vi ser går den gjennomsnittlige kravstørrelsen som ønsket oppover i utviklingsårene 1, 2, 3 og 4, mens den i utviklingsår 5 går ned igjen for så å fortsette oppover. Det er også et ganske stort sprang mellom utviklingsår 7 og 8. La oss nå se på hvordan δ_e^* ble:

Tabell B.2: δ-oversikt

Utviklings-år	δ^*	$\hat{\delta}$	Gjennomsnittlig differanse	Standardavvik differansen
1	2.999	3.038	0.039	0.3997
2	1.624	1.619	-0.005	0.0481
3	1.271	1.275	0.004	0.0363
4	1.172	1.172	0.000	0.0114
5	1.113	1.111	-0.002	0.0142
6	1.042	1.043	0.001	0.0108
7	1.033	1.033	0.000	0.0022
8	1.017	1.017	0.000	0.0062
9	1.009	1.009	0.000	0.0024

Differanse mellom $\hat{\delta}_e$ og δ_e^* er ikke så stor, og heller ikke standardavvikene er veldig store. Eneste er standardavviket for det første utviklingsåret, som muligens kunne vært litt mindre. Vi fortsetter nå med å se på reservene.

Tabell B.3: Estimerte vs simulerte reserver

Skade år	Estimerte reserver	Simulerte reserver	Differanse	Pred feil simulering	Pred feil estimering
1	–	–	–	–	–
2	154	162	8	129	206
3	617	566	−51	419	623
4	1,636	1,668	32	631	747
5	2,747	2,857	110	1,346	1,469
6	3,649	3,517	−132	1,681	2,002
7	5,435	5,403	−32	1,876	2,209
8	10,907	11,598	691	5,146	5,358
9	10,650	10,251	−399	5,794	6,333
10	16,339	18,804	2,465	28,714	24,566
Overall	52,134	54,826	2,692	31,022	26,909

De simulerte reservene ble totalt sett litt høyere enn de estimerte, og prediksjonsfeilen ble også en del større. Da dette valget av c_e-vektor ikke var helt bra under noen av de tre kriteriene, valgte vi å ikke bruke denne. La oss nå se på det neste alternativet.

Alternativ 2:
$c_e = (0.62,\ 0.21,\ 0.03,\ 0.01,\ 0.006,\ 0.001,\ 0.0009,\ 0.0008,\ 0.0006)$

Tabell B.4: Gjennomsnittlig størrelse på kravene

	0	1	2	3	4	5	6	7	8	9
1										
2										15.2
3									24.6	15.4
4								36.8	19.5	15.6
5							39.3	36.8	21.9	14.1
6						18.9	39.7	37.8	25.4	16.9
7					16.4	18.7	45.5	37.0	22.6	13.0
8				7.6	16.6	18.9	43.2	36.6	18.7	18.7
9			2.9	9.3	17.9	21.2	43.8	36.5	22.1	15.6
10		3.0	2.6	10.6	16.0	15.9	28.7	35.4	19.7	14.2

I dette tilfellet har vi en økning på gjennomsnittlig kravstørrelse, frem til utviklingsår 7. Det er imidlertid en ganske kraftig økning fra utviklingsår 3 til 4, og en kraftig nedgang fra utviklingsår 7 til 8. La oss se på hvordan det går med $\hat{\delta}_e$:

Tabell B.5: δ-oversikt

Utviklings- år	δ^*	$\hat{\delta}$	Gjennomsnittlig differanse	Standardavvik differansen
1	2.999	2.987	-0.012	0.2412
2	1.624	1.622	-0.002	0.0428
3	1.271	1.266	-0.005	0.0405
4	1.172	1.170	-0.002	0.0110
5	1.113	1.113	0.000	0.0158
6	1.042	1.041	-0.001	0.0093
7	1.033	1.033	0.000	0.0022
8	1.017	1.017	0.000	0.0062
9	1.009	1.009	0.000	0.0021

Heller ikke i dette tilfellet er det noen store differanser, og her har vi også stort sett lavere standardavvik enn ved det første alternativet for c_e-vektor. Vi fortsetter med reservene:

Tabell B.6: Estimerte vs simulerte reserver

Skade år	Estimerte reserver	Simulerte reserver	Differanse	Pred feil simulering	Pred feil estimering
1	–	–	–	–	–
2	154	151	-3	157	206
3	617	669	52	528	623
4	1,636	1,623	-13	481	747
5	2,747	2,656	-91	1,170	1,469
6	3,649	3,598	-51	1,848	2,002
7	5,435	5,309	-126	2,024	2,209
8	10,907	10,054	-853	4,819	5,358
9	10,650	10,939	289	5,916	6,333
10	16,339	14,811	$-1,528$	17,264	24,566
Overall	52,134	49,810	$-2,324$	19,850	26,909

Her ender vi ut med litt mindre reserver ved simuleringen, i forhold til estimeringen. Vi ser også at prediksjonsfeilen er lavere enn ved det første alternativet.

Alternativ 3:
$c_e = (0.76, 0.21, 0.055, 0.03, 0.015, 0.0045, 0.003, 0.001, 0.00055)$

Tabell B.7: Gjennomsnittlig størrelse på kravene

	0	1	2	3	4	5	6	7	8	9
1										
2										18.9
3									18.8	17.0
4								11.0	14.6	17.9
5							8.6	11.1	14.7	19.0
6						7.6	10.5	11.2	14.1	18.2
7					5.8	6.7	10.6	10.8	20.4	17.3
8				4.7	5.8	7.5	9.2	11.1	16.6	12.7
9			2.9	5.6	5.7	7.3	9.6	11.4	20.3	18.4
10		2.7	3.0	5.1	5.5	7.4	9.9	11.5	16.1	19.5

Her har vi den ønskede utviklingen. Den gjennomsnittlige skadestørrelsen øker jevnt etter som årene går. La oss se på hvordan δ_e^* ser ut:

Tabell B.8: δ-oversikt

Utviklings-år	δ_e^*	$\bar{\delta}_e$	Gjennomsnittlig differanse	Standardavvik differansen
1	2.999	3.001	0.002	0.2762
2	1.624	1.623	-0.001	0.0559
3	1.271	1.270	-0.001	0.0384
4	1.172	1.172	0.000	0.0140
5	1.113	1.112	-0.001	0.0163
6	1.042	1.043	0.001	0.0117
7	1.033	1.033	0.000	0.0023
8	1.017	1.017	0.000	0.0053
9	1.009	1.009	0.000	0.0025

Det ser bra ut også her. Vi ser at $\hat{\delta}_e$ holder seg ganske nært δ_e^*, samtidig som standardavviket er ganske lavt. La oss til slutt se på reservene:

Tabell B.9: Differanse mellom simulerte og estimerte gjenstående reserver

Skade-år	Estimerte reserver	Simulerte reserver	Differanse	Pred feil simulering	Pred feil estimering
1	–	–	–	–	–
2	154	178	24	151	206
3	617	659	42	421	623
4	1,636	1,603	−33	540	747
5	2,747	2,618	−129	1,199	1,469
6	3,649	3,754	105	1,841	2,002
7	5,435	5,400	−35	2,229	2,209
8	10,907	10,601	−306	4,356	5,358
9	10,650	10,786	136	6,352	6,333
10	16,339	16,225	−114	18,649	24,566
Overall	52,134	51,824	−310	20,211	26,909

De simulerte reservene avviker ikke så mye fra de estimerte, og prediksjonsfeilen ved simuleringen er lavere enn ved estimeringen.

Konklusjon: Ut fra disse tre alternativene mener vi at alternativ 3 er det beste. Den kommer bra ut ved alle tre kriteriene, og det er denne c_e-vektoren vi har valgt å bruke under simuleringen i kapittel 4.

C

Simulering i R av den sammensatte Poisson prosessen

Vi ønsker å simulere data etter de krav som ble beskrevet i kapittel 4, samt sjekke om disse passer med Mack sin modell.

Vi antar at den øvre delen av matrisen er kjent, og ønsker å simulere nedre halvdel. Når dette er gjort vil vi sammenligne resultatet med Mack sin modell. Det første som må gjøres er at programmet som er beskrevet i appendiks A må kjøres, da vi trenger input matrisen, samt δ_e^* og γ_e^*. Når dette er gjort kan simuleringen starte.

Vi begynner med å definere c_e vektoren, se diskusjon i appendiks B om valg av denne, samt finne α_e og β_e.

```
c_vektor=c(0.76, 0.21, 0.055, 0.03, 0.015, 0.0045, 0.003, 0.001, 0.00055)
alfa = (delta-1)^2 / (c_vektor*gamma - (delta-1)^2)
beta = (c_vektor*gamma-(delta-1)^2) / (c_vektor*(delta-1))
```

Hver gang simuleringen gjennomføres ønsker vi å ta vare på $\hat{\delta}_e$, som er utviklingsfaktorene beregnet på grunnlag av hele matrisen, samt de totale simulerte reservene for hvert skadeår. Vi lager derfor to matriser, der dette skal plasseres.

```
delta_matrise = matrix(rep(NA),J-1,100)
tot_sim_matrise = matrix(rep(NA),J,100)
```

Når dette er gjort kan selve simuleringen starte. Den skal gjøres 100 ganger, og vi lar derfor k gå fra 1 til 100. Det første som gjøres så er å lage en matrise der det simulerte antall skader plasseres, dette vil være en 10x10 matrise, men der kun nedre triangel vil bli simulert. Det lages også en 10x10 matrise der de summerte kravbeløpene for hvert skadeår og utviklingsår plasseres. Til slutt lages det en matrise for de akkumulerte kravene. Denne vil være lik *X_tilde_matrise* fra appendiks A, og det er nedre triangel av denne som skal fylles ut. Når dette er gjort blir antall skader og kravbeløp simulert, og plassert i sine tilhørende matriser.

```
for (k in 1:100){
ant_skader = matrix(rep(NA),J,J)
belop = matrix(rep(NA),J,J)
akk_belop = X_tilde_matrise

for (j in 2:J){
   for (e in (J-j+2):J){
      lambda = c_vektor[e-1] * akk_belop[j,(e-1)]
      ant_skader[j,e] =r pois(1,lambda)
      belop[j,e] = sum(rgamma(ant_skader[j,e],shape=alfa[e-1],
                   scale=beta[e-1]))
      akk_belop[j,e] = akk_belop[j,e-1]+belop[j,e]
   }
}
```

Vi ønsker nå å beregne hvor mye de simulerte reservene er for hvert skadeår. Dette blir plassert i matrisen *tot_sim_matrise* som vi laget tidligere. Siden all historie for første skadeår er kjent, blir reservene der satt lik 0.

```
tot_sim_matrise[1,] = 0
for (j in 2:J){
   tot_sim_matrise[j,k] = akk_belop[j,J] - akk_belop[j,J-j+1]
}
```

Nå som hele matrisen er fylt ut, kan vi estimere en $\hat{\delta}_e$, basert på alle kravene. Når dette er gjort er vi ferdig med det som skal gjøres for hver av de 100 simuleringene, og avslutter da for-løkken som ble startet i begynnelsen av programmet.

```
for (e in 2:J){
   delta_matrise[e-1,k]=sum(akk_belop[1:J,e])/sum(akk_belop[1:J,(e-1)])
}
}
```

For å kunne vurdere hvor mye δ_e^* ble påvirket av de tilnærmelser som er blitt gjort, ønsker vi å finne et gjennomsnitt av alle de 100 ulike $\hat{\delta}_e$ som ble beregnet på grunnlag av hele matrisen. Vi ønsker også å se på standardavviket for de ulike skadeårene. Det er også av interesse å finne gjennomsnittlig reserve for hvert skadeår, samt standardavviket.

```
ny_delta = c(rep(NA,J-1))
sd_delta=c(rep(NA,J-1)
mean_sim=c(rep(NA,J))
sd_sim=c(rep(NA,J))

for (i in 1:J){
   if (i<J){
```

```
        ny_delta[i] = mean(delta_matrise[i,])
        sd_delta[i] = sd(delta_matrise[i,])
        mean_sim[i] = mean(tot_sim_matrise[i,])
        sd_sim[i] = sd(tot_sim_matrise[i,]-reserve[i])
    }
    else {
        mean_sim[i] = mean(tot_sim_matrise[i,])
        sd_sim[i] = sd(tot_sim_matrise[i,]-reserve[i])
    }
}
```

Vi har nå alt vi trenger, bortsett fra den totale prediksjonsfeilen for simuleringen. Denne finner vi ved å se på standardavviket til de totale reservene mellom hver simulering.

```
temp = c(rep(0,100))
for (i in 1:100){
    temp[i] = sum(tot_sim_matrise[,i])
    overall_sd_sim = sd(temp)
}
```

Nå har vi regnet ut alt vi trenger, og det eneste som gjenstår er å sette alt sammen til to tabeller, slik at det blir lettere å få en oversikt over resultatene.

```
utv_år = seq(1:9)
delta = round(delta,3)
ny_delta = round(ny_delta,3)
mean_delta = round(ny_delta-delta,3)
sd_delta = round(sd_delta,4)
delta_oversikt = as.data.frame(cbind(utv_år, delta, ny_delta,
                 mean_delta, sd_delta))

skadeår = seq(1:11)
est_res = round(reserve,0)
est_res[1] = 0
est_res[J+1] = sum(est_res[1:J])
gj_sim_res = round(mean_sim,0)
gj_sim_res[J+1] = sum(gj_sim_res[1:J])
gj_diff = gj_sim_res-est_res
pred_feil = round(sd_sim,0)
pred_feil[J+1] = round(overall_sd_sim,0)
pred_feil_mack = round(sqrt(forventning1),0)
pred_feil_mack[1] = 0
pred_feil_mack[J+1] = round(feilestimat,0)
total_oversikt = as.data.frame(cbind(skadeår, est_res, gj_sim_res,
                 gj_diff, pred_feil, pred_feil_mack))
```

D

De Silvas program i R for utregning av MSEP

I artikkelen til England og Verrall [Engl 02] kommer de i avsnittet om overdispensert Poisson modell frem til følgende modell:

$$E(X_{j,e}) = m_{j,e} = e^{\eta_{j,e}} = e^{\alpha_j + \beta_e + c}$$

Videre ønsker vi å bruke en GLM til å estimere totale krav, samt det totale feilestimatet. De Silva har laget et program som gjør akkurat dette, og det er dette programmet vi skal se nærmere på nå. Eksempelet han har valgt å se på er det samme som vi har sett på tidligere, og er hentet fra side 22 i England og Verrall sin artikkel:

Incremental Claims Data

	Development Period									
Origin Period	1	2	3	4	5	6	7	8	9	10
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	-103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

Man ønsker å ha dataene på matriseform, så i dette eksempelet lager man da en 10×10 matrise, der man putter inn de oppgitte verdiene, samtidig som det skrives NA på de plassene uten observasjoner.

Før man kan tilpasse en GLM til disse dataene må det gjøres noen forberedelser. Dataene må være på vektorform, så det første man gjør er å putte matrisen inn i en vektor. Vektoren vil da bestå av alle kolonnene i matrisen, først første kolonne, så andre kolonne, osv. Man trenger også informasjon om hvor mange skadeår (origin years) og utviklingsår (development years) dataene man ser på består av, og dette kan man finne ved å se på lengden av kolonnene og radene.

```
claims = as.vektor(data)
n.origin = nrow(data)
n.dev = ncol(data)
```

Han fortsetter med å lage to nye vektorer, "origin" og "dev", som skal inneholde henholdsvis det skadeåret og utviklingsåret som hører sammen med verdien som står på tilsvarende plass i vektoren "claims". Dersom man ser på plass nummer 13 i vektoren "claims" vil man finne verdien som kommer fra skadeår 3, utviklingsår 2. I vektoren "origin" vil det da stå 3 på plass nummer 13, mens det i vektoren "dev" vil stå 2. Til slutt setter han disse tre vektorene sammen i en data.frame, noe som ikke er nødvendig, men det gjør at det blir lettere å se hvilket skadeår og utviklingsår de ulike kravene hører sammen med.

```
origin = factor(row = rep(1:n.origin, n.dev))
dev = factor(col = rep(1:n.dev, each=n.origin))
mack = data.frame(claims=claims, origin=origin, dev=dev)
```

Man er nå nesten klar til å tilpasse en GLM, men må først se på et lite teknisk problem. Siden man ser på en overdispensert Poisson modell, må man bruke quasipoisson familien når det skal tilpasses en GLM. I eksempelet som De Silva ser på, som er det samme som vi har sett på tidligere, har han imidlertid noen negative verdier i dataene. Som vi diskuterte i kapittel 5 tillater ikke den opprinnelige quasipoisson funksjonen i R dette. Han bruker derfor David Firth sin modifiserte versjon av quasipoisson, og da er alt klart for tilpassing av GLM:

```
model = glm(claims ~ origin + dev, family= quasipoisson(),
        subset = !is.na(claims), data=mack)
summary(model)
```

Output vil nå gi et sammendrag av modellen, med de tilpassede koeffisientene, standardavvikene deres osv. For å kunne bruke disse dataene videre må de imidlertid hentes ut fra modellen. I eksempelet som ses på her har man 10 skadeår og 10 utviklingsår. Han vil da få ut 19 koeffisienter; et skjæringspunkt c, $\alpha_2, \cdots, \alpha_{10}$ og $\beta_2, \cdots, \beta_{10}$, og disse vil nå bli plassert i en vektor som da vil ha 19 verdier. Man ønsker også å hente ut dispersjonsfaktoren, samt kovariansmatrisen til koeffisientene.

```
coef = model$coefficients
disp = summary(model)$dispersion
cov.param = disp * summary(model)$cov.unscaled
```

Hvert element i den øvre diagonale matrisen med kravene hører, som tidligere sagt, sammen med et bestemt skadeår og utviklingsår. Det er disse som har blitt brukt til å tilpasse modellen, og man kan nå bruke de tilpassede koeffisientent til å predikere krav i fremtidige utviklingsår. Man må imidlertid være sikker på at man adderer de korrekte koeffisientene. Dersom man f.eks. ønsker å predikere kravkostnaden for skadeår 2, utviklingsår 10, må man bruke formelen:

$$\mathrm{E}(X_{2,10}) = e^{\hat{\alpha}_2+\hat{\beta}_{10}+\hat{c}} = e^{F\hat{\beta}}$$

der

$\hat{\beta}$ = vektor med tilpassede koeffisienter $(\hat{c}, \hat{\alpha}_j, \hat{\beta}_e)$
F = fremtidig design matrise

Den fremtidige design matrisen brukes til å hente ut de relevante koeffisientene. I eksempelet ovenfor ville F ha en enkel rad og nitten kolonner, en for hver koeffisient. Alle verdiene vil være null, bortsett fra de som korresponderer til $\hat{c}$, $\hat{\alpha}_2$, $\hat{\beta}_{10}$. For å estimere de resterende kravene må han altså lage en fremtidig design matrise som har en rad for hvert krav som skal estimeres, og der hver rad har nitten kolonner. Dette må gjøres i flere steg. Først må han finne ut hvor mange krav som skal estimers, og så lager han en tom fremtidig design matrise med så mange rader, og like mange kolonner som vi har parametre:

```
n.fut.points = length(claims[is.na(claims)])
fut.design = matrix(0, nrow=n.fut.points, ncol=length(coef))
```

Videre må han nå finne ut hvilke krav som skal estimers. Han begynner med å lage en vektor som er lik "claims", der han setter 0 på alle kravene som er kjent (der det ikke står NA), før han nummerere de ukjente kravene fra 1 og oppover. Dersom man tenker seg at man fremdeles ser på en matrise vil nummereringen begynne i nederste venstre hjørne, og fortsette mot høyre, ovenfra og ned.

```
fut.points = claims
fut.points[!is.na(claims)] = 0
fut.points[is.na(claims)] = 1:n.fut.points
```

De Silva ønsker nå å fylle ut den fremtidige design matrisen. Hver rad tilsvarer ett krav, og i de nitten kolonnene bortover skal vi nå plassere 1-tall på de parameterene som skal brukes for å estimere kravet. Siden alle kravene skal ha med skjæringspunktet i predikeringen, begynner han med å sette et 1-tall i hele første kolonne. Videre må han finne ut hvilkte skadeår og utviklingsår hvert av kravene tilhører. Det er nå man får bruk for "origin" og "dev" vektorene som ble laget tidligere. For hvert av kravene går man inn i "fut.points", og legger merke til plasseringen. Videre går man så inn i "origin" på den samme plasseringen, og ser hvilket tall som står der. Dette vil være et tall mellom 1 og 10, og vil representere skadeåret. Dette tallet brukes videre til å bestemme hvilken parameter i fremtidig design matrisen som skal få et 1-tall. Står det 1 i "origin" hører kravet til første skadeår, og der er alle krav antatt kjente, noe som medfører at ingen parametre vil få tallet 1. Dersom det står 2 tilhører kravet skadeår 2, og det er parameteren α_2 i kolonne 2 som får et 1-tall. Etter at dette er gjort for alle kravene er man “ferdig” med kolonne 2 til 10 i “fut.design”, som representerer α parametrene. Da må man gå videre til “dev” vektoren, og gjøre akkurat det samme. Her vil tallet som står i “dev” representere utviklingsåret, og det er kolonne 11 til 19 i “fut.design” som representerer β_2 til β_{10}, og som kan få 1-tall. Når dette er gjort for alle kravene vil hver av radene ha ett, to eller tre 1-tall, fordelt på de nitten kolonnene.

```
for (p in 1:n.fut.points){
   fut.design[p,1] = 1¨
   fut.design[p, 1 + as.numeric(origin[match(p, fut.points)]) - 1] = 1
   fut.design[p, 1 + (n.origin-1) +
      as.numeric(dev[match(p, fut.points)]) - 1] = 1
}
```

Man kan nå predikere hvert av de ukjente kravene, og plasserer disse i en diagonalmatrise. Når man summerer denne vil man få forventet totale krav.

```
fitted.values = diag(as.vector(exp(fut.design %*% coef)))
total.reserve = sum(fitted.values)
```

Grunnen til at han puttet kravene i en diagonalmatrise er for å gjøre den kommende beregningen av varians enklere. I artikkelen til England og Verrall [Engl 02] kom de frem til følgende uttrykk for “Mean Squared Error of Prediction” for hele reserven, som også er det samme uttrykket som vi har regnet oss frem til i likning 5.10:

$$
\begin{aligned}
\mathrm{MSEP}[\hat{X}_{++}] &= \sum_{j,e\in\Delta} \phi \bar{m}_{j,e} + \sum_{j,e\in\Delta} \bar{m}_{j,e}^2 \mathrm{Var}\left[\bar{\eta}_{j,e}\right] + \sum_{\substack{j_1,e_1\in\Delta \\ j_2,e_2\in\Delta \\ j_1,e_1\neq j_2,e_2}} \bar{m}_{j_1,e_1}\bar{m}_{j_2,e_2}\mathrm{Cov}\left(\bar{\eta}_{j_1,e_1}, \bar{\eta}_{j_2,e_2}\right) \\
&= \sum_{j,e\in\Delta} \phi \bar{m}_{j,e} + \sum_{\substack{j_1,e_1\in\Delta \\ j_2,e_2\in\Delta}} \bar{m}_{j_1,e_1}\bar{m}_{j_2,e_2}\mathrm{Cov}\left(\bar{\eta}_{j_1,e_1}, \bar{\eta}_{j_2,e_2}\right)
\end{aligned}
\tag{D.1}
$$

For å kunne gjøre denne utregningen må han først finne kovariansmatrisen til de lineære parametrene, før han bruker denne til å finne kovariansmatrisen til de ukjente kravene. Når dette er gjort har man alt man trenger for å regne ut likningen, som man så tar kvadratroten av for å finne det totale feilestimatet:

```
cov.pred = fut.design %*% cov.param %*% t(fut.design)
cov.fitted = fitted.values %*% cov.pred %*% fitted.values
total.rmse = sqrt(disp * total.reserve + sum(cov.fitted))
total.predictionerror = round(100 * total.rmse / total.reserve)
```

Siden vi også er interessert i forventede reserver og prediksjonsfeil fordelt på skadeår, har vi laget en ekstra programmeringsdel, som er gjengitt her. Vi begynner med å hente ut de estimerte kravene, og plasserer de i en vektor. Når dette er gjort ønsker vi å plassere kravene på riktig plass i “data” matrisen. Vi vet at kravene er estimert i denne rekkefølgen:

```
                   37
                29 38
             22 30 39
          16 23 31 40
       11 17 24 32 41
     7 12 18 25 33 42
   4  8 13 19 26 34 43
 2 5  9 14 20 27 35 44
1 3 6 10 15 21 28 36 45
```

Plasserer så kravene i "data" matrisen, og lager en matrise med de akkumulerte beløpene.

```
krav=as.vector(exp(fut.design %*% coef))
temp=1
for (e in 2:10){
   for (j in (10-e+2):10){
      data[j,e]=krav[tt]
      temp=temp+1
   }
}
totale_krav_matrise = matrix(rep(NA),J,J)
for (i in 1:J){
   if (i==1){totale_krav_matrise[,i] = data[,i]}
   else totale_krav_matrise[,i] = totale_krav_matrise[,i-1]+data[,i]
}
```

Henter ut de estimerte kravene for hvert skadeår, og plasserer dem i en vektor.

```
tot_est_vektor = c(rep(NA),J)
for (j in 1:J){
   tot_est_vektor[j] = totale_krav_matrise[j,J]
                     - totale_krav_matrise[j,J-j+1]
}
```

Vi ønsker nå å finne prediksjonsfeilen for hvert skadeår. Lager først en vektor der disse kan plasseres. Videre lager vi 9 vektorer, en for hvert skadeår der krav er blitt estimert, og i denne vektoren skriver vi inn hvilke krav som hører til i dette skadeåret. Disse vektorene plasserer vi så i en matrise, der hvert skadeår har sin egen kolonne.

```
prediksjonsfeil = c(rep(0,10))
subset = matrix(0,10,10)
subset[1,2] = c(37)
subset[1:2,3] = c(29,38)
subset[1:3,4] = c(22,30,39)
subset[1:4,5] = c(16,23,31,40)
subset[1:5,6] = c(11,17,24,32,41)
```

```
subset[1:6,7] = c(7,12,18,25,33,42)
subset[1:7,8] = c(4,8,13,19,26,34,43)
subset[1:8,9] = c(2,5,9,14,20,27,35,44)
subset[1:9,10] = c(1,3,6,10,15,21,28,36,45)
```

Det vi ønsker er å finne prediksjonsfeil for hvert enkelt skadeår. Ved å sette alle de kravene som ikke er med i det aktuelle skadeåret lik 0, kan det andre leddet i likning D.1 summeres på samme måte som tidligere, da det kun er de kravene som skal være med, som vil gi bidrag. Vi lager derfor en 45×10 matrise, der hvert skadeår har sin kolonne. Vi setter så 0 på alle kravene som ikke hører hjemme i dette skadeåret. Når dette er gjort lager vi en ny løkke, som for hvert av skadeårene beregner prediksjonsfeilen:

```
mod_matrise = matrix(0,length(krav),10)
for (j in 2:10){
   mod_matrise[subset[,j],j] = krav[subset[,j]]
}

for(j in 2:10){
   fitted.values = diag(as.vector(mod_matrise[,j]))
   cov.fitted = fitted.values%*% cov.pred %*% fitted.values
   prediksjonsfeil[j] = sqrt(disp*tot_est_vektor[j]+sum(cov.fitted))
}
```

Vi har nå funnet de forventede kravene for hvert skadeår, vektoren "tot_est_vektor", og prediksjonsfeilen for hvert skadeår, vektoren "prediksjonsfeil".