



Stockholms
universitet

Skadebalansernas effekt på reserv- och kassa- flödesmodellering för kortsvansad sakförsäkring

Viktor Ellerud

Masteruppsats 2025:12
Försäkringsmatematik
Juni 2025

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm



Skadebalansernas effekt på reserv- och kassaflödesmodellering för kortsvansad sakförsäkring

Viktor Ellerud*

Juni 2025

Sammanfattning

Denna masteruppsats syftar till att undersöka hur variationer i skadebalanser, det vill säga antalet ohanterade skador hos skadereglerare, påverkar reservsättning och kassaflödesmodellering inom kortsvansad sakförsäkring. Studien baseras på data från Länsförsäkringar AB som innehåller information om behandlade skador och backlognivåer. Däremot saknas uppgifter om antalet rapporterade skador, därför härleds rapporterade skador. Därefter struktureras de skattade rapporterade skadorna i triangelformat genom att lösa ett optimeringsproblem, vilket möjliggör en separat analys av rapporterade skador.

Anledningen till detta är att chain ladder antagandet om stabil skadestegs utveckling över tid inte uppfylls för behandlade skador, eftersom dessa påverkas av behandlingskapacitet och backlognivåer. Det leder till systematiska fel i reservberäkningar vid varierande backlog. Rapporterade skador styrs däremot främst av försäkringstagarnas skadeanmälningar, vilka förväntas vara mer stabila över tid och därmed bättre uppfylla modellens antaganden. Därför undersöks om chain ladder ger mer tillförlitliga skattningar när den tillämpas på de härledda skattade rapporterade skadorna.

Utöver denna analys utvecklas modeller för att förbättra prediktionen av utvecklingsfaktorer baserade på behandlade skador. Olika GLM modeller används för att skatta utvecklingsfaktorer utifrån backlognivåer. Modellvarianter utvecklas med diagonaleffekt, radeffekt samt kombinerade modeller som även inkluderar förändringar i backlognivå. Som jämförelse implementeras en random forest modell med motsvarande struktur. Samtliga modeller utvärderas både med avseende på modellanpassning och prediktiv förmåga. Samtliga GLM modellvarianter har mycket hög träffsäkerhet i prediktionerna. Random forest presterar generellt något sämre.

Vidare presenteras en metod för att prediktera framtida kassaflöden baserat på en given backlognivå vid bokslutstidpunkten. Denna metod bidrar till en mer effektiv, repeterbar och tillförlitlig bokslutsprocess.

Slutsatsen är att den kombinerade GLM modellvarianten bör användas för att skatta utvecklingsfaktorer baserade på behandlade skador samt att Mack chain ladder tillämpat på rapporterade skador kan utföra ett lämpligt alternativ.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: Viktor@ellerud.se. Handledare: Filip Lindskog.

Abstract

This master's thesis aims to examine how variations in backlog level, defined as the number of unprocessed claims, affect reserving and cash flow modelling in short-tail non-life insurance. The analysis is based on data from Länsförsäkringar, which includes monthly information on processed claims, backlog levels and portfolio exposure. However, the dataset lacks information on the number of reported claims. To address this, reported claims are derived and estimated using information on backlog levels and processed claims. The number of reported claims is then estimated in triangle format by solving an optimization problem. This estimation enables a separate analysis of reported claims.

This approach is motivated by the fact that the chain ladder assumption of stable claim development over accident months does not hold for processed claims that are influenced by processing capacity and backlog levels. This results in systematic errors in reserve estimates when backlogs fluctuate. Reported claims, on the other hand, are primarily driven by policyholders reporting behaviour, which is expected to be more stable over time and therefore more likely to satisfy the model assumptions. The thesis therefore investigates whether applying the chain ladder method to the derived reported claims provides more reliable estimations.

Additionally, models are developed to improve the prediction of development factors based on processed claims. Generalized linear models (GLM) with log link function are used to estimate development factors as a function of backlog levels. GLM model variants with diagonal and row effects are developed, along with combined models that also include changes in backlog levels, both in absolute and relative terms. A random forest model with a corresponding structure is implemented for comparison.

All models are evaluated regarding model fit and predictive performance. The claim count reserve is also calculated and compared for each model variant. For processed claims, all GLM model variants show very high accuracy. The random forest model generally performs slightly worse.

Furthermore, a method is presented to predict future cash flows based on a given backlog level. The method is based on fitted GLM row effect models, one for each development month. This contributes to a more efficient, repeatable and reliable financial reporting process.

The conclusion is that the combined GLM model variant should be used to estimate development factors based on processed claims. Applying the chain ladder method to reported claims may also be a suitable alternative.

Keywords: Reserving, Backlog, Mack chain ladder model, Generalized linear models (GLM), Random forest, Cash flow modelling, Estimation of reported claims, Development factors, Short-tail non-life insurance

Förord

Denna uppsats utgör mitt masterarbete i försäkringsmatematik vid Stockholms universitet och omfattar 30 högskolepoäng.

Jag vill börja med att rikta ett varmt tack till min externa handledare Melker Wallén vid Länsförsäkringar AB för hans inspirerande tankar, engagemang och stöd under arbetets gång. Ett stort tack riktas även till min handledare Filip Lindskog vid Stockholms universitet för hans värdefulla idéer och vägledning. Jag vill dessutom tacka Länsförsäkringar AB för att ha tillhandahållit det datamaterial som ligger till grund för denna uppsats.

Sist men inte minst vill jag uttrycka min tacksamhet till min familj, som genom hela livet har varit ett ovärderligt stöd.

AI användning

I denna uppsats har ChatGPT använts för att generera LaTeX tabeller baserade på tabeller från R . AI har även använts som stöd vid visuella förbättringar av figurer, exempelvis genom val av färger och stilar, samt vid felsökning av kod.

Innehållsförteckning

1	Introduktion	1
1.1	Bakgrund	1
1.2	Problematik	2
1.3	Syfte	3
1.4	Förväntade resultat	3
1.5	Översikt	3
2	Data	4
3	Teori	5
3.1	Reservsättning	5
3.2	Skadebalanser	7
3.3	Mack chain ladder	8
3.4	Generaliserade linjära modeller (GLM)	10
3.5	Random forest	11
4	Resultat	13
4.1	Modeller	14
4.1.1	Chain ladder metoden	15
4.1.2	GLM modell med diagonaleffekt	16
4.1.3	GLM modell med radeffekt	18
4.1.4	GLM modell med kombineradeffekt	20
4.1.5	Random forest kombinerad modell med backlogförändring i antal	24
4.2	Val av antalet utvecklingsfaktorer (n) för GLM och random forest modellering	26
4.3	Jämförelse av modellvarianter	31
4.4	Jämförelse av modellvarianternas prediktiva förmåga	36
4.5	Antalsreserv	40
4.6	Kassaflöden för framtida perioder	42
5	Skattning av antalet rapporterade skador	49
5.1	Skattning av rapporterade skador genom optimering	49
5.2	Chain ladder baserat på skattade rapporterade skador	52
6	Diskussion	57
7	Slutsatser	60
	Referenser	62

1 Introduktion

I detta avsnitt ges en grundläggande förståelse för ämnet. Först presenteras bakgrunden till försäkringsbolagens reservsättning och de utmaningar som finns inom området. Därefter beskrivs problematiken med variationer i skadebalanser och hur dessa påverkar reservsättningsmodeller som chain ladder. Syftet med rapporten klargörs, följt av en redogörelse för de förväntade resultaten. Slutligen ges en översikt över rapportens struktur och de olika kapitlen som följer.

1.1 Bakgrund

En försäkringsverksamhets grundprincip är att intäkterna från premier under en viss tidsperiod ska täcka de skadekostnader som uppstår under samma period. Reserven definieras som det belopp som avsätts för att täcka skadekostnader som inträffat under perioden, men ännu inte är fullt reglerade.

En av utmaningarna för försäkringsbolag är att det vid periodens slut kan vara svårt att exakt uppskatta de totala skadekostnaderna. Detta beror dels på att många skador fortfarande kan vara orapporterade vid periodens slut och dels på bristen på detaljerad information om kostnaderna för redan rapporterade skador.

För att skatta reserven finns ett flertal olika metoder som kan användas. Den vanligaste är chain ladder, en metod som antar att framtida utbetalningar följer samma mönster som de historiska. GLM representationen av chain ladder är en vidareutveckling av metoden och används också för reservsättning. Chain ladder fungerar dock sämre vid strukturella förändringar i utbetalningsmönster, till exempel på grund av säsongsvariationer. Dessutom är metoden mindre tillförlitlig i inledningen av en utvecklingsmånad, när datamängden är begränsad. Utöver chain ladder finns flera metoder, såsom Bornhuetter-Ferguson metoden, som kombinerar historiska data med a priori uppskattning för att skatta det slutliga skadebeloppet. En ytterligare metod är Benktander metoden, som utgör en kombination av chain ladder och Bornhuetter-Ferguson.

För försäkringsbolag är korrekt reservsättningen en viktig del för att säkerställa finansiell stabilitet och uppfylla både interna och externa krav. För det första, om reserven sätts för lågt, kan det leda till en brist på kapital som är nödvändigt för att täcka framtida utbetalningar för skador som redan har inträffat. Dessutom kan en felaktig reserv påverka premierna inom försäkringsverksamheten, eftersom de ofta beräknas utifrån den totala skadekostnaden. Detta kan resultera i en under- eller överskattning av framtida skadekostnader som ännu inte har inträffat.

Reservsättning spelar en avgörande roll för stabiliteten inom försäkringssektorn. Under finanskrisen blev det tydligt hur viktigt det är för försäkringsbolag och framförallt för banker att ha tillräckliga resurser för att kunna hantera oväntade skador och utbetalningar. Många företag stod inför svårigheter när de inte

hade tillräckliga resurser för att täcka sina åtaganden, vilket ledde till en ökad oro för deras finansiella stabilitet och långsiktiga överlevnad. Finanskrisen underströk vikten av att försäkringsbolagen avsätter tillräckliga reserver för att klara av plötsliga förändringar och oväntade skador.

I syfte att stärka sektorns motståndskraft och säkerställa att försäkringsbolag kan möta sina framtida förpliktelser, infördes Solvens II regelverket. Detta regelverk ställer krav på att försäkringsbolag ska ha tillräckliga reserver och kapital för att uppfylla sina åtaganden gentemot försäkringstagarna för både förväntade och oväntade händelser. Solvens II regelverket bygger på en helhetssyn av riskerna, där alla risker försäkringsbolaget utsätts för tas hänsyn till. Detta innebär att bolagen måste ha en djupare förståelse för sina finansiella tillgångar och hur dessa påverkar både deras risknivå och de försäkringsrisker de är exponerade för.

Genom att säkerställa att försäkringsbolagen har lämpliga reservnivåer och tillräckligt med kapital kan Solvens II bidra till att förbättra stabiliteten inom försäkringsbranschen och förhindra att företag hamnar i ekonomiska svårigheter vid plötsliga förändringar. Detta regelverk främjar en mer robust och hållbar försäkringssektor, vilket i sin tur stärker förtroendet för branschen som helhet. Det understryker vikten av korrekt reservsättning.

Inom sakförsäkring kan skador delas in i två olika typer, kortsvansade och långsvansade. Kortsvansade produkter, såsom djurförsäkring, präglas av att skador vanligtvis hanteras och avslutas på en relativt kort tid. Å andra sidan omfattar långsvansade produkter, som personförsäkring, de ärenden där det kan dröja betydligt längre från det att skadan rapporteras tills dess att den är helt slutreglerad.

1.2 Problematik

Inom kortsvansad sakförsäkring, såsom djur- och sjukvårdsförsäkring, är korrekt reservsättning och kassaflödesmodellering avgörande för att säkerställa finansiell stabilitet och lönsamhet. Skadebalanser, definierade som antalet ohanterade skador hos skadereglerarna, spelar en central roll i dessa processer. Variationer i dessa kan påverka mönstren i reservsättningsmodeller som chain ladder och därmed påverka finansiella resultat och kassaflödesprognoser. Chain ladder metoden är en av de mest använda teknikerna för att skatta framtida skadekostnader baserat på historiska data. Metoden bygger på att skador rapporteras och betalas ut över tid i ett förutsägbart mönster. När skadebalanserna varierar, kan detta mönster störas. Detta kan leda till felaktiga uppskattningar av framtida skadekostnader och därmed påverka både reservsättningen och kassaflödesprognoserna negativt. En korrekt hantering av skadebalansernas påverkan på reserver och kassaflöden är därför avgörande för att förbättra noggrannheten i finansiella prognoser, vilket i sin tur stärker beslutsunderlaget inom företaget.

1.3 Syfte

Syftet med examensarbetet är att undersöka och kvantifiera effekten av variationer i skadebalanser på reserv- och kassaflödesmodellering inom kortsvansad sakförsäkring. Genom att utveckla och tillämpa avancerade statistiska modeller så strävar arbetet mot att förbättra noggrannheten i dessa processer. Mer konkret är syftet att skapa en förståelse för hur variationer i antalet ohanterade skador påverkar reservsättningsmetoder som chain ladder, vilket i sin tur påverkar framtida kassaflöden. Vidare är avsikten att utveckla olika modeller för reservsättning och utvärdera hur de hanterar situationer med varierande skadebalanser. Slutligen, ta fram en metod för att snabbt kunna ta fram tillförlitliga prognoser över framtida kassaflöde baserat på skadebalansnivån vid månadens slut.

1.4 Förväntade resultat

Det förväntas att detta examensarbete kommer att leda till utveckling av mer precisa och robusta modeller för reservsättning och kassaflödesprognoser inom kortsvansad sakförsäkring under varierande skadebalansnivåer. Det väntas därmed förbättra försäkringsbolagets förmåga att avsätta tillräckliga reserver för framtida skadekostnader.

1.5 Översikt

Rapporten inleds med en introduktion där bakgrund, problemformulering och syfte presenteras. Därefter följer en beskrivning av det datamaterial som ligger till grund för analysen, samt ett teorikapitel där relevanta modeller och begrepp förklaras. Huvuddelen av rapporten består av resultat från modellutveckling och jämförelser, följt av ett avsnitt om skattning av rapporterade skador. Rapporten avslutas med en diskussion av resultaten samt slutsatser.

2 Data

Denna rapport baseras på datamaterial som tillhandahållits av Länsförsäkringar AB. Datamaterialet innehåller information om antalet skador, backlognivåer och exponeringsmått för kortsvansade sakförsäkringsprodukter.

Vi har tillgång till information om skadebalanserna (backlog) månadsvis. Skadebalanserna är definierade på så vis att den inte endast innehåller obehandlade skador, utan skadebalanserna kan även innehålla ärenden som inte är skador, det vill säga nollskador. Skadebalansen är specificerad så att en skada som ligger i skadebalansen ej är behandlad, vilket innebär att när skadan behandlas av en skadereglerare lämnar skadan skadebalansen.

Antalet behandlade skador definieras i datamaterialet som de skador där skadereglerare har påbörjat handläggning av, de behöver alltså inte vara färdigreglerade. Eftersom data hämtas från kortsvansade produkter är dock skillnaden mellan antalet behandlade och antalet betalda (färdigreglerade) skador generellt mycket liten. Data över behandlade skador är månadsvisa och utgör grunden för analysen i denna rapport. Det är viktigt att notera att skadebalanserna och antalet behandlade skador är definierade på olika sätt. Skadebalanserna omfattar både skador och andra typer av ärenden, medan antalet behandlade skador enbart avser faktiska skador, det vill säga skador med ett positivt skadebelopp (icke-nollskador).

För att kvantifiera portföljens storlek och dess förändringar används exponeringsdata i form av duration. Durationen avser den sammanlagda aktiva tiden för samtliga kontrakt vid månadens slut, vilket innebär att kontrakten i portföljen kan vara aktiva under delar av månaden. Detta exponeringsmått syftar till att kvantifiera portföljens storlek över tid och möjliggör analys av dess förändringar. Denna data är presenterad månadsvis.

3 Teori

I detta avsnitt ges en genomgång av reservsättning och dess betydelse för försäkringsbolagens finansiella stabilitet. Reservsättning definieras, och det förklaras hur inkrementella och kumulativa skadetrianglar används för att strukturera historisk data. Vidare presenteras definitionen av skadebalanser samt deras inverkan på reservskattningar. Därefter redogörs för de huvudsakliga reservkategorierna IBNR och RBNS samt för centrala modeller såsom Mack chain ladder, GLM och random forest, inklusive respektive metods nyckelantaganden.

3.1 Reservsättning

Försäkringsbranschen är uppbyggd på så sätt att intäkterna, i form av premier från kunder, samlas in innan de verkliga kostnaderna för skador kan fastställas. Detta innebär att försäkringsbolag får betalt av kunder för försäkringsskydd innan de har möjlighet att veta hur stora utbetalningarna för skador kommer att bli. Denna ordning skiljer sig från tillverkningsindustrin, där produktionskostnaderna är kända innan produkten säljs och priset kan sättas utifrån det.

För att hantera osäkerheten och upprätthålla ekonomisk stabilitet behöver försäkringsbolag noggrant skatta och avsätta tillräckliga reserver. Vid en given tidpunkt, görs beräkningar av bolagets förpliktelser som är kopplade till tidigare sålda försäkringsavtal. Dessa förpliktelser kallas skadereserver och representerar det kapital som krävs för att täcka skador som ännu inte är färdigreglerade. Det omfattar både skador som redan har rapporterats men ännu inte är slutreglerade och skador som har inträffat men ännu inte har rapporterats till försäkringsbolaget.

Inom reservsättning är data vanligtvis representerad i triangelform, vilket brukar benämnas skadetriangel. Den kan innehålla antingen inkrementella eller kumulativa skadekostnader. Beteckningen som använts för att beskriva inkrementella skadeutbetalningar är $X_{i,j}$. Vi betraktar $X_{i,j}$ som den skadeutbetalning som görs för skador som inträffat under skademånad i och utvecklingsmånad j . Varje rad i representerar en skademånad, medan varje kolumn j representerar en utvecklingsmånad. Vi antar att det finns ett maximalt antal skademånader I och att vi beaktar utvecklingsmånader upp till J . Figuren nedan illustrerar den inkrementella skadetriangeln. Den kumulativa skadetriangeln är uppbyggd på samma sätt, men med de kumulativa skadekostnaderna $C_{i,j}$. Den kumulativa skadekostnaden är summan av de inkrementella skadekostnaderna, vilket ges av

$$C_{i,j} = \sum_{l=0}^j X_{i,l}.$$

Tabell 1: Inkrementell skadetriangel med skadeutbetalningar $X_{i,j}$, där varje rad motsvarar en skademånad och varje kolumn en utvecklingsmånad. Den övre triangeln, ovanför diagonalen, innehåller observerade värden. Den undre delen består av ännu okända utbetalningar som behöver predikteras. Det antas att $J \leq I$.

Skademånad i	Utvecklingsmånad j					
	1	2	...	j	...	J
1	$X_{1,1}$	$X_{1,2}$	...	$X_{1,j}$	...	$X_{1,J}$
$\vdots$	$\vdots$	$\vdots$				$\vdots$
$I - J$	$X_{I-J,1}$	$X_{I-J,2}$	...			$X_{I-J,J}$
$\vdots$						
i			Observationer			
$\vdots$						
$I - 1$			Skadekostnader att prediktera			
I	$X_{I,1}$	$X_{I,2}$	...			$X_{I,J}$

Av tabell 1 framgår att den övre vänstra triangeln innehåller de kända inkrementella skadeutbetalningarna, medan den nedre högra triangeln innehåller de okända inkrementella skadekostnaderna. Vi antar att de kända inkrementella skadeutbetalningarna $X_{i,j}$ motsvarar observationer då $i + j - 1 \leq I$ samt okända skadekostnader då $i + j - 1 > I$. Grunden i reservsättning är att man vill estimerar den undre triangeln, de okända skadeutbetalningarna. Detta med utgångspunkt från de kända observationerna i den övre triangeln. Denna ansats bygger på antagandet att skador utvecklas enligt ett stabilt mönster över tid, vilket gör det möjligt att använda historiska utbetalningsdata för att prediktera framtida kostnader. [16, ss. 96–98]

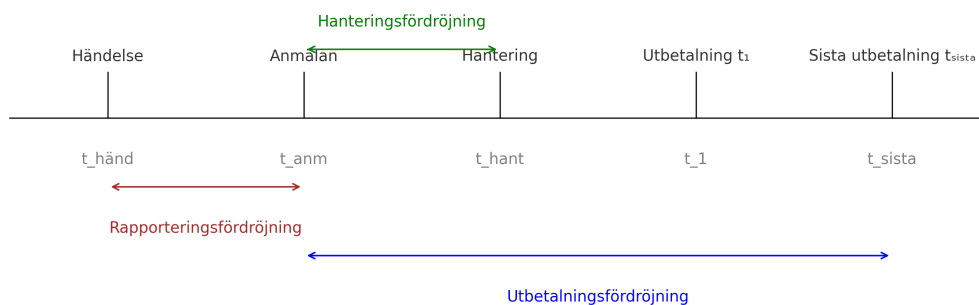
Reserverna delas vanligtvis in i två kategorier, IBNR och RBNS. IBNR, som står för incurred but not reported, syftar på skador som har inträffat men ännu inte rapporterats. Det vill säga sådana skador som försäkringsbolaget saknar kännedom om. RBNS, reported but not settled, avser skador som redan har rapporterats men att samtliga utbetalningar ännu inte är utförda, inte färdigreglerade.

Det finns flera olika metoder för att skatta dessa reserver, den vanligaste metoden är chain ladder. Till skillnad från vissa andra metoder beräknar Chain Ladder den totala reserven som en helhet, utan att särskilja mellan IBNR och RBNS. [3, ss.181–182]

3.2 Skadebalanser

När försäkringshändelser inträffar har försäkringstagaren rätt till ersättning från försäkringsbolaget. De faktiska kostnaderna för sådana skador är dock ofta okända för försäkringsbolaget direkt vid händelsen. Det är vanligt att det dröjer innan skador rapporteras in, vilket är en naturlig del av skadeutvecklingscykeln. Även när en skada väl rapporteras, innebär det inte nödvändigtvis att skadeärendet avslutas snabbt. Detta eftersom handläggningen av skadeärenden ofta tar tid. Dessutom kan det ta lång tid till skadan är färdigreglerad (samtliga utbetalningar till försäkringstagaren är gjorda).

Detta förlopp illustreras i figur 1, där en händelse inträffar vid tidpunkten $t_{\text{händ}}$. Händelsen anmäls till försäkringsbolaget vid tidpunkten t_{anm} , efter en viss rapporteringsfördröjning. Skadeärenden som befinner sig mellan händelsetidpunkten och anmälningstidpunkten klassificeras som IBNR (Incurred But Not Reported), eftersom skadan har inträffat men ännu inte anmälts till försäkringsbolaget, se avsnitt 3.1. För dessa skador finns ingen tillgänglig information vid det aktuella tillfället. Tidsintervallet mellan anmälan och tidpunkten t_{hant} motsvarar hanteringsfördröjningen, det vill säga den tid det tar för skadereglerarna att påbörja behandlingen av skadeärendet. Denna fördröjning är nära kopplad till skadebalanserna, en längre hanteringstid indikerar generellt en högre skadebalansnivå, medan en kortare hanteringstid tyder på en lägre skadebalansnivå. Efter hanteringsfördröjningen kan utbetalningar påbörjas. I figuren sker en första utbetalning vid tidpunkten t_1 och en sista vid t_{sista} , då skadeprocessen avslutas. Tiden mellan anmälan och sista utbetalning definieras som utbetalningsfördröjningen, vilket på engelska ofta benämns som settlement delay. Skador inom detta intervall tillhör kategorin RBNS (Reported But Not Settled), eftersom de är anmälda men ännu inte färdigreglerade. Efter tidpunkten t_{sista} betraktas skadan som avslutad, och inga ytterligare utbetalningar förväntas ske under antagandet att skador inte kan återöppnas.



Figur 1: Illustration av skadans utvecklingscykel och fördröjningar inom skadereglering.

Hur snabbt ett skadeärende kan hanteras efter att det har rapporterats beror på om det omedelbart kan behandlas eller måste vänta på grund av kapacitetsbegränsningar. Fördröjningar uppstår när det inte finns tillräcklig kapacitet för att hantera alla inkommande ärenden. De ohanterade skadeärenden benämns som skadebalans, eller backlog på engelska. Skadebalansen definieras här som antalet ärenden som ännu inte har hanterats av skadereglerare. Att sådana balanser kan byggas upp beror på att försäkringsbolag arbetar med begränsade resurser, till exempel finns ett visst antal anställda skadereglerare, eftersom det medför kostnader att öka kapaciteten. [11, ss. 1–2] Skadebalanser tenderar därför att tillfälligt öka under perioder på grund av variationer i kapaciteten och antalet rapporterade skador. Detta kan orsakas av exempelvis säsongsvariationer eller hög sjukfrånvaro. Omvänt kan skadebalansen minska under perioder när inflödet av skadeanmälningar är lägre än kapaciteten. [2, ss. 444–445]

3.3 Mack chain ladder

Chain ladder metoden är en av de mest använda och välkända metoderna inom reservsättning. Modellen bygger på historisk data och används för att skatta framtida reserver genom att identifiera mönster i hur skador utvecklas över tid. Metoden är relativt enkel att tillämpa men vilar på ett antal viktiga modellantaganden som är avgörande för dess tillförlitlighet.

Mack chain ladder modellen är en stokastisk vidareutveckling av den klassiska chain ladder metoden. Till skillnad från den deterministiska versionen möjliggör Mack chain ladder kvantifiering av osäkerheten i skattningarna utan att förutsätta en specifik fördelning för skadekostnaderna. Mack chain ladder är fördelningsfri,

det vill säga den kräver inga antaganden om någon specifik fördelning för skadebeloppen.

Modellen bygger på att data är på triangelformat med kumulativa skadekostnader $C_{i,j}$, där i betecknar skademånad och j utvecklingsmånad. Där observationerna $C_{i,j}$ är kända då $i + j - 1 \leq I$. Målet är att skatta ultimo skadekostnaderna $C_{i,J}$ samt reserverna.

De strukturella antaganden som modellen grundar sig på att det för varje utvecklingsmånad j finns en deterministisk utvecklingsfaktor f_j sådan att det betingade väntevärdet för $C_{i,j+1}$ ges av skadekostnaderna i utvecklingsmånad $j + 1$ är proportionellt mot föregående utvecklingsmånads skadekostnader

$$\mathbb{E}[C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}] = C_{i,j} \cdot f_j.$$

Vidare antas att den betingade variansen för $C_{i,j+1}$ är proportionell mot skadekostnaden i utvecklingsmånad j

$$\text{Var}(C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}) = \sigma_j^2 \cdot C_{i,j},$$

där σ_j^2 är variansparametern för utvecklingsmånad j .

Utöver dessa strukturella antaganden görs följande modellantaganden:

- **Stabila utvecklingsmönster:** Modellen antar att utvecklingsfaktorerna f_j är konstanta över skademånader. Stabila utvecklingsmönster innebär att skadekostnaden över tid utvecklas på liknande sätt mellan olika skademånader.
- **Oberoende mellan skademånader:** Utvecklingen för olika skademånader antas vara oberoende. Det vill säga

$$\{C_{i,1}, \dots, C_{i,I+1-i}\} \perp \{C_{k,1}, \dots, C_{k,I+1-k}\}, \quad \text{för } i \neq k.$$

Utvecklingsfaktorerna $\hat{f}_j$ skattas som kvoten mellan de totala skadeutbetalningarna i utvecklingsmånad $j + 1$ och de i utvecklingsmånad j , enligt

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{i,j}}.$$

Dessa utvecklingsfaktorer används för att skatta framtida utbetalningar i den undre triangeln. Ultimobeloppet, det vill säga den totala förväntade skadekostnaden för skademånad i . Beräknas genom att multiplicera det senast observerade beloppet med de relevanta utvecklingsfaktorerna

$$\hat{C}_{i,J} = C_{i,I+1-i} \cdot \hat{f}_{I+1-i} \cdots \hat{f}_{J-1}.$$

Därefter kan den skattade reserven beräknas som skillnaden mellan ultimobeloppet och den senast observerade kumulativa skadekostnaden för skademånad i

$$\hat{R}_i = \hat{C}_{i,J} - C_{i,I+1-i}.$$

Mack chain ladder modellen möjliggör dessutom beräkning av osäkerhet i form av medelkvadratfel (mean squared error) för varje skademånadsreserv (se [12, ss. 216–218]). Variansparametrarna skattas enligt

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=1}^{I-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, \quad 1 \leq j \leq I-2.$$

En av de stora fördelarna med chain ladder metoden är dess enkelhet och tolkbarhet. Metoden är effektiv när det finns en stor mängd historisk data som kan användas för att förutsäga framtida skadekostnader. Nackdelen med metoden är att den inte tar hänsyn till om historiska mönster förändras, vilket kan leda till missvisande skattningar. [12, ss. 214–217]

3.4 Generaliserade linjära modeller (GLM)

Generaliserade linjära modeller (GLM) utgör en utvidgning av klassiska linjära regressionsmodeller och möjliggör modellering av responsvariabler som inte är normalfördelade. GLM bygger på tre huvudsakliga komponenter, en slumpmässig komponent (responsvariabel), en systematisk komponent (förklarande variabler) och en länkfunktion. Ett grundläggande antagande inom den generella teorin för GLM är att responsvariablerna $Y_1 \dots Y_n$ antas vara oberoende. En viktig klass av fördelningar inom denna teori är exponential dispersion models (EDM), vilka generaliserar flera vanliga fördelningar, såsom normal-, Poisson- och gammafördelningen.

Sannolikhetsfördelningen för en responsvariabel som följer en EDM ges av

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{\phi / w_i} + c(y_i, \phi, w_i) \right).$$

Här är θ_i en parameter som kan variera mellan observationer, medan spridningsparametern ϕ är konstant över alla observationer. Funktionen $b(\theta_i)$, ofta kallad kumulantfunktionen, är två gånger kontinuerligt deriverbar med en inverterbar första derivata. För en given kumulantfunktion $b(\cdot)$ är fördelningen entydigt bestämd av parametrarna θ_i och ϕ . Funktionen $c(\cdot, \cdot, \cdot)$ är däremot av mindre intresse inom GLM teorin eftersom den inte påverkar fördelningens första och andra moment.

Denna formulering av EDM utgör grunden för den generaliserade linjära modellen, där sambandet mellan de förklarande variablerna och responsvariabeln kan beskrivas via en länkfunktion [15, ss. 15 – 18].

Den systematiska komponenten relaterar de förklarande variablerna till responsvariabeln genom en linjär prediktor

$$\eta_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}.$$

Den linjära prediktorn fungerar som ett mellanled mellan de observerade förklarande variablerna och den fördelning som modellen antar för responsvariabeln. Eftersom responsvariabeln i en GLM varken behöver vara normalfördelad eller linjärt relaterad till de förklarande variablerna. Därav används en länkfunktion för att koppla den linjära prediktorn till väntevärdet $\mu_i = \mathbb{E}[Y_i]$ av responsvariabeln

$$g(\mu_i) = \eta_i.$$

Länkfunktionen $g(\cdot)$ transformerar väntevärdet μ_i , till den linjära prediktorn η_i . Funktionen är monoton och deriverbar. Ett vanligt val är identitetslänkfunktionen $g(\mu_i) = \eta_i$, vilket leder till en linjär modell för väntevärdet. Ett annat vanligt val är log-länkfunktionen, $\log(\mu_i) = \eta_i$ som ofta används när responsvariabeln antas vara Poissonfördelad. Den kanoniska länkfunktionen definieras som den funktion som direkt relaterar den naturliga parametern θ_i till väntevärdet μ_i , vilket innebär att

$$g(\mu_i) = \theta_i.$$

Användningen av en kanonisk länk säkerställer att modellen utnyttjar egenskaperna hos den exponentiella familjen fullt ut och leder ofta till förenklade uttryck för sannolikhetsfunktionens derivator, vilket i sin tur underlättar parameterskattning.[1, ss. 113–114]

3.5 Random forest

Random forest är en maskininlärningsmetod som bygger vidare på beslutsträd i syfte att öka modellens prediktiva noggrannhet och minska varians. Random forest är ett kraftfullt verktyg för både klassificerings- och regressionsproblem. Den centrala idén bakom random forest är att kombinera många svagt korrelerade beslutsträd till en enda modell, vars prediktioner aggregeras genom medelvärde vid regressionsproblem. Detta leder till förbättrad generaliseringsförmåga jämfört med enskilda träd.

För att förstå random forest är det först relevant att kort beskriva grunden, vilket är beslutsträd. Beslutsträd är modeller som delar upp data i olika regioner genom en serie binära beslut baserade på värden i förklarande variablerna. Ett binärt beslut innebär exempelvis att modellen delar upp observationerna baserat på om ett visst värde i en förklarande variabel underskrider ett tröskelvärde, såsom

$x_j < c$, vilket leder till två grenar, en för de observationer som uppfyller villkoret och en för de som inte gör det. Dessa modeller är intuitiva och lättförståeliga, men lider ofta av hög varians. Små förändringar i indata kan leda till stora förändringar i den slutliga modellen. Detta gör dem instabila och potentiellt känsliga för överanpassning. Random forest adresserar denna svaghet genom att införa två centrala källor till slumpmässighet i modellbyggandet.

Den första källan för slumpmässighet är bootstrap urval, vilket innebär att varje enskilt träd tränas på ett slumpmässigt urval (med återläggning) av den ursprungliga datamängden. Detta tillvägagångssätt kallas även bagging (bootstrap aggregating) och syftar till att minska variansen i modellen. Den andra källan till slumpmässighet införs vid varje nodelning i träden, där enbart ett slumpmässigt urval av förklarande variabler övervägs vid varje potentiell split. Genom att på detta sätt minska korrelationen mellan träden förbättras den aggregerade modellens prestanda ytterligare.

En fördelaktig egenskap hos random forest är att medelvärden av många svagt korrelerade modeller kan kraftigt reducera modellens varians. Enligt principen bakom baggingmetoder kan detta uttryckas matematiskt. För en ny observation $\mathbf{x}$, beräknas prediktionen från en random forest modell med B träd enligt

$$\hat{f}_{\text{RF}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{(b)}(\mathbf{x}),$$

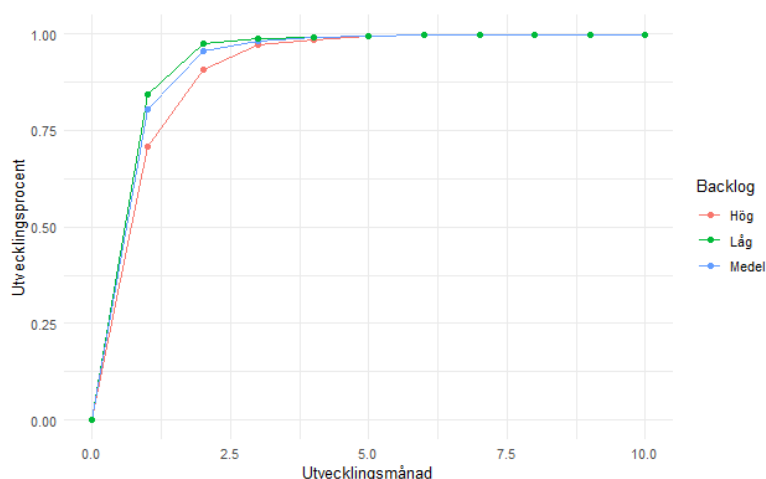
där $\hat{f}^{(b)}(\mathbf{x})$ är prediktionen från det b :te trädet. Med andra ord, att ta medelvärdet av flera prediktioner från olika träningsuppsättningar leder till en mer stabil modell med lägre varians, vilket förbättrar modellens prestanda på testdata.

Random forest är särskilt effektiv vid prediktionsproblem där det föreligger komplexa, icke linjära samband samt i situationer där antalet prediktorer är stort. Fördelar med metoden är hög prediktionsnoggrannhet, robusthet mot överanpassning och förmågan att hantera både numeriska och kategoriska data.

Samtidigt finns vissa begränsningar med random forest. Framför allt är modellen svår att tolka jämfört med enklare metoder såsom generaliserade linjära modeller, chain ladder eller traditionella beslutsträd. Eftersom prediktionerna bygger på genomsnittet av ett stort antal träd, förloras möjligheten att följa ett tydligt beslutsflöde. Detta kan vara en nackdel i sammanhang där tolkbarhet är viktig, exempelvis inom vissa aktuariella tillämpningar. Ytterligare en begränsning är att random forest kräver en stor mängd data för att ge tillförlitliga resultat och kan således ha sämre prestanda när datamängden är begränsad. [9, ss. 340–345]

4 Resultat

Mack chain ladder metoden bygger på antagandet att utvecklingsmönstret för skador är stabilt över tid samt oberoende av skademånad. Detta innebär att övergången från en utvecklingsmånad till nästa kan beskrivas med samma utvecklingsfaktorer, oavsett vilken skademånad som avses (se avsnitt 3.3).



Figur 2: Utvecklingsprocent för olika backlognivåer. Utvecklingsprocenten är definierad enligt $\frac{1}{f_j}$.

Figur 2 visar hur utvecklingsprocenten varierar beroende på backlognivån. Figuren illustrerar att utvecklingsmönstret förändras mellan perioder med låg respektive hög backlog. Detta tyder på att chain ladder antagandet om konsekventa och utvecklingsfaktorer oberoende av skademånad, inte är uppfyllt vid varierande backlognivåer. Metoden fungerar väl under förhållanden där backlognivån är relativt konstant över tid, men vid större variationer i backlog bryts grundantagandet om stabila utvecklingsmönster.

I figuren kan vi exempelvis observera att utvecklingsprocenten vid låg backlognivå ligger omkring 0.85, medan den vid hög backlognivå är cirka 0.7 under den första utvecklingsmånaden. Det innebär att en hög backlog tenderar att ge ett lägre antal behandlade skador tidigt i utvecklingen, vilket påverkar pålitligheten i skattningen av utvecklingsfaktorer.

För att konkretisera denna problematik presenteras nedan ett illustrativt exempel. Det kumulativa antalet behandlade skador per skademånad och utvecklingsmånad visas i tabell 2.

Tabell 2: Kumulativa antalet behandlade skador per skademånad och utvecklingsmånad.

Skademånad	Utvecklingsmånad 1	Utvecklingsmånad 2	Utvecklingsmånad 3	Utvecklingsmånad 4	Utvecklingsmånad 5
1	100	150	180	190	200
2	120	170	200	210	
3	110	180	190		
4	150	170			
5	40				

Utifrån dessa data skattas chain ladder utvecklingsfaktorerna enligt tabell 3.

Tabell 3: Skattade chain ladder utvecklingsfaktorer utifrån exempeldata

Utvecklingsfaktor	Värde
$\hat{f}_1$	1.396
$\hat{f}_2$	1.140
$\hat{f}_3$	1.053
$\hat{f}_4$	1.053

Om vi exempelvis tillämpar faktorn $\hat{f}_1$ på den femte skademånaden, där endast 40 skador är behandlade i första utvecklingsmånaden, fås $40 \cdot \hat{f}_1 = 55.84$. Prognosen för ultimotalet, baserat på samtliga utvecklingsfaktorer, blir

$$40 \cdot \hat{f}_1 \cdot \hat{f}_2 \cdot \hat{f}_3 \cdot \hat{f}_4 = 70.58.$$

Detta är betydligt lägre än vad som rimligen kan förväntas, givet historiska mönster och det inledande låga antalet behandlade skador som orsakas av den höga backlognivån. Detta exempel belyser tydligt hur chain ladder metoden underskattar ultimo när backlognivån avviker från den genomsnittliga nivån som råder i den historiska data, vilken utvecklingsfaktorerna är baserade på.

Sammanfattningsvis fungerar chain ladder metoden bäst när backlognivån är stabil över tid. Vid kraftigt varierande backlog bryts metodens grundläggande antaganden, vilket leder till felaktiga prognoser. Problematiken är central i rapporten och ligger till grund för behovet av att undersöka alternativa metoder.

4.1 Modeller

I detta avsnitt presenteras modellvarianterna som kan användas för att skatta utvecklingsfaktorer. Inledningsvis illustreras den klassiska chain ladder metoden och GLM diagonaleffektsmodellvarianten. Därefter introduceras alternativa metoder som kan hantera variationer i backlognivå och därmed kringgå de begränsningar som chain ladder medför.

4.1.1 Chain ladder metoden

Chain ladder metoden utgår från att skadeutvecklingen följer ett stabilt mönster över tid, vilket innebär att förändringen mellan utvecklingsmånader kan beskrivas med samma utvecklingsfaktorer, oberoende av vilken skademånad skadan inträffade. Denna utveckling illustreras ofta med en triangulär form där både observerade utvecklingsfaktorer och skattade utvecklingsfaktorer framgår. Ett generellt exempel på en sådan struktur för utvecklingsfaktorerna ges i tabell 4.

Vi definierar i som skademånad, där $1 \leq i \leq I$, och j som utvecklingsmånad, där $1 \leq j \leq J$. Kalendermånaden ges av $t = i + j - 1$. Datamaterialet som använts är definierat enligt avsnitt 2. Låt $N_{i,j}$ beteckna de kumulativa antalet behandlade skador. Data för skadebalanserna (backlog) månadsvis benämns b

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_I \end{pmatrix}.$$

Backlog elementet b_i representerar antalet ohanterade skadeärenden i slutet av skademånad i . Det definieras som antalet skador som har inträffat före eller under skademånad i , men som fortfarande är oreglerade vid slutet av samma månad. Data är tillgängliga för de kumulativa antalet behandlade skador $N_{i,j}$ i triangelformat enligt

$$N = \begin{pmatrix} n_{1,1} & n_{1,2} & \cdots & n_{1,J} \\ n_{2,1} & \cdots & n_{2,J-1} & \\ \vdots & \ddots & & \\ n_{I,1} & & & \end{pmatrix}.$$

De observerade värdena används för att skatta utvecklingsfaktorer $\hat{f}_j$ enligt

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} N_{i,j+1}}{\sum_{i=1}^{I-j} N_{i,j}}.$$

Dessa faktorer används därefter för att stegvis fylla i den nedre högra delen av triangeln med skattade utvecklingsfaktorer.

Tabell 4: Utvecklingsfaktorer enligt chain ladder metoden med generell indexering. Observerade utvecklingsfaktorer ($f_{i,j}$) presenteras inom den övre vänstra triangeln, medan skattade utvecklingsfaktorer ($\hat{f}_j$) visas i den nedre högra delen av tabellen.

$i \setminus j$	1	2	3	$\dots$	n	$n+1$	$\dots$	$J-1$
1	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$\dots$	$f_{1,n}$	$f_{1,n+1}$	$\dots$	$\hat{f}_{J-1}$
2	$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$\dots$	$f_{2,n}$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
3	$f_{3,1}$	$f_{3,2}$	$f_{3,3}$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\dots$		$\dots$		
	$\vdots$	$\vdots$	$\hat{f}_3$	$\dots$		$\dots$	$\dots$	$\hat{f}_{J-1}$
$I-1$	$f_{I-1,1}$	$\hat{f}_2$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
I	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$

Som visats i föregående avsnitt (avsnitt 4) är denna metod känslig för förändringar i backlognivå, vilket påverkar hur representativa de skattade utvecklingsfaktorerna är. Därför presenteras andra alternativa metoder som explicit beaktar dessa variationer i backlognivån.

4.1.2 GLM modell med diagonaleffekt

Modellvarianten GLM modell med diagonaleffekt ersätter chain ladder utvecklingsfaktorer längs diagonalen för att fånga effekten av varierande backlognivåer längs diagonalen. Grundtanken är att en hög backlog vid en viss kalendermånad påverkar utvecklingsfaktorer för den kalendermånaden (diagonalen). Syftet med modellen är att ersätta chain ladder utvecklingsfaktorerna för att ta hänsyn till att en hög backlognivå vid en viss kalendermånad kan påverka hur snabbt skador behandlas.

För att förtydliga hur utvecklingsfaktorerna $f_{i,j}^{\text{Diag}}$ skattas i GLM modellen med diagonaleffekt, används de två matriserna

- **Antalstriangeln** $N_{i,j}$, de kumulativa antalet behandlade skador för en viss skademånad i och utvecklingsmånad j .
- **Skadebalanskvoten** $SR_{i,j} = \frac{b_{i+j-1}}{N_{i,j}}$, definierad som backlognivå i förhållande till totala antalet behandlade skador. Detta för att kompensera för förändringar i portföljstorlek.

Nedan illustreras skadebalanskvoten $SR_{i,j}$ i triangelformat, där varje element motsvarar en kalendermånad $i + j - 1$

$$\begin{bmatrix} SR_{1,1} & SR_{1,2} & SR_{1,3} & SR_{1,4} & SR_{1,5} \\ SR_{2,1} & SR_{2,2} & SR_{2,3} & SR_{2,4} & \\ SR_{3,1} & SR_{3,2} & SR_{3,3} & & \\ SR_{4,1} & SR_{4,2} & & & \\ SR_{5,1} & & & & \end{bmatrix}.$$

Utvecklingsfaktorn definieras som kvoten mellan kumulativa antalet behandlade skador vid två på varandra följande utvecklingsmånader

$$Y_{i,j} = \frac{N_{i,j+1}}{N_{i,j}}.$$

Målet med modellen är att modellera det betingade väntevärdet av $Y_{i,j}$ givet $SR_{i,j}$, vilket motsvarar utvecklingsfaktorn $f_{i,j}^{\text{Diag}}$.

Modellen är en generaliserad linjär modell (GLM) med log-länkfunktion, där $Y_{i,j}$ utgör responsvariabeln och skadebalanskvoten $SR_{i,j}$ används som förklarande variabel. På så vis skattas hur antalet behandlade skador utvecklas från utvecklingsmånad j till $j + 1$ för en given skademånad i . Till skillnad från modeller baserade på skadekostnader är denna modell definierad för antalsdata. Eftersom backlognivån påverkar samtliga skador som regleras under en viss kalendermånad, baseras modellen på kalendermånad snarare än skade- eller utvecklingsmånad, vilket gör den kalendertidsspecifik. Utvecklingsfaktorn $f_{i,j}^{\text{Diag}}$ estimeras med hjälp av historisk data för både behandlade skador och skadebalanskvoter.

Tabell 5: GLM modell med diagonaleffekt uttryckt med generell indexering för att illustrera vilka utvecklingsfaktorer som ersätts med GLM modellen ($\hat{f}_{i,j}^{\text{Diag}}$) samt vilka som bestäms av Mack chain ladder ($\hat{f}_j$).

$i \setminus j$	1	2	3	$\dots$	n	$n+1$	$\dots$	$J-1$
1	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$\dots$	$f_{1,n}$	$f_{1,n+1}$	$\dots$	$\hat{f}_{J-1}$
2	$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$\dots$	$f_{2,n}$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
3	$f_{3,1}$	$f_{3,2}$	$f_{3,3}$	$\dots$	$\hat{f}_{I-n,n}^{\text{Diag}}$	$\hat{f}_{n+1}$	$\dots$	$\vdots$
	$\vdots$	$\vdots$		$\dots$		$\dots$		
	$\vdots$	$\vdots$	$\hat{f}_{I-2,3}^{\text{Diag}}$	$\dots$		$\dots$	$\dots$	$\hat{f}_{J-1}$
$I-1$	$f_{I-1,1}$	$\hat{f}_{I-1,2}^{\text{Diag}}$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
I	$\hat{f}_{I,1}^{\text{Diag}}$	$\hat{f}_2$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$

I tabell 5 illustreras hur denna modellvariant implementeras. De diagonala utvecklingsfaktorer som ersätts med GLM modellen är benämnda $\hat{f}_{i,j}^{\text{Diag}}$, medan

övriga utvecklingsfaktorer sätts enligt chain ladder ($\hat{f}_j$). Ett förutbestämt antal diagonala utvecklingsfaktorer, n , ersätts enligt denna modellstruktur. Där samtliga motsvarar samma kalendertid och påverkas av motsvarande skadebalanskvot $SR_{i,j}$.

Den specifika GLM modellen är formulerad enligt följande

- **Responsvariabel:**

$$Y_{i,j} = \frac{N_{i,j+1}}{N_{i,j}}, \quad \text{där} \quad f_{i,j}^{\text{Diag}} = E[Y_{i,j} \mid SR_{i,j}].$$

- **Systematisk komponent:**

$$\eta_{i,j} = \beta_1 + \beta_2 \cdot SR_{i,j}.$$

- **Länkfunktion och fördelning:** Responsvariabeln antas vara gammafördelad och modellen använder en log-länkfunktion:

$$\log(f_{i,j}^{\text{Diag}}) = \eta_{i,j} = \beta_1 + \beta_2 \cdot SR_{i,j}.$$

Detta innebär att den skattade utvecklingsfaktorn ges av

$$\hat{f}_{i,j}^{\text{Diag}} = \exp(\hat{\beta}_1 + \hat{\beta}_2 \cdot SR_{i,j}).$$

Här betecknar

- β_1 : interceptparametern.
- β_2 : regressionskoefficienten som fångar effekten av skadebalanskvoten $SR_{i,j}$.

Parametrarna β_1 och β_2 skattas med maximum likelihood, baserat på historisk data från antalstriangeln samt matrisen för skadebalanskvoter.

4.1.3 GLM modell med radeffekt

I GLM modellen med radeffekt tillämpas en radvis modellstruktur för att fånga effekten av hög backlog för en viss skademånad. En förhöjd backlognivå antas påverka behandlingstakten för samtliga skador från denna skademånad, vilket i sin tur påverkar utvecklingsfaktorerna. Till skillnad från diagonaleffektsmodellen som modellerar påverkan per kalendermånad. Innebär den radvisa ansatsen att skadebalansnivån har en direkt inverkan på utvecklingsfaktorerna kopplade till just den skademånad som backlogen härrör.

Tabell 6 visar hur denna modellvariant implementeras. De utvecklingsfaktorer som ersätts vid tillämpning av radeffektsmodellen betecknas med $\hat{f}_{i,j}^{\text{Rad}}$, medan $\hat{f}_j$

representerar utvecklingsfaktorerna från chain ladder metoden. Precis som i avsnitt 4.1.2 ersätts ett givet antal utvecklingsfaktorer, n , men i denna modellvariant längs raderna i triangeln istället för längs diagonalerna. Övriga utvecklingsfaktorer sätts enligt chain ladder metoden.

Tabell 6: GLM modell med radeffekt uttryckt med generell indexering för att illustrera vilka utvecklingsfaktorer som ersätts med GLM modellen ($\hat{f}_{i,j}^{\text{Rad}}$) samt vilka som bestäms av Mack chain ladder ($\hat{f}_j$).

$i \setminus j$	1	2	3	$\dots$	n	$n+1$	$\dots$	$J-1$
1	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$\dots$	$f_{1,n}$	$f_{1,n+1}$	$\dots$	$\hat{f}_{J-1}$
2	$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$\dots$	$f_{2,n}$	$\hat{f}_{n+1}$	$\ddots$	$\hat{f}_{J-1}$
3	$f_{3,1}$	$f_{3,2}$	$f_{3,3}$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\ddots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$			$\ddots$		
	$\vdots$	$\vdots$	$\hat{f}_3$	$\ddots$		$\ddots$	$\ddots$	$\hat{f}_{J-1}$
$I-1$	$f_{I-1,1}$	$\hat{f}_2$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
I	$\hat{f}_{I,1}^{\text{Rad}}$	$\hat{f}_{I,2}^{\text{Rad}}$	$\hat{f}_{I,3}^{\text{Rad}}$	$\dots$	$\hat{f}_{I,n}^{\text{Rad}}$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$

De n utvecklingsfaktorer som ersätts på den sista raden i triangeln motsvarar utvecklingsfaktorer för en specifik skademånad i . Detta beror på att det antas att samtliga oregrerade skador påverkas av skadebalansnivån b_i för den aktuella skademånaden. Matrisen $B_{i,j}$ definieras enligt följande

$$\begin{bmatrix} b_1 & b_1 & b_1 & b_1 & b_1 \\ b_2 & b_2 & b_2 & b_2 & \\ b_3 & b_3 & b_3 & & \\ b_4 & b_4 & & & \\ b_5 & & & & \end{bmatrix}.$$

Modellen specificeras genom det betingade väntevärdet av responsvariabeln givet skadebalansnivån, vilket uttrycks som

$$f_{i,j}^{\text{Rad}} = E[Y_{i,j} \mid B_{i,j}].$$

Den systematiska komponenten formuleras enligt

$$\eta_{i,j} = \beta_1 + \beta_2 \cdot B_{i,j},$$

där $B_{i,j}$ är skadebalansnivån för skademånad i . Övriga modellkomponenter är identiska med dem som använts i modellen för diagonaleffekt. Detta gäller responsvariabeln, länkfunktionen, antagandet om gammafördelning samt skattningsmetoden (maximum likelihood), se avsnitt 4.1.2.

4.1.4 GLM modell med kombineradeffekt

I detta avsnitt presenteras en modellstruktur som kombinerar både radeffekten och diagonaleffekten från de tidigare modellvarianterna i avsnitt 4.1.2 och 4.1.3. Först presenteras grundmodellen för kombineradeffekt. Därefter introduceras två vidareutvecklade varianter som även beaktar backlogförändring, både i antal och procent.

Syftet är att fånga hur en varierande backlognivå påverkar utvecklingsmönstret på två sätt. För det första genom radeffekten, där backlognivån för en viss skademånad påverkar utvecklingen under de n första utvecklingsmånaderna för just den skademånaden. För det andra genom diagonaleffekten, där backlognivån vid en viss kalendermånad påverkar utvecklingsfaktorerna för de n första utvecklingsfaktorerna längs den sista diagonalen för den givna kalendermånaden, oavsett skademånad.

Tabell 7: Kombinerad modellvariant som tillämpar både diagonaleffekt och radeffekt kombinerat. Generell indexering används för att illustrera vilka utvecklingsfaktorer som ersätts med GLM modellen för radeffekten ($\hat{f}_{i,j}^{\text{Rad}}$), för diagonaleffekten ($\hat{f}_{i,j}^{\text{Diag}}$) samt vilka utvecklingsfaktorer som bestäms av Mack chain ladder ($\hat{f}_j$).

$i \setminus j$	1	2	3	$\dots$	n	$n+1$	$\dots$	$J-1$
1	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$\dots$	$f_{1,n}$	$f_{1,n+1}$	$\dots$	$\hat{f}_{J-1}$
2	$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$\dots$	$f_{2,n}$	$\hat{f}_{n+1}$	$\ddots$	$\hat{f}_{J-1}$
3	$f_{3,1}$	$f_{3,2}$	$f_{3,3}$	$\dots$	$\hat{f}_{I-n,n}^{\text{Diag}}$	$\hat{f}_{n+1}$	$\ddots$	$\vdots$
	$\vdots$	$\vdots$		$\ddots$		$\ddots$		
	$\vdots$	$\vdots$	$\hat{f}_{I-2,3}^{\text{Diag}}$	$\ddots$		$\ddots$	$\ddots$	$\hat{f}_{J-1}$
$I-1$	$f_{I-1,1}$	$\hat{f}_{I-1,2}^{\text{Diag}}$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
I	$\hat{f}_{I,1}^{\text{Diag}}$	$\hat{f}_{I,2}^{\text{Rad}}$	$\hat{f}_{I,3}^{\text{Rad}}$	$\dots$	$\hat{f}_{I,n}^{\text{Rad}}$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$

Tabell 7 visar hur modellen implementeras. Utvecklingsfaktorerna $\hat{f}_{i,j}^{\text{Rad}}$ fångar radeffekten, medan $\hat{f}_{i,j}^{\text{Diag}}$ fångar diagonaleffekten. Utvecklingsfaktorer som inte påverkas av backlog skattas enligt chain ladder metoden och betecknas med $\hat{f}_j$.

GLM modellerna för radeffekt och diagonaleffekt anpassas var för sig och oberoende av varandra och bygger på att modellera det betingade väntevärdet av responsvariabeln givet respektive backlogmått.

För radeffekten specificeras modellen enligt

$$\hat{f}_{i,j}^{\text{Rad}} = E[Y_{i,j} \mid B_{i,j}].$$

För radeffektsmodellen är $Z_{i,j}^{(\text{Rad})} = B_{i,j}$ den enda förklarande variabeln, där $B_{i,j}$ anger skadebalansnivån för skademånad i . Modellen antar därmed att backlog för en viss skademånad påverkar behandlingstakten för tillhörande skador.

Den systematiska komponenten formuleras som

$$\eta_{i,j}^{(\text{Rad})} = \beta_1^{(\text{Rad})} + \beta_2^{(\text{Rad})} \cdot B_{i,j}.$$

För diagonaleffekten specificeras modellen enligt

$$f_{i,j}^{\text{Diag}} = E[Y_{i,j} \mid SR_{i,j}],$$

där $SR_{i,j}$ representerar skadebalanskvoten vid kalendertid $i + j - 1$. I detta fall utgörs $Z_{i,j}^{(\text{Diag})}$ enbart av $SR_{i,j}$ som förklarande variabel.

Den systematiska komponenten formuleras enligt

$$\eta_{i,j}^{(\text{Diag})} = \beta_1^{(\text{Diag})} + \beta_2^{(\text{Diag})} \cdot SR_{i,j}.$$

Detta utgör grunden för den kombinerade GLM modellen. Utöver denna grundstruktur kan modellen utvidgas för att även fånga dynamiken i hur backlogförändras över tid. Det är nämligen inte bara nivån på skadebalanskvoten påverkar utvecklingsfaktorerna, utan även förändringen i backlog kan signalera temporära belastningstoppar eller avlastning i skadehanteringen. Därför introduceras två vidareutvecklade varianter av den kombinerade GLM modellen som beaktar förändringen i backlog, både i antal och procent.

Kombinerad GLM modell med backlogförändring i antal

Denna modellvariant benämns kombinerad GLM modell med backlogförändring i antal. I tabell 8 specificerar vilka utvecklingsfaktorer som ersätts med utvecklingsfaktorer från GLM modellen för diagonaleffekten $\hat{f}_{i,j}^{\text{Diag}+\Delta b}$ samt vilka som ersätts med GLM modellen för radeffekten $\hat{f}_{i,j}^{\text{Rad}}$. De resterande utvecklingsfaktorerna sätts med chain ladder metoden $\hat{f}_j$.

Tabell 8: Kombinerad GLM modell med backlogförändring i antal som tillämpar både diagonaleffekt och radeffekt kombinerat. Generell indexering används för att illustrera vilka utvecklingsfaktorer som ersätts med GLM modellen för radeffekten ($\hat{f}_{i,j}^{\text{Rad}}$), för diagonaleffekten inklusive backlogförändring ($\hat{f}_{i,j}^{\text{Diag}+\Delta b}$) samt vilka utvecklingsfaktorer som bestäms av Mack chain ladder ($\hat{f}_j$).

$i \setminus j$	1	2	3	$\dots$	n	$n+1$	$\dots$	$J-1$
1	$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$\dots$	$f_{1,n}$	$f_{1,n+1}$	$\dots$	$\hat{f}_{J-1}$
2	$f_{2,1}$	$f_{2,2}$	$f_{2,3}$	$\dots$	$f_{2,n}$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
3	$f_{3,1}$	$f_{3,2}$	$f_{3,3}$	$\dots$	$\hat{f}_{I-n,n}^{\text{Diag}+\Delta b}$	$\hat{f}_{n+1}$	$\dots$	$\vdots$
	$\vdots$	$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	$\vdots$	$\vdots$	$\hat{f}_{I-2,3}^{\text{Diag}+\Delta b}$	$\dots$	$\dots$	$\dots$	$\dots$	$\hat{f}_{J-1}$
$I-1$	$f_{I-1,1}$	$\hat{f}_{I-1,2}^{\text{Diag}+\Delta b}$	$\hat{f}_3$	$\dots$	$\hat{f}_n$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$
I	$\hat{f}_{I,1}^{\text{Diag}+\Delta b}$	$\hat{f}_{I,2}^{\text{Rad}}$	$\hat{f}_{I,3}^{\text{Rad}}$	$\dots$	$\hat{f}_{I,n}^{\text{Rad}}$	$\hat{f}_{n+1}$	$\dots$	$\hat{f}_{J-1}$

I denna variant bibehålls radeffekten oförändrad jämfört med grundmodellen för kombineradeffekt. Däremot inkluderar diagonaleffektsmodellen även den förklarande variabeln, förändringen i backlog mellan två på varandra följande kalendermånader

$$\Delta b_i = b_i - b_{i-1}.$$

Vidare presenteras $\Delta B_{i,j}$ på matrisform

$$\begin{bmatrix} \Delta b_1 & \Delta b_2 & \Delta b_3 & \Delta b_4 & \Delta b_5 \\ \Delta b_2 & \Delta b_3 & \Delta b_4 & \Delta b_5 & \\ \Delta b_3 & \Delta b_4 & \Delta b_5 & & \\ \Delta b_4 & \Delta b_5 & & & \\ \Delta b_5 & & & & \end{bmatrix}.$$

För diagonaleffekten med backlogförändring specificeras modellen enligt

$$f_{i,j}^{\text{Diag}+\Delta b} = E[Y_{i,j} \mid SR_{i,j}, \Delta B_{i,j}],$$

där $SR_{i,j}$ representerar skadebalanskvoten vid kalendertid $i+j-1$ och $\Delta B_{i,j}$ är förändringen i backlog mellan två på varandra följande kalendermånader. För modellvarianten som inkluderar backlogförändring i antal utgör $Z_{i,j}^{(\text{Diag}+\Delta b)} = (SR_{i,j}, \Delta B_{i,j})$ vektorn av förklarande variabler i modellen.

Detta ger följande systematiska komponent för diagonaleffekten

$$\eta_{i,j}^{\text{Diag}+\Delta b} = \beta_1^{\text{Diag}+\Delta b} + \beta_2^{\text{Diag}+\Delta b} \cdot SR_{i,j} + \beta_3^{\text{Diag}+\Delta b} \cdot \Delta B_{i,j}.$$

Kombinerad GLM modell med backlogförändring i procent

Denna modellvariant bygger vidare på samma modellstruktur som föregående modellvariant men ersätter backlogförändringen i antal med den procentuella förändringen. Syftet med denna modell är att neutralisera effekten av varierande portföljstorlek. Detta genom att använda en relativ förändring istället för en förändring i antal. Därav fångas den underliggande dynamiken i backlogutvecklingen på ett mer jämförbart sätt över tid.

Radeffekten i denna modellvariant bibehålls oförändrad från grundmodellen. Däremot ändras diagonaleffekten till att inkludera den procentuella förändringen i backlog mellan två på varandra följande kalendermånader. Låt oss benämna denna modellvarianten den kombinerade GLM modellen med backlogförändring i procent. Den procentuella förändringen beräknas som

$$\Delta b_i^{\%} = \frac{b_i - b_{i-1}}{b_{i-1}} \cdot 100.$$

Där vi uttrycker $\Delta B_{i,j}^{\%}$ på matrisformat enligt

$$\begin{bmatrix} \Delta b_1^{\%} & \Delta b_2^{\%} & \Delta b_3^{\%} & \Delta b_4^{\%} & \Delta b_5^{\%} \\ \Delta b_2^{\%} & \Delta b_3^{\%} & \Delta b_4^{\%} & \Delta b_5^{\%} & \\ \Delta b_3^{\%} & \Delta b_4^{\%} & \Delta b_5^{\%} & & \\ \Delta b_4^{\%} & \Delta b_5^{\%} & & & \\ \Delta b_5^{\%} & & & & \end{bmatrix},$$

och kan ersätta $\Delta B_{i,j}$ i modellen ovan med $\Delta B_{i,j}^{\%}$ för att istället ersätta utvecklingsfaktorerna baserade på procentuell backlogförändring som förklarande variabel.

För diagonaleffekten med procentuell backlogförändring specificeras modellen enligt

$$f_{i,j}^{\text{Diag}+\Delta b^{\%}} = E[Y_{i,j} \mid SR_{i,j}, \Delta B_{i,j}^{\%}],$$

där $SR_{i,j}$ representerar skadebalanskvoten vid kalendertid $i+j-1$ och $\Delta B_{i,j}^{\%}$ är den procentuella förändringen i backlog mellan två på varandra följande kalendermånader. I modellen med procentuell förändring används $Z_{i,j}^{(\text{Diag}+\Delta B^{\%})} = (SR_{i,j}, \Delta B_{i,j}^{\%})$.

Den systematiska komponenten för diagonaleffekten med procentuell förändring ges av

$$\eta_{i,j}^{\text{Diag}+\Delta b} = \beta_1^{\text{Diag}+\Delta b^{\%}} + \beta_2^{\text{Diag}+\Delta b^{\%}} \cdot SR_{i,j} + \beta_3^{\text{Diag}+\Delta b^{\%}} \cdot \Delta B_{i,j}^{\%}.$$

Gemensamma modellantaganden

De kombinerade GLM modellvarianterna ovan bygger samtliga på samma modellstruktur. Responsvariabeln antas följa en gammafördelning och definieras som

kvoten mellan två på varandra följande kumulativa antal behandlade skador

$$Y_{i,j} = \frac{N_{i,j+1}}{N_{i,j}}.$$

Modellen specificeras genom det betingade väntevärdet av responsvariabeln givet en vektor av förklarande variabler

$$E[Y_{i,j} \mid Z_{i,j}^{(\cdot)}] = f_{i,j}^{(\cdot)},$$

där $Z_{i,j}^{(\cdot)}$ representerar de förklarande variabler som ingår i modellvarianten $(\cdot)$. Notationen $(\cdot)$ anger vilken typ av modell som tillämpas, exempelvis Rad, Diag, Diag + ΔB eller Diag + $\Delta B^{\%}$.

För att modellera väntevärdet används en log-länkfunktion, vilket ger

$$\log(f_{i,j}^{(\cdot)}) = \eta_{i,j}^{(\cdot)},$$

där $\eta_{i,j}^{(\cdot)}$ är den systematiska komponenten för respektive modellvariant.

Utvecklingsfaktorerna skattas enligt

$$\hat{f}_{i,j}^{(\cdot)} = \exp(\eta_{i,j}^{(\cdot)}).$$

Modellparametrarna $\beta_k^{(\cdot)}$ skattas med maximum likelihood, baserat på historiska data och de relevanta förklarande variablerna som ingår i $Z_{i,j}^{(\cdot)}$. Denna generella modellstruktur möjliggör flexibel och konsistent implementation av olika modellvarianter inom samma ramverk.

4.1.5 Random forest kombinerad modell med backlogförändring i antal

Utöver de GLM baserade modellvarianterna utvärderas även en mer flexibel maskinlärningsmodell för att prediktera utvecklingsfaktorer. Vi benämner denna modellvariant random forest kombinerad modell med backlogförändring. Som framgår av tabell 8, tillämpas utvecklingsfaktorerna på samma vis som för kombinerade GLM modellen med backlogförändring. För att random forest modellvarianten ska fånga både radeffekten och diagonaleffekten används samma struktur som i den kombinerade GLM modellen med backlogförändring. Diagonaleffekten modelleras som en funktion av både skadebalanskvoten ($SR_{i,j}$) och förändringen i backlog ($\Delta B_{i,j}$), medan radeffekten modelleras som en funktion av backlognivån ($B_{i,j}$). Övriga utvecklingsfaktorer sätts enligt chain ladder metoden.

Diagonaleffektsmodell med backlogförändring appliceras för celler längs den sista diagonalen, där en separat random forest regressionsmodell anpassas för varje utvecklingsmånad j med två förklarande variabler

$$SR_{i,j}, \quad \Delta B_{i,j},$$

där förklarande variablerna är definierade enligt avsnitt 4.1.4. Prediktionen ges som ett genomsnitt över samtliga träd i modellen

$$\hat{f}_{i,j}^{\text{Diag}} = \widehat{RF}^{\text{Diag},j}(SR_{i,j}, \Delta B_{i,j}) = \frac{1}{T} \sum_{t=1}^T T_t^{\text{Diag},j}(SR_{i,j}, \Delta B_{i,j}).$$

$\widehat{RF}^{\text{Diag},j}$ betecknar den random forest modell som tränats specifikt för utvecklingsmånad j . Där T är det totala antalet beslutsträd och $T_t^{\text{Diag},j}$ anger det t :te beslutsträdet för utvecklingsmånad j .

Radeffektmodell med backlognivå appliceras för celler på den sista raden ($i = I$). En separat random forest regressionsmodell tränas per utvecklingsmånad j , där backlognivån $B_{i,j}$ använts som enda förklarande variabel

$$\hat{f}_{i,j}^{\text{Rad}} = \widehat{RF}^{\text{Rad},j}(B_{i,j}) = \frac{1}{T} \sum_{t=1}^T T_t^{\text{Rad},j}(B_{i,j}).$$

För övriga celler används utvecklingsfaktorer enligt chain ladder metoden, det vill säga $\hat{f}_j$.

För att tillämpa random forest har paketet *randomForest* i *R* använts. Samtliga argument har behållits som standard, förutom antalet träd (**ntree**) som har varierats för att minimera OOB MSE (out of bag medelkvadratfel). Detta är ett mått på modellens förmåga att ge träffsäkra prediktioner på data som inte ingick i det enskilda trädets träningsdata. Vid en systematisk sökning över **ntree** värden mellan 100 och 2000 erhöles följande individuella optimala antal träd för utvecklingsmånad 1—5

$$[100, 1700, 1400, 100, 600].$$

Utvecklingsmånad två och tre var de utvecklingsmånader som krävande flest träd, med 1700 respektive 1400 träd för att OOB medelkvadratfelet skulle minimeras. Därför valdes 1700 träd konsekvent för alla fem modeller. Detta ger stabila resultat även för de mest krävande utvecklingsmånaderna, utan att beräkningstiden blir ohanterlig. För radeffektmodellerna krävdes generellt något färre träd för att uppnå minimalt OOB medelkvadratfel. Eftersom beräkningstiden inte ökade nämnvärt, användes 1700 träd för samtliga modeller, både för diagonaleffektmodellerna och radeffektmodellerna.

Att använda fler träd minskar modellens varians och ger robustare, mer reproducerbara utvecklingsfaktorer. Med för få träd blir OOB medelkvadratfelet högre och skattningarna mer känsliga för slumpmässiga variationer. Ökningen i beräkningstid från 1000 till 1700 träd är marginell i denna applikation men resulterar i förbättrad stabilitet i prediktionerna.

4.2 Val av antalet utvecklingsfaktorer (n) för GLM och random forest modellering

Som framgår av modellerna i avsnitt 4.1 ersätts de n första utvecklingsfaktorerna med skattningar från en GLM eller random forest modell, istället för att använda chain ladder. För att avgöra hur många perioder bakåt backlognivån har en påverkan undersöks sambandsdiagram mellan backlognivån och det inkrementella antalet behandlade skador per utvecklingsmånad.

För att kunna kvantifiera när ett samband anses tillräckligt starkt för att inkluderas i modellen, väljs ett tröskelvärde för Spearman korrelationen vid $|\rho| \geq 0.5$. Det vill säga att korrelationer med absolutvärde under denna gräns betraktas inte som tillräckligt starka och anses inte bidra med tillräcklig stark korrelation för att ersättas. Den här avgränsning används för att avgöra hur många utvecklingsmånader (n) som ska ersättas i respektive modell. [14]

Från figur 3 analyseras sambandet för radeffekten mellan det inkrementella antalet behandlade skador för olika utvecklingsmånader och backlognivån. Låt $P_{i,j}$ vara definierat som det inkrementella antalet behandlade skador för respektive skademånad i och utvecklingsmånad j och b_t som backlognivån per kalendermånad $t = i + j - 1$. Med samband avses här främst de visuellt observerbara mönstren i sambandsdiagrammen, vilket kan tolkas som antingen positiva eller negativa samband mellan backlognivån och det inkrementella antalet behandlade skador.

I figur 3a, visar sambandet mellan backlognivån och det inkrementella antalet behandlade skador första utvecklingsmånaden. Figuren visar på att det framgår ett negativt samband. En hög backlognivå tenderar att vara associerad med ett lägre antal behandlade skador, medan en låg backlognivå sammanfaller med ett högre antal behandlade skador. Detta indikerar att en hög backlognivå initialt (första utvecklingsmånaden) kan hämma skaderegleringstakten. Korrelationerna i tabell 9 bekräftar sambandet eftersom Spearmans korrelation är -0.5749 och Kendall är -0.4184 .

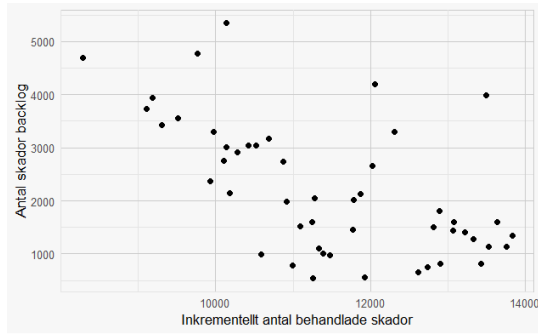
I den andra utvecklingsmånaden, figur 3b, observeras däremot ett positivt samband, där en hög backlognivå samvarierar med ett högre inkrementellt antal behandlade skador. Mönstret upprepas även i figur 3c och figur 3d för utvecklingsmånad 3 respektive 4. Korrelationerna i tabell 9 förstärker detta med höga positiva värden, Spearman korrelationen är 0.7651 för utvecklingsmånad 3 och 0.6427 för utvecklingsmånad 4. Det antyder att effekten av backlogen på skadereglering förändras över tid. Från en hämmande faktor i första utvecklingsmånaden till en möjlig drivkraft för ökat inkrementellt antal behandlade skador i senare utvecklingsmånader.

Figur 3e, motsvarande utvecklingsmånad 5, är sambandet fortsatt positivt, men inte lika tydligt som i de föregående utvecklingsmånaderna. Även detta speglas i korrelationerna (Spearman 0.4916 och Kendall 0.3360). För den sjätte utvecklings-

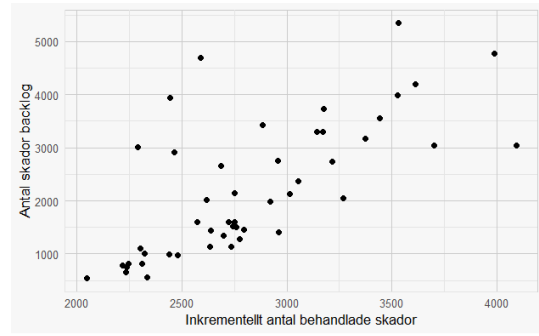
månaden, figur 3f, är inget tydligt observerbart samband. Även om mönstret inte förefaller helt slumpmässigt, är det svårt att identifiera någon klar trend i sambandet. Detta stöds även av korrelationerna (Spearman 0.2955 och Kendall 0.2040), se tabell 9. Vilka är klart lägre än för de övriga utvecklingsmånaderna. Under de kommande utvecklingsmånaderna fortsätter korrelationerna att sjunka, och sambandet framstår som allt mer slumpmässigt (figurer eller korrelationer för dessa utvecklingsmånader är inte inkluderade i rapporten).

Dessa figurer visar att en hög backlognivå initialt leder till färre behandlade skador. I de efterföljande utvecklingsmånaderna ökar däremot antalet behandlade skador. Detta antyder att en hög backlognivå initialt hämmar utvecklingsmönstret, men att denna effekt vänds i senare månader, vilket medför ett förskjutet utvecklingsmönster.

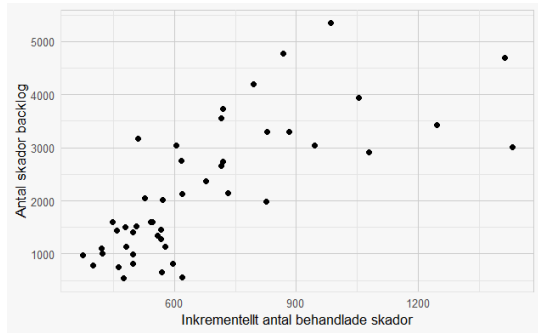
Utifrån figur 3 kan slutsatsen dras att $n = 5$ är rimligt för radeffekten, eftersom vi observerar ett tydligt samband mellan backlognivån och det inkrementella antalet behandlade skador. Spearman korrelationerna för utvecklingsmånad 1–4 ligger över 0.5, vilket tyder på ett starkt samband. Korrelationen i utvecklingsmånad 5 ligger nära gränsen och anses tillräcklig för att motivera inkludering. Detta resultat motiverar att de första fem utvecklingsfaktorerna ersätts i modellerna, för att fånga backlogeffekten.



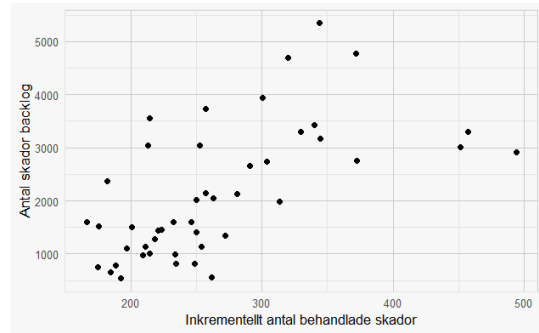
(a) Utvecklingsmånad 1



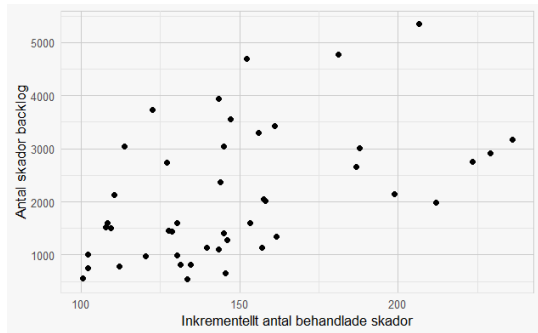
(b) Utvecklingsmånad 2



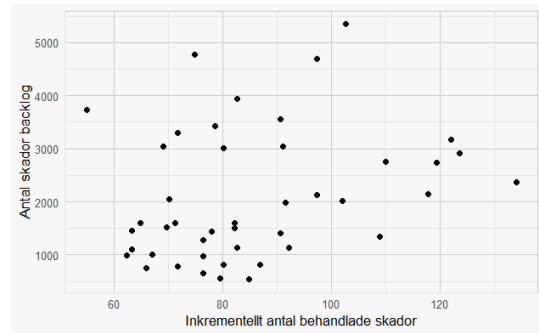
(c) Utvecklingsmånad 3



(d) Utvecklingsmånad 4



(e) Utvecklingsmånad 5



(f) Utvecklingsmånad 6

Figur 3: Visualisering av radeffekten, sambandsdiagrammen visar sambandet mellan backlognivån (b_t) per kalendermånad och det inkrementella antalet behandlade skador ($P_{i,j}$) per utvecklingsmånad (j).

Tabell 9: Korrelationer mellan backlognivån (b_t) per kalendermånad och det inkrementella antalet behandlade skador ($P_{i,j}$) per utvecklingsmånad (j), vilket illustrerar radeffekten.

Utvecklingsmånad	Matematisk representation	Spearman korrelation	Kendall korrelation
1	$\text{cor}(b_t, P_{i,1})$	-0.574898	-0.4183673
2	$\text{cor}(b_t, P_{i,2})$	0.7226444	0.5620567
3	$\text{cor}(b_t, P_{i,3})$	0.7651344	0.5509262
4	$\text{cor}(b_t, P_{i,4})$	0.6427162	0.4632497
5	$\text{cor}(b_t, P_{i,5})$	0.4915667	0.3360331
6	$\text{cor}(b_t, P_{i,6})$	0.2954634	0.2040411

I figur 4 analyseras istället diagonaleffekten, vilket motsvarar backlognivåns påverkan på inkrementella antalet behandlade skador för olika utvecklingsmånader vid en given kalendertid.

För den första och andra utvecklingsmånaden, se figur 4a och figur 4b, överensstämmer sambanden med de som observerats för radeffekten. Det vill säga, ett negativt samband i den första utvecklingsmånaden och ett positivt i den andra. Detta är väntat, eftersom radeffekt och diagonaleffekt sammanfaller under de två inledande utvecklingsmånaderna. Korrelationerna i tabell 10 bekräftar detta.

I de senare utvecklingsmånaderna skiljer sig sambanden från dem som observerats under radeffekten, eftersom diagonaleffekten speglar kalendertid snarare än skademånad. I figur 4c observeras ett svagt positivt samband mellan backlognivån och det inkrementella antalet behandlade skador. Mot bakgrund av att Spearman korrelationen för utvecklingsmånad 3 uppgår till 0.471667, inkluderas även denna utvecklingsmånad som ett tillräckligt starkt samband mellan backlognivå och inkrementella antalet behandlade skador. Bedömningen sker med liknande motivering som för utvecklingsmånad 5 för radeffekten. Däremot tyder figur 4d, figur 4e och figur 4f inte på något tydligt samband. Detta speglas i tabell 10, där korrelationerna i utvecklingsmånad 4 är nära noll eller och utvecklingsmånad 5 och 6 är svagt negativa (t.ex. Spearman -0.3368 för utvecklingsmånad 5).

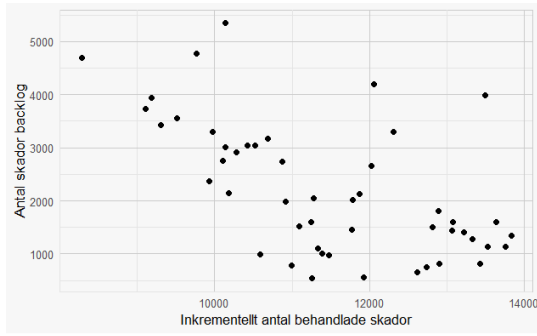
För radeffekten observeras ett samband mellan backlognivån och det inkrementella antalet behandlade skador under flera utvecklingsmånader. Detta till skillnad från diagonaleffekten, där sambandet främst är begränsat till de tre inledande utvecklingsmånaderna. Detta tyder på att radeffekten är starkare och bättre fångar backlognivåns påverkan för olika utvecklingsmånader.

Baserat på sambandsdiagrammen för både radeffekt och diagonaleffekt kommer $n = 5$ att användas i modellerna. Spearman korrelationerna för utvecklingsmånad 1 och 2 ligger över 0.5 och utvecklingsmånad 3 är nära, vilket tyder på starka samband. För utvecklingsmånaderna 4–6 är korrelationerna svagare, de svagare sambandet kan även ses i figur 4. Därför är $n = 3$ ett rimligt val för diagonaleffekten, eftersom endast de tre första utvecklingsmånaderna visar tillräckligt starka samband för att motivera inkludering i analysen.

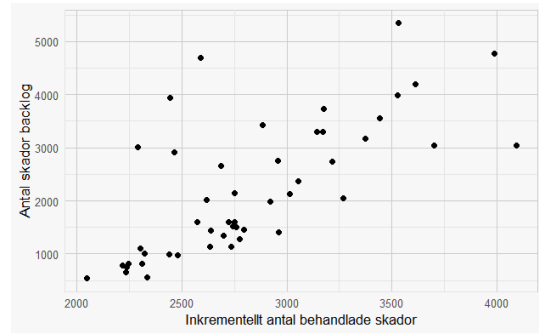
För att säkerställa jämförbarhet mellan radeffekt och diagonaleffekt används samma antal utvecklingsmånader i båda modellerna, det vill säga $n = 5$. Valet grundas på att backlognivån uppvisar starka samband med inkrementella skador under de fem första utvecklingsmånaderna för radeffekten. Även om sambanden är något svagare för diagonaleffekten i de senare utvecklingsmånaderna, möjliggör detta en enhetlig modellstruktur och underlättar en rättvisande jämförelse mellan modellerna.

Tabell 10: Korrelationer mellan backlognivån (b_t) per kalendermånad och det inkrementella antalet behandlade skador ($P_{i,j}$) per utvecklingsmånad (j), vilket illustrerar diagonaleffekten.

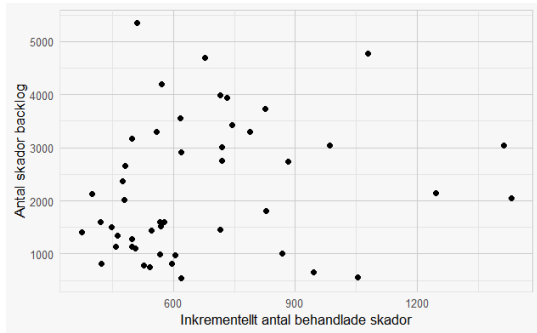
Utvecklingsmånad	Matematisk representation	Spearman korrelation	Kendall korrelation
1	$\text{cor}(b_t, P_{i,1})$	-0.574898	-0.4183673
2	$\text{cor}(b_t, P_{i,2})$	0.7226444	0.5620567
3	$\text{cor}(b_t, P_{i-1,3})$	0.471667	0.2561669
4	$\text{cor}(b_t, P_{i-2,4})$	0.02324371	0.02129631
5	$\text{cor}(b_t, P_{i-3,5})$	-0.3368389	-0.224698
6	$\text{cor}(b_t, P_{i-4,6})$	-0.3198999	-0.2224514



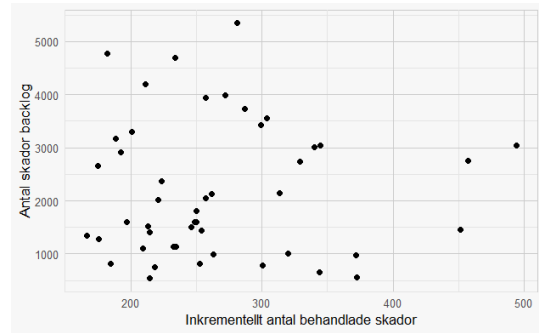
(a) Utvecklingsmånad 1



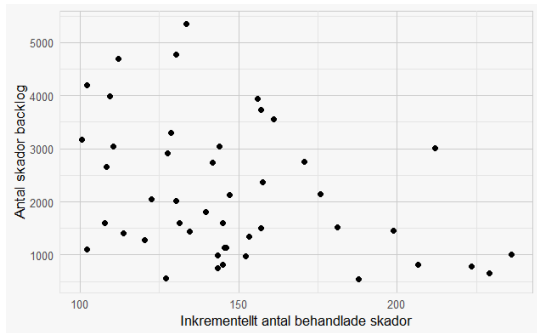
(b) Utvecklingsmånad 2



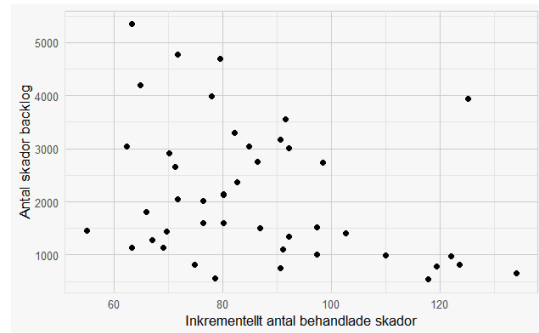
(c) Utvecklingsmånad 3



(d) Utvecklingsmånad 4



(e) Utvecklingsmånad 5



(f) Utvecklingsmånad 6

Figur 4: Visualisering av diagonaleffekten, sambandsdiagrammen visar sambandet mellan backlognivån (b_t) per kalendermånad och det inkrementella antalet behandlade skador ($P_{i,j}$) per utvecklingsmånad (j).

4.3 Jämförelse av modellvarianter

För att utvärdera och jämföra modellvarianternas prestanda används flera olika jämförelsemått. Log likelihood (för en detaljerad genomgång se [13, ss.285–296]) mäter hur väl modellen passar de observerade data, ju högre värde desto bättre

anpassning. Akaike information criterion (för en djupare beskrivning se [17, s. 284]) och bayesian information criterion (se [17, s. 284]) bygger vidare på log likelihood men innehåller en straffterm för antalet parametrar. Detta innebär att lägre värden tyder på en bättre modell, med god balans mellan enkelhet och förklaringsförmåga. Pseudo R^2 (se [7, ss.59–60]) är ett mått för modellens förklaringsgrad, ett högre värde innebär att modellen bättre fångar variationen i data. Residualavvikelse (för ytterligare detaljer se [1, ss. 215–216]) speglar modellens avvikelse från det observerade utfallet, lägre värden indikerar bättre träffsäkerhet.

I avsnitt 4.1.2 – 4.1.4 beskrivs modellvarianternas uppbyggnad. Radeffekt och diagonaleffekt modelleras separat, och separat för varje utvecklingsfaktor. För den första utvecklingsmånaden modelleras samma utvecklingsfaktor, $f_{I,1}$, i både radeffektsmodellen och diagonaleffektsmodellen. För övriga utvecklingsmånader skiljer sig de modellerade utvecklingsfaktorerna åt mellan radeffekt och diagonaleffekt.

Exempelvis för utvecklingsmånad 2 modelleras utvecklingsfaktorn $f_{I,2}$ i samtliga radeffektsmodellvarianter med samma modelluppbyggnad. Detta beror på att alla modellvarianter, även de kombinerade har identiska radeffektsmodeller, där backlognivån är den enda förklarande variabeln. Därav kommer jämförelsemåtten för radeffekten inte att skilja sig mellan modellvarianterna, vilket framgår i tabell 11 – tabell 15. För diagonaleffekten i samma utvecklingsmånad modelleras utvecklingsfaktorn $f_{I-1,2}$ med olika uppsättningar av förklarande variabler, beroende på modellvariant. Exempelvis använder den kombinerade modellvarianten med backlogförändring både skadebalanskvoten och backlogförändring som förklarande variabler för diagonaleffekten, medan andra modellvarianter använder andra kombinationer av förklarande variabler.

En viktig begränsning i denna modellstruktur är att radeffekt och diagonaleffekt modelleras helt oberoende av varandra. Det innebär att varje kombinerad modellvariant i praktiken består av två separata modeller, en för radeffekten och en för diagonaleffekten. Eftersom dessa delar inte modelleras gemensamt går det inte att beräkna samlade utvärderingsmått som exempelvis AIC (Akaike information criterion), BIC (Bayesian information criterion) eller pseudo R^2 för den kombinerade modellen som helhet. Sådana mått bygger på en sammanhängande sannolikhetsfunktion, vilket saknas när effektstrukturerna anpassas var för sig. Därmed är det inte möjligt att objektivt rangordna hela modellvarianter mot varandra utifrån ett enda mått.

Med utgångspunkt i denna särskilda uppdelning redovisas därför samtliga utvärderingsmått separat för varje enskild modellvariant, det vill säga för varje utvecklingsfaktor inom respektive effektstruktur. Det är i huvudsak diagonaleffekten som möjliggör jämförelser mellan modellvarianter, eftersom radeffektsmodellen är densamma oavsett modellvariant. Variationer uppstår i diagonaleffektens modelluppställning eftersom de olika modellvarianterna använder olika kombina-

tioner av förklarande variabler. Detta medför att modellernas förklaringsgrad och anpassning kan skilja sig, vilket gör denna effektstruktur intressant att analysera närmare.

Syftet med jämförelserna är att analysera hur olika jämförelsemått skiljer sig mellan modellvarianter, effektstrukturer och utvecklingsmånader. Modellvarianterna jämförs därför inom, men inte mellan, olika utvecklingsmånader. Jämförelsen baseras på jämförelsemåtten residualavvikelse, AIC, BIC, log likelihood och pseudo R^2 . Dock kan inte random forest modeller jämföras utifrån dessa jämförelsemått med anledning av att det är en icke-parametrisk modell, en jämförelse med denna modellvariant utförs vid utvärdering av modellvarianternas prediktiva förmåga.

Tabell 11: De olika modellvarianterna för radeffekt och diagonaleffekt är anpassade för utvecklingsfaktorn $f_{I,1}$. Eftersom radeffekt och diagonaleffekt modelleras separat samt separat för varje utvecklingsfaktor, presenteras utvärderingsmått enbart för varje specifik utvecklingsfaktor och effektstruktur.

Modellvariant	AIC	BIC	Log likelihood	Pseudo R^2	Residualavvikelse
Diagonaleffekt	-97.4754	-94.0689	51.7377	0.2838	1.498×10^{-2}
Radeffekt	-96.2115	-92.8050	51.1058	0.2315	1.075×10^{-2}
Kombinerad (diagonal)	-97.4754	-94.0689	51.7377	0.2838	1.498×10^{-2}
Kombinerad + Δ Backlog (diagonal)	-91.6237	-87.2595	49.8118	0.3318	1.391×10^{-2}
Kombinerad + $\Delta\%$ Backlog (diagonal)	-90.0217	-85.6576	49.0109	0.2814	1.496×10^{-2}
Kombinerad (rad)	-96.2115	-92.8050	51.1058	0.2315	1.075×10^{-2}
Kombinerad + Δ Backlog (rad)	-96.2115	-92.8050	51.1058	0.2315	1.075×10^{-2}
Kombinerad + $\Delta\%$ Backlog (rad)	-96.2115	-92.8050	51.1058	0.2315	1.075×10^{-2}

Tabell 12: De olika modellvarianterna för radeffekt är anpassade för utvecklingsfaktorn $f_{I,2}$, medan diagonaleffekten är anpassad för utvecklingsfaktorn $f_{I-1,2}$. Eftersom radeffekt och diagonaleffekt modelleras var för sig och enskilt för varje utvecklingsfaktor, presenteras utvärderingsmått separat för varje specifik utvecklingsfaktor och effektstruktur.

Modellvariant	AIC	BIC	Log likelihood	Pseudo R^2	Residualavvikelse
Diagonaleffekt	-160.1007	-156.8276	83.0504	0.3506	6.777×10^{-4}
Radeffekt	-163.2879	-160.0148	84.6439	0.4348	5.467×10^{-4}
Kombinerad (diagonal)	-160.1007	-156.8276	83.0504	0.3506	6.777×10^{-4}
Kombinerad + Δ Backlog (diagonal)	-160.6688	-156.3046	84.3344	0.4222	6.030×10^{-4}
Kombinerad + $\Delta\%$ Backlog (diagonal)	-158.8720	-154.5078	83.4360	0.3730	6.543×10^{-4}
Kombinerad (rad)	-163.2879	-160.0148	84.6439	0.4348	5.467×10^{-4}
Kombinerad + Δ Backlog (rad)	-163.2879	-160.0148	84.6439	0.4348	5.467×10^{-4}
Kombinerad + $\Delta\%$ Backlog (rad)	-163.2879	-160.0148	84.6439	0.4348	5.467×10^{-4}

Tabell 13: De olika modellvarianterna för radeffekt är anpassade för utvecklingsfaktorn $f_{I,3}$, medan diagonaleffekten är anpassad för utvecklingsfaktorn $f_{I-2,3}$. Eftersom radeffekt och diagonaleffekt modelleras var för sig och enskilt för varje utvecklingsfaktor, presenteras utvärderingsmåten separat för varje specifik utvecklingsfaktor och effektstruktur.

Modellvariant	AIC	BIC	Log likelihood	Pseudo R^2	Residualavvikelse
Diagonaleffekt	-200.8729	-197.7393	103.4364	0.3603	6.478×10^{-5}
Radeffekt	-197.9338	-194.8003	101.9669	0.2633	7.241×10^{-5}
Kombinerad (diagonal)	-200.8729	-197.7393	103.4364	0.3603	6.478×10^{-5}
Kombinerad + Δ Backlog (diagonal)	-201.8555	-197.6774	104.9278	0.4450	5.621×10^{-5}
Kombinerad + $\Delta\%$ Backlog (diagonal)	-199.5244	-195.3463	103.7622	0.3799	6.281×10^{-5}
Kombinerad (rad)	-197.9338	-194.8003	101.9669	0.2633	7.241×10^{-5}
Kombinerad + Δ Backlog (rad)	-197.9338	-194.8003	101.9669	0.2633	7.241×10^{-5}
Kombinerad + $\Delta\%$ Backlog (rad)	-197.9338	-194.8003	101.9669	0.2633	7.241×10^{-5}

Tabell 14: De olika modellvarianterna för radeffekt är anpassade för utvecklingsfaktorn $f_{I,4}$, medan diagonaleffekten är anpassad för utvecklingsfaktorn $f_{I-3,4}$. Eftersom radeffekt och diagonaleffekt modelleras var för sig och enskilt för varje utvecklingsfaktor, presenteras utvärderingsmåten separat för varje specifik utvecklingsfaktor och effektstruktur.

Modellvariant	AIC	BIC	Log likelihood	Pseudo R^2	Residualavvikelse
Diagonaleffekt	-206.7760	-203.7888	106.3880	0.1136	2.807×10^{-5}
Radeffekt	-211.6629	-208.6757	108.8314	0.3053	2.161×10^{-5}
Kombinerad (diagonal)	-206.7760	-203.7888	106.3880	0.1136	2.807×10^{-5}
Kombinerad + Δ Backlog (diagonal)	-210.6627	-206.6798	109.3314	0.3396	2.091×10^{-5}
Kombinerad + $\Delta\%$ Backlog (diagonal)	-206.4199	-202.4369	107.2099	0.1835	2.585×10^{-5}
Kombinerad (rad)	-211.6629	-208.6757	108.8314	0.3053	2.161×10^{-5}
Kombinerad + Δ Backlog (rad)	-211.6629	-208.6757	108.8314	0.3053	2.161×10^{-5}
Kombinerad + $\Delta\%$ Backlog (rad)	-211.6629	-208.6757	108.8314	0.3053	2.161×10^{-5}

Tabell 15: De olika modellvarianterna för radeffekt är anpassade för utvecklingsfaktorn $f_{I,5}$, medan diagonaleffekten är anpassad för utvecklingsfaktorn $f_{I-4,5}$. Eftersom radeffekt och diagonaleffekt modelleras var för sig och enskilt för varje utvecklingsfaktor, presenteras utvärderingsmåten separat för varje specifik utvecklingsfaktor och effektstruktur.

Modellvariant	AIC	BIC	Log likelihood	Pseudo R^2	Residualavvikelse
Diagonaleffekt	-229.1156	-226.2823	117.5578	0.3486	4.700×10^{-6}
Radeffekt	-222.6523	-219.8190	114.3261	0.0842	6.539×10^{-6}
Kombinerad (diagonal)	-229.1156	-226.2823	117.5578	0.3486	4.700×10^{-6}
Kombinerad + Δ Backlog (diagonal)	-229.6233	-225.8456	118.8117	0.4291	4.119×10^{-6}
Kombinerad + $\Delta\%$ Backlog (diagonal)	-227.1156	-223.3378	117.5578	0.3486	4.700×10^{-6}
Kombinerad (rad)	-222.6523	-219.8190	114.3261	0.0842	6.539×10^{-6}
Kombinerad + Δ Backlog (rad)	-222.6523	-219.8190	114.3261	0.0842	6.539×10^{-6}
Kombinerad + $\Delta\%$ Backlog (rad)	-222.6523	-219.8190	114.3261	0.0842	6.539×10^{-6}

För utvecklingsmånad 1 (tabell 11) uppvisar den kombinerade modellvarianten lägst AIC och BIC samt högst log likelihood bland diagonaleffektsmodellerna. Dessutom uppnår *Kombinerad + Δ Backlog* högst pseudo R^2 och lägst residualavvikelse för diagonaleffekten. Av tabellen framgår också att radeffektsmodellerna uppvisar lägst residualavvikelse av samtliga modellvarianter. Detta tyder på att modellen passar data bäst och ger den mest träffsäkra prediktionen för utvecklingsfaktorn $f_{I,1}$. Eftersom både rad- och diagonaleffektsmodellerna i utvecklingsmånad 1 predikterar samma utvecklingsfaktor $f_{I,1}$, kan en direkt jämförelse mellan effekterna göras. Resultaten visar att diagonaleffektsmodellerna generellt ger lägre AIC och BIC samt högre log likelihood och pseudo R^2 , vilket är önskvärt. Däremot ger radeffektsmodellen lägre residualavvikelse. Sammanfattningsvis presterar diagonaleffektsmodellerna bättre än radeffektsmodellerna i fyra av fem jämförelsemått, vilket tyder på att dessa modellvarianter bättre fångar variationen i data, trots något högre residualavvikelse. Baserat på samtliga utvärderingsmått bedöms den kombinerade modellvarianten vara att föredra i utvecklingsmånad 1.

I utvecklingsmånad 2 (tabell 12) uppvisar modellvarianten *Kombinerad + Δ Backlog* de mest fördelaktiga värdena på samtliga jämförelsemått för diagonaleffekten. Ett undantag är BIC, där en annan modell har ett marginellt lägre värde. Vad gäller radeffekten är samtliga modellvarianter fortfarande identiska, vilket innebär att de ger exakt samma värden för de olika jämförelsemåtten. Radeffektsmodellerna uppvisar överlag bättre prestanda än diagonaleffektsmodellerna, sett till samtliga jämförelsemått. Slutsatsen för utvecklingsmånad 2 är att *Kombinerad + Δ Backlog* bör föredras för modellering av diagonaleffekten, eftersom den ger bäst övergripande resultat.

I utvecklingsmånad 3 till 5 (tabell 13–tabell 15) bekräftas det tidigare observerade mönstret i modellprestanda. För diagonaleffekten presterar modellvarianten *Kombinerad + Δ Backlog* bäst i utvecklingsmånad 3–5 enligt samtliga jämförelsemått, med undantag för BIC. I utvecklingsmånad 4 uppvisar modellen däremot bäst resultat för samtliga mått, inklusive BIC.

Det är dock viktigt att understryka att de använda jämförelsemåtten inte möjliggör en utvärdering av modellvarianterna som helheter, eftersom radeffekten och diagonaleffekten modelleras separat. Det innebär att gemensamma mått såsom AIC eller pseudo R^2 inte kan beräknas för de kombinerade modellerna i sin helhet. Analysen genomförs därför separat för varje modelleffekt, det vill säga radeffekt respektive diagonaleffekt.

Resultaten i avsnitt 4.2 visar på en tydlig radeffekt, med ett starkare samband mellan backlognivån och det inkrementella antalet behandlade skador än vad som observeras för diagonaleffekten. Detta tyder på att radeffekten har en faktisk inverkan på utvecklingsfaktorerna och bör därför inkluderas i modelleringen. En modellvariant som kombinerar både radeffekt och diagonaleffekt är således motiverad.

Valet står därför mellan de två kombinerade modellvarianterna *Kombinerad + Δ Backlog* och *Kombinerad*. Den förstnämnda presterar bäst i utvecklingsmånad 2 till 5, medan den sistnämnda uppvisar något bättre resultat i utvecklingsmånad 1. Utvecklingsmånad 1 står för en stor andel av det totala antalet skador. Det inkrementella antalet skador är avsevärt högre i denna utvecklingsmånad än i övriga utvecklingsmånader. Därför kan det argumenteras för att utvecklingsmånad 1 bör tilldelas större vikt vid modellvalet.

Samtidigt uppvisar *Kombinerad + Δ Backlog* lägst residualavvikelse även i utvecklingsmånad 1, vilket tyder på en starkare prediktiv förmåga. Detta stärker ytterligare modellens tillförlitlighet i den mest kritiska perioden. Dessutom presterar modellen bäst för diagonaleffekten i fyra av fem utvecklingsmånader, och den underliggande korrelationen i data ger ett tydligt stöd för att radeffekten bör inkluderas i modellen. Eftersom radeffekten är identisk i samtliga modellvarianter är det diagonaleffekten som avgör modellvalet. Sammanfattningsvis väger den överlägsna prestandan i flertalet utvecklingsmånader, tyngre än den marginella fördelen hos *Kombinerad* i just denna utvecklingsmånad. Därför väljs *Kombinerad + Δ Backlog* som slutlig modellvariant. Dock är det värt att notera att modellvarianten *Kombinerad* är något starkare just under den första utvecklingsmånaden, den utvecklingsfaktor som har störst betydelse för reservsättningen.

De estimerade utvecklingsfaktorerna för samtliga modellvarianter redovisas i appendix (tabell 29 – tabell 34). Tabellerna visar vilka utvecklingsfaktorer som har skattats med respektive modellvariant (radeffekt, diagonaleffekt eller kombinerad) för de fem första utvecklingsmånaderna, samt vilka som har skattats med chain ladder metoden. Färgkodningen anger vilken effektstruktur som har använts, blå celler motsvarar chain ladder, gröna avser diagonaleffekt och lila radeffekt. Samtliga modellvarianter använder GLM som regressionsmetod, förutom den sista modellvarianten (tabell 34) där random forest har använts.

4.4 Jämförelse av modellvarianternas prediktiva förmåga

Syftet med denna delsektion är att utvärdera och jämföra de olika modellvarianternas prediktiva förmåga, med fokus på hur väl de kan prediktera utvecklingsfaktorer relaterade till radeffekt och diagonaleffekt. Eftersom modellernas huvudsakliga syfte är att prediktera utvecklingsfaktorer, därav är modellernas prediktiva förmåga av stor betydelse. De prediktiva måtten som använts i utvärderingen är root mean

squared error (för en mer detaljerad genomgång se [8, ss. 5481—5482]) som mäter storleken på prediktionsfelen med högre vikt på större fel. Mean absolute error (se [8, ss. 5481—5482]) anger prediktionernas genomsnittliga absoluta avvikelse från faktiska värden. Mean absolute percentage error (se [4, ss.1–2]) vilket beskriver det genomsnittliga prediktionsfelet uttryckt i procent av faktiska värden. Sedan R^2 (förklaringsgrad) som anger andelen av den totala variationen i observerade data som förklaras av modellen, för mer information se [10, ss. 66–74].

Denna utvärdering skiljer sig från tidigare jämförelser i avsnitt 4.3 eftersom den möjliggör jämförelse mellan samtliga modellvarianter, inklusive random forest modellen som tidigare inte kunde inkluderas. Detta beror på att jämförelsemått såsom AIC och log likelihood inte är tillämpliga på icke-parametriska modeller som random forest. Utvärderingen är därför inte bara viktig för att bedöma modellernas relativa prestanda, utan även för att avgöra vilka varianter som är bäst lämpade för prediktiva syften. En särskilt intressant aspekt är att analysera hur modellernas prediktiva prestanda påverkas av både val av modellstruktur och införandet av olika förklarande variabler, såsom förändringar i backlog.

För att utvärdera modellernas prediktionsförmåga beräknas ett antal jämförelsemått baserade på utvecklingsfaktorer från observerad data. Analysen genomförs genom att en effekt i taget undersöks, vilket innebär att modellernas förmåga att prediktera testas separat för utvecklingsfaktorer längs rader och diagonaler. Detta görs genom att en hel rad eller diagonal i antalstriangeln maskeras (sätts till NA), vilket simulerar att dessa värden saknas. De borttagna värdena motsvarar de utvecklingsfaktorer modellerna i praktiken behöver kunna prediktera. För att uppnå en mer robust och representativ utvärdering upprepas denna procedur för flera olika rader respektive diagonaler. För varje maskning beräknas jämförelsemått mellan predikterade och faktiska värden, varefter ett medelvärde tas fram över samtliga maskningar. De värden som presenteras i tabellerna utgör således medelvärden av jämförelsemåtten över flera maskade rader respektive diagonaler. Denna metod minskar risken för att slutsatser påverkas av enskilda, mer osäkra rader eller diagonaler. Detta annars kan leda till en snedvriden bild på grund av tillfälliga variationer.

Enligt resultaten i tabell 16 uppvisar modellvarianten *GLM Kombinerad* den bästa prediktiva förmågan för utvecklingsfaktorer diagonalvis, baserat på samtliga utvärderingsmått. Modellen har lägst RMSE (Root Mean Squared Error), vilket indikerar att den generellt har mindre avvikelser och har färre stora avvikelser. Den har även lägst MAE (Mean Absolute Error), något som bekräftar att det genomsnittliga felet är lågt. Därtill har modellen det lägsta MAPE (Mean Absolute Percentage Error), detta innebär att de relativa felen är mindre jämfört med övriga modeller. Slutligen uppnår modellen även den högsta förklaringsgraden (R^2), vilket tyder på att modellen fångar en större andel av variationen i de observerade

utvecklingsfaktorerna.

Tabellen visar även att samtliga modeller uppvisar mycket låga prediktionsfel. Att MAPE ligger under 0.4 procent för alla modeller indikerar en överlag hög prediktiv precision i förhållande till de observerade värdena. Det innebär att modellernas prediktioner i genomsnitt endast avviker marginellt i relativa termer. Denna höga träffsäkerhet antyder att utvecklingsfaktorerna till stor del kan förklaras av de ingående förklarande variablerna, samt att modellerna lyckas fånga de huvudsakliga strukturella mönstren i data. Sammantaget, med stöd i jämförelsemåtten som helhet, tyder resultaten på att samtliga modellvarianter uppvisar en mycket god prediktiv förmåga med låga fel både i absoluta och relativa termer. Därmed kan samtliga modellvarianterna anses vara lämpliga att använda i prediktivt syfte.

Random forest modellen, som utvärderats utifrån samma förklarande variabler men med en icke-linjär modellstruktur, uppvisar genomgående sämre resultat än GLM modellvarianterna. Jämförelsemåtten visar tydligt att random forest inte lyckas utnyttja informationen i backlognivån lika effektivt som GLM. En möjlig förklaring är att den linjära ansatsen i GLM modellerna är bättre lämpad för att modellera utvecklingsfaktorerna.

Trots att samtliga modeller bygger på samma radeffektsmodell samt enbart använder backlognivån som förklarande variabel, skiljer sig utvärderingsmåtten åt. Detta beror på hur utvecklingsfaktorerna ersätts i de kombinerade modellerna. I dessa fall bestäms utvecklingsfaktorerna för den första utvecklingsmånaden och den sista skademånaden med hjälp av diagonaleffektsmodellen, vilket introducerar skillnader mellan modellvarianternas resultat. Hade istället samtliga utvecklingsfaktorer på sista raden ersatts med radeffektsmodellen, som är identisk mellan modellerna, skulle även jämförelsemåtten varit desamma för samtliga modellvarianter.

Radeffektsmodellen (GLM), som enbart bygger på radeffektkomponenten, presterar marginellt sämre, med något högre MAE och MAPE samt en något lägre förklaringsgrad. Denna skillnad tyder på att diagonaleffektsmodellen i den kombinerade modellvarianten är bättre lämpad för att prediktera den första utvecklingsfaktorn. Eftersom samtliga modellvarianter använder samma förklarande variabler i radeffektsmodellen för övriga utvecklingsfaktorerna, bör jämförelsemåtten inte skilja sig mellan modellvarianterna. Skillnaden i prediktion uppstår därför främst i den första utvecklingsfaktorn, där diagonaleffektsmodellen har en direkt påverkan. Vi kan därmed dra slutsatsen att diagonalkomponenten i den kombinerade modellen ger en mer träffsäker prediktion just för denna faktor. Detta är särskilt betydelsefullt eftersom den första utvecklingsfaktorn har störst påverkan med tanke på försäkringsproduktens karaktär. Resultaten indikerar att diagonaleffektsmodellen fångar ett mer fullständigt mönster i data än vad radeffektsmodellen gör för denna specifika komponent.

Tabell 16: Jämförelsemått för att utvärdera de olika modellvarianternas prediktiva förmåga för utvecklingsfaktorerna längs raden. Måtten beräknas för varje rad baserat på prediktionsfelet, värdena representerar medelvärdet av dessa mått över flera rader.

Modell	RMSE	MAE	MAPE	R^2
GLM Radeffekt	0.00605	0.00330	0.2949	0.9865
GLM Kombinerad	0.00556	0.00294	0.2638	0.9909
GLM Kombinerad + Δ Backlog	0.00670	0.00346	0.3057	0.9857
GLM Kombinerad + $\Delta\%$ Backlog	0.00669	0.00330	0.2929	0.9846
Random forest Kombinerad + Δ Backlog	0.00718	0.00373	0.3295	0.9843

Tabell 17 visar att samtliga GLM baserade modellvarianter uppvisar stark prediktiv förmåga med mycket likvärdiga resultat. Både *GLM Diagonaleffekt* och *GLM Kombinerad* uppnår identiskt och bäst resultat, med lägst RMSE, lägst MAE och högst förklaringsgrad. Detta tyder på att diagonaleffekten fångas väl även utan ytterligare backlogrelaterade förklarande variabler. I likhet med resultaten för radeffekten kan det konstateras att random forest modellvarianten generellt uppvisar lägre prediktiv precision jämfört med GLM modellvarianterna. Vidare tyder resultaten på att inkludering av backlogförändringar, oavsett om de uttrycks i antal eller procent inte leder till någon förbättring av modellernas prediktiva förmåga.

Tabell 17: Jämförelsemått för att utvärdera de olika modellvarianternas prediktiva förmåga för utvecklingsfaktorerna längs diagonalen. Måtten beräknas för varje diagonal baserat på prediktionsfelet, värdena representerar medelvärdet av dessa mått över flera diagonaler.

Modell	RMSE	MAE	MAPE	R^2
GLM Diagonaleffekt	0.00488	0.00170	0.1495	0.9880
GLM Kombinerad	0.00488	0.00170	0.1495	0.9880
GLM Kombinerad + Δ Backlog	0.00512	0.00178	0.1550	0.9864
GLM Kombinerad + $\Delta\%$ Backlog	0.00499	0.00174	0.1510	0.9873
Random forest Kombinerad + Δ Backlog	0.00578	0.00197	0.1713	0.9831

Sammanfattningsvis visar resultaten att samtliga GLM modellvarianter ger träffsäkra prediktioner. Bland dessa uppnår dock modellvarianten *GLM Kombinerad* konsekvent den bästa prestandan, vilket gör den till det mest fördelaktiga valet. Den uppvisar genomgående bäst resultat för samtliga prediktiva mått vid utvärdering av modellernas förmåga att prediktera utvecklingsfaktorer längs både

diagonaler och rader. Specifikt presterar den bäst för radeffekten och delar förstaplatsen med *GLM Diagonaleffekt* vad gäller diagonaleffekten. Modellvarianter som inkluderar förändringsvariabler i backlog ger inte någon ytterligare förbättrad prediktiv förmåga i denna tillämpning. Vidare uppvisar random forest ingen fördel jämfört med GLM. Sammantaget talar detta för att GLM modellvarianter fortsatt bör föredras för prediktion av utvecklingsfaktorer, åtminstone baserat på det tillgängliga datat.

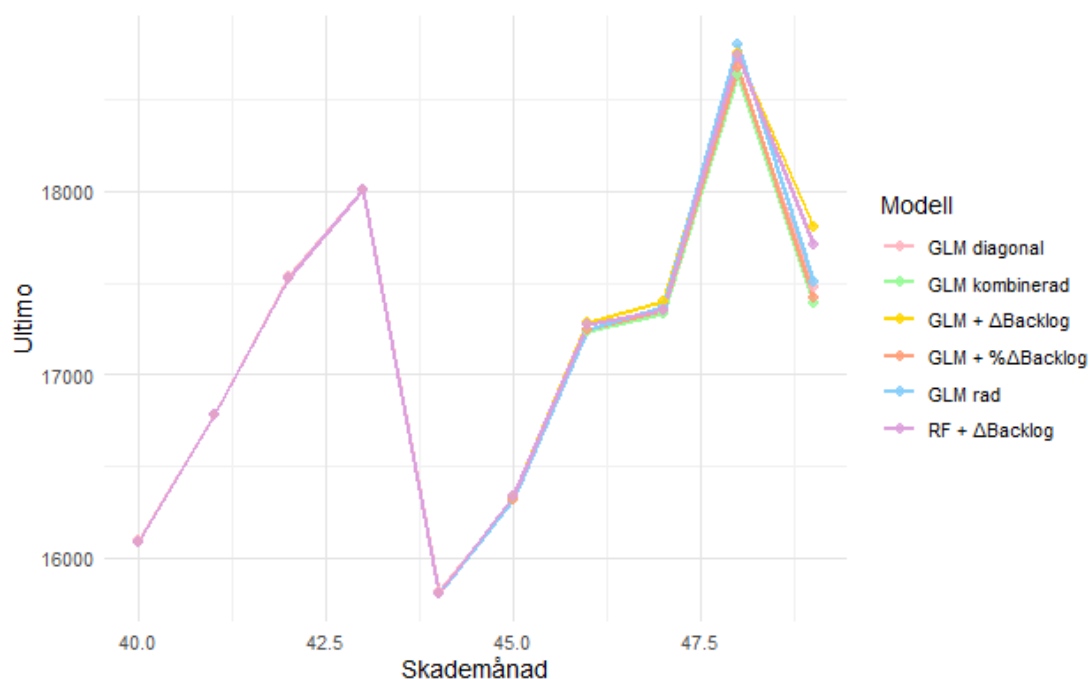
4.5 Antalsreserv

I avsnitt 3.1 introduceras grunden för reservsättning. Med anledning av att rapporten endast fokuserar på antal behandlade skador är reserven angiven i antal skador och inte skadekostnad. För att skatta de okända cellerna i antalstriangeln tillämpas samtliga sex modellvarianter. Dessa skiljer sig åt både vad gäller vilka utvecklingsfaktorer som ersätts samt vilka metoder (GLM eller random forest) som använts. Därav genererar varje modell olika skattningar av såväl ultimoreserv som kassaflöden.

I appendix (tabell 23—tabell 28) redovisas det kumulativa antalet behandlade skador för de tolv första utvecklingsmånaderna samt ultimo, begränsat till de femton senaste skademånaderna. Denna avgränsning motiveras av att samtliga modeller enbart modifierar de fem första utvecklingsfaktorerna, längs sista raden, diagonalen eller en kombination av dessa. Eftersom resterande utvecklingsfaktorer beräknas med samma chain ladder faktorer, blir skattningarna av ultimo för tidigare skademånader identiska mellan modellvarianterna.

Figur 5 visualiserar skattningarna av ultimo för samtliga modellvarianter. Figuren visar att skillnaderna mellan modellerna är obetydliga för de skademånader där majoriteten av utvecklingen redan har inträffat, men att skillnaderna ökar för de senare skademånaderna. Denna ökade skillnad kan förklaras av att den aktuella försäkringsprodukten är kortsvansad, där merparten av skadorna behandlas tidigt i utvecklingsförloppet och i stor utsträckning är fullt reglerade efter tre utvecklingsmånader. Därmed får de första utvecklingsfaktorerna, särskilt den första, ett betydande genomslag på skattningen av ultimo.

Av figur 5 framgår att de två modellvarianterna, GLM och random forest, som båda använder en kombinerad modellstruktur med backlogförändring som förklarande variabel, ger olika ultimoskattningar. GLM modellen tenderar att generera något högre ultimo jämfört med random forest. Dessutom är GLM modellen med backlogförändring den modellvariant som oftast ger den högsta reservskattningen för majoriteten av skademånaderna. Samtidigt är det GLM modellen med kombinerad struktur utan förändringsvariabler som generellt ger de lägsta ultimoskattningarna.



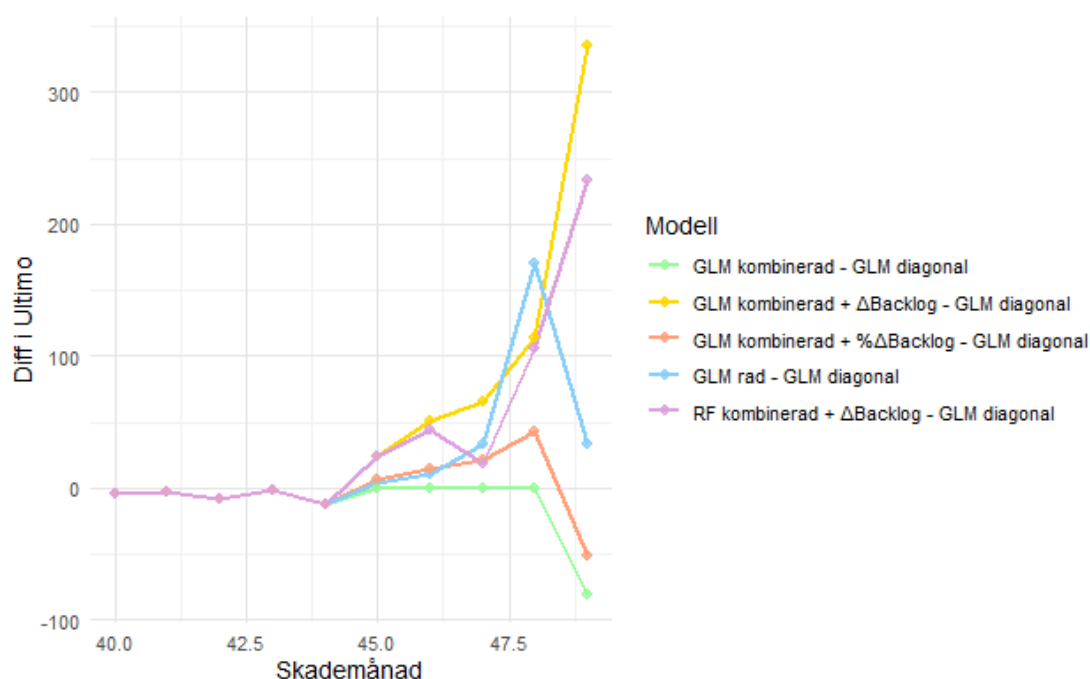
Figur 5: Jämförelse av ultimoskattningen för samtliga modellvarianter, illustrerad för de sista 15 skademånaderna.

I figur 6 visas differensen mellan ultimoskattningarna för respektive modellvariant och referensmodellen (diagonalmodellen). Figuren ger en tydligare översikt av hur modellvarianterna skiljer sig från varandra, jämfört med figur 5. Det framgår att samtliga modellvarianter, förutom de som inkluderar backlogförändring i antal ger ultimoskattningar som ligger nära referensmodellen. Detta indikerar att just denna förklarande variabel har en tydlig inverkan på resultaten. Skillnaderna är särskilt framträdande för den sista skademånaden, representerad av den sista punkten i figuren. Här är avvikelseerna mellan modellvarianterna som störst. Den GLM modellvariant som inkluderar backlogförändring i procent ger en ultimoskattning som i stort sett sammanfaller med den kombinerade modellstrukturen utan backlogförändring som förklarande variabel. Detta indikerar att den procentuella förändringen inte har någon större effekt på skattningen av utvecklingsfaktorer och därmed även på ultimoskattningen.

Modellvarianterna som inkluderar förändring i backlogantalet avviker tydligt från övriga modellvarianter, vilket framgår både av figur 5 och figur 6. Denna avvikelse är särskilt anmärkningsvärd eftersom den sista skademånaden generellt präglas av störst osäkerhet i ultimoskattningen. Detta beror på att den skademånaden omfattar minst tillgänglig information samt en hög andel oreglerade skador. Att modellvarianterna med backlogförändring i antal uppvisar en sådan avvikelse

indikerar att denna förklarande variabel har en betydande påverkan på de skattade utvecklingsfaktorerna. Det faktum att dessa modeller skiljer sig från övriga kan tolkas på två sätt. Det kan dels tyda på en viss grad av överanpassning till de observerade backlognivåerna, vilket i så fall skulle förklara modellens avvikelse från övriga varianter. Dels kan det indikera en förbättrad modellpassning, där modellen i högre utsträckning fångar ett faktiskt samband mellan backlogförändring och det inkrementella antalet behandlade skador.

Med utgångspunkt i avsnitt 4.4, där de prediktiva jämförelsemått för modellvarianterna presenteras, framgår att GLM modellvarianter generellt uppvisar lägre prediktionsfel än random forest. Detta stärker tolkningen att GLM i större utsträckning fångar ett verkligt samband i data, snarare än att överanpassa. Det indikerar därmed på att GLM modellen är bättre på att identifiera ett faktiskt samband mellan backlognivå och utvecklingsfaktorerna.



Figur 6: Illustration av differensen i ultimoskattningen mellan samtliga modellvarianter och GLM diagonalmodellen för de 15 sista skademånaderna.

4.6 Kassaflöden för framtida perioder

Vid bokslut är det ofta önskvärt att snabbt kunna ta fram tillförlitliga prognoser över framtida antalskassaflöde baserat på backlognivån vid kalendermånadens slut. Den metodik som idag använts saknar dock ett effektivt sätt att direkt koppla en

given backlognivå till ett utvecklingsmönster. Detta skulle resultera i ett värdefullt verktyg inför bokslutsarbetet som skulle effektivisera skattningen av framtida antalskassaflöde när backlog siffran presenteras.

För att förenkla processen och förbättra träffsäkerheten i skattningen av framtida antalskassaflöden har en modell utvecklats där backlognivån används som ingångsvärde. Modellen syftar till att vid bokslutstidpunkten kunna skatta utvecklingsmönstret för den framtida perioden, enbart baserat på den aktuella backlognivån.

Beräkningen av antalskassaflödet baseras på modeller som härletts från tidigare anpassade GLM eller random forest modeller, en för varje utvecklingsmånad. Dessa modeller har sparats för att möjliggöra en smidig och repeterbar tillämpning vid prognostillfällen. Istället för att enbart använda chain ladder utvecklingsfaktorer, tillämpas GLM eller random forest baserade utvecklingsfaktorer från radeffektsmodellen för de fem första utvecklingsmånaderna. För de senare utvecklingsmånaderna används däremot chain ladder utvecklingsfaktorer, eftersom dessa perioder bedöms vara mer stabila och saknar tydlig koppling till backlognivån. I modellen för framtida period används enbart radeffektsmodellen för att prediktera antalskassaflödet. Då den explicit beskriver hur utvecklingsmönstret för en viss skademånad påverkas av backlognivån.

I denna tillämpning antas att radeffektsmodellen, skattad via en uppsättning random forest- eller GLM modeller, kan användas för att översätta backlognivåer till antalskassaflöden. Skattningen baseras på ett utvecklingsmönster uttryckt genom utvecklingsfaktorer f_j , där j representerar utvecklingsmånad. De inkrementella antalskassaflödena i respektive utvecklingsmånad betecknas med X_j , och beräknas utifrån ett uppskattat ultimoantal $\bar{U}$ samt de skattade utvecklingsfaktorerna $\hat{f}_j$.

Vi definierar den kumulativa utvecklingsfaktorn från utvecklingsmånad j till slutet som

$$\prod_{k=j}^{J-1} \hat{f}_k.$$

Den inkrementella andelen skador som behandlas i utvecklingsmånad j definieras med hjälp av uttryck från Engler och Lindskog [6, s. 315]. Enligt deras resultat konvergerar följande sekvens nästan säkert mot de sanna inkrementella andelarna $(p_1, p_2, \dots, p_J)$

$$\left(\frac{1}{\prod_{s=1}^{J-1} \hat{f}_s}, \frac{\hat{f}_1 - 1}{\prod_{s=1}^{J-1} \hat{f}_s}, \frac{(\hat{f}_2 - 1)\hat{f}_1}{\prod_{s=1}^{J-1} \hat{f}_s}, \dots, \frac{(\hat{f}_{J-1} - 1) \prod_{s=1}^{J-2} \hat{f}_s}{\prod_{s=1}^{J-1} \hat{f}_s} \right) \xrightarrow{\text{a.s.}} (p_1, p_2, \dots, p_J).$$

Utifrån detta uttryck kan den estimerade inkrementella andelen $\hat{p}_j$ omformuleras och delas upp i två fall. Eftersom det för den första utvecklingsperioden ($j = 1$) inte finns någon föregående utvecklingsmånad att koppla en faktor $\hat{f}_{j-1}$ till (eftersom $j - 1 = 0$ är odefinierad), och produkten $\prod_{s=1}^{j-2} \hat{f}_s$ blir då en tom produkt med värde 1. Därav beräknas andelen för den första utvecklingsmånaden enligt

$$\hat{p}_1 = \frac{1}{\prod_{k=1}^{J-1} \hat{f}_k}.$$

För efterföljande utvecklingsmånader ($j = 2, \dots, J$) beräknas andelen enligt

$$\hat{p}_j = \frac{(\hat{f}_{j-1} - 1) \cdot \prod_{s=1}^{j-2} \hat{f}_s}{\prod_{s=1}^{J-1} \hat{f}_s}.$$

Slutligen multipliceras varje andel $\hat{p}_j$ med den skattade ultimnivån $\bar{U}$ för att få det inkrementella antalskassaflödet

$$\hat{X}_j = \bar{U} \cdot \hat{p}_j.$$

Där $\bar{U}$ är den skattade ultimnivån. Det är värt att notera att summan av de skattade inkrementella antalskassaflödena över samtliga utvecklingsmånader återger ultimnivån, det vill säga

$$\sum_{j=1}^J \hat{X}_j = \bar{U},$$

vilket följer av att summan av andelarna $\hat{p}_j$ summerar till 1. De beräknade antalskassaflödena $\hat{X}_j$ kan samtidigt tolkas som en approximation av storleken på den inkrementella reserven för respektive framtida utvecklingsmånad.

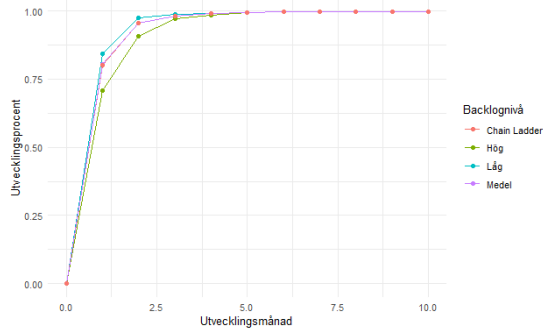
Ultimoantalet skattas som ett medelvärde baserat på de 12 senaste utvecklingsmånadernas historiska data. Denna metod har valts eftersom variationen i

observerade ultimoantal är relativt låg, vilket gör medelvärdet till ett stabilt och rimligt antagande. Det bör dock understrykas att alternativa skattningsmetoder kan användas, exempelvis baserade på intjänad premie.

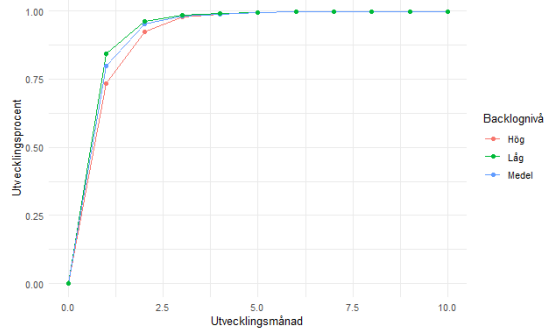
I denna modell används enbart radeffekten från olika modellvarianter för att prediktera utvecklingsmönstret givet en ny angiven backlognivå. Eftersom samtliga GLM modellvarianter har anpassats med backlognivån som enda förklarande variabel är radeffektsmodellen identisk mellan dem. Därför jämförs radeffektsmodellen i GLM direkt med motsvarande modellstruktur i random forest. Som visas i figur 7 ger modellerna mycket liknande resultat vid låga till medelhöga backlognivåer. Vid höga backlognivåer uppstår däremot tydligare skillnader, där GLM modellen tenderar att generera större variationer i utvecklingsprocent än random forest. Detta tyder på att GLM är mer känslig för variationer i backlognivån, medan random forest uppvisar ett mer dämpat och stabilt mönster.

I figuren illustreras hur utvecklingsprocenten varierar med backlognivån. Utvecklingsprocenten beräknas som $\frac{1}{f_j}$, där f_j är den skattade utvecklingsfaktorn från modellen baserad på backlognivå. De tre valda backlognivåerna är baserade på historiska observationer, låg motsvarar den approximativt lägsta observerade backlognivån (500), medel motsvarar backlogens medelvärde (2000), och hög den approximativt högsta backlognivån (6000).

Figuren visar att en hög backlognivå leder till en lägre utvecklingsprocent i den första utvecklingsmånaden jämfört med lägre backlognivåer. Däremot är förändringen i utvecklingsprocent mellan första och andra utvecklingsmånaden är större vid hög backlog. Detta innebär att en lägre initial utvecklingsprocent kompenseras under nästkommande utvecklingsmånader. Efter fyra utvecklingsmånader är utvecklingsprocenten i stort sett densamma oavsett backlognivå. Utvecklingsprocenten för chain ladder metoden överensstämmer med den som ges av modellen för medel backlognivån (2000), vilket framgår av figur 7a. Detta är i linje med teorin eftersom chain ladder metoden implicit baserar sina skattade utvecklingsfaktorer på ett historiskt genomsnitt av backlognivåerna.



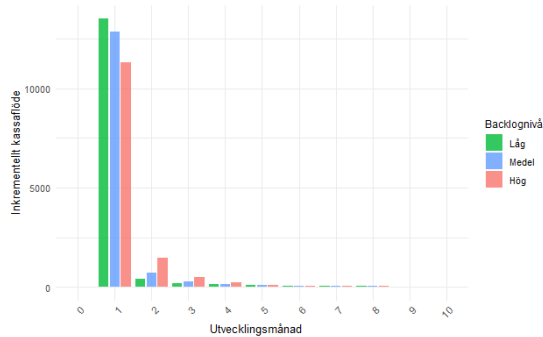
(a) Utvecklingsprocent för olika backlognivåer utifrån GLM modell.



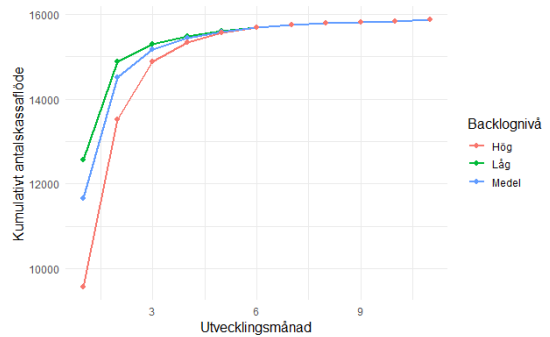
(b) Utvecklingsprocent för olika backlognivåer utifrån random forest modell.

Figur 7: Jämförelse av utvecklingsprocent vid olika backlognivåer, där hög, medel och låg nivå definieras som 6000, 2000 respektive 500. Utvecklingsprocenten är definierad enligt $\frac{1}{f_j}$.

I figur 8 och figur 9 illustreras hur antalskassaflödet varierar med backlognivån. Mönstret överensstämmer med det som observerats för utvecklingsprocenten, en hög backlognivå medför ett lägre antalskassaflöde i den första utvecklingsmånaden, men ett högre antalskassaflöde i den andra perioden jämfört med lägre backlognivåer. Detta mönster kan även observeras i historiska data och bekräftas av korrelationerna som redovisas i avsnitt 4.2.



(a) Inkrementellt antalskassaflöde.



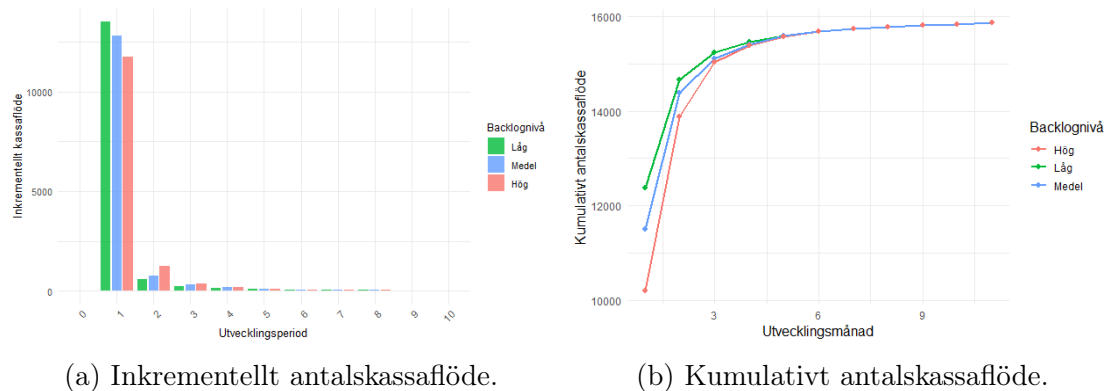
(b) Kumulativt antalskassaflöde.

Figur 8: Jämförelse av antalskassaflöden vid olika backlognivåer, baserat på GLM radeffektmodellen. Hög, medel och låg backlognivå definieras som 6000, 2000 respektive 500.

Vidare kan det vara informativt att analysera hur det kumulativa antalskassaflödet utvecklas. Detta illustreras i figur 8b, där endast de första tio utvecklingsmånaderna visas för att tydligare visualisera antalskassaflödenas initiala utveck-

ling. Trots att skillnaderna i antalskassaflöde mellan backlognivåerna är betydande under de första utvecklingsmånaderna, visar det sig att det kumulativa antalskassaflödena konvergerar efter ungefär fyra utvecklingsmånader. Det innebär att backlognivån främst påverkar tidsfördelningen av kassaflödet snarare än den totala volymen. För praktiska tillämpning, exempelvis reservsättning och kassaflödesmodellering, är denna information central. En hög backlognivå signalerar visserligen en initial förskjutning, men inte nödvändigtvis ett lägre ultimo.

Denna utveckling speglar även modellens underliggande egenskaper. Eftersom antalskassaflödena återhämtar sig snabbt efter initiala fördröjningar tyder det på att modellen fångar temporära kapacitetsbegränsningar snarare än långsiktiga strukturella förändringar. Den totala volymen påverkas inte, utan endast tidpunkten för när skadorna hanteras. Detta stärker modellens trovärdighet och användbarhet som prediktivt verktyg i situationer där det är avgörande att fånga när i tiden kassaflödena inträffar, exempelvis vid kvartalsrapportering eller reservanalys. Det bekräftar att modellen beskriver en omfördelning över tid snarare än en kvantitativ skillnad, vilket ligger i linje med det mönster som även observeras i data.



Figur 9: Jämförelse av antalskassaflöden vid olika backlognivåer, baserat på random forest radeffektmodellen. Hög, medel och låg backlognivå definieras som 6000, 2000 respektive 500.

Vid en jämförelse mellan figur 8 och figur 9 kan en liknande slutsats dras som vid jämförelsen mellan GLM och random forest. Resultaten från de två metoderna är överlag likartade, men random forest uppvisar något mer stabila skattningar. Detta beror på att variationer i den förklarande variabeln, backlognivån, har mindre påverkan på skattningen av utvecklingsfaktorerna i random forest jämfört med GLM, vilket framgår av figurerna. Den minskade känsligheten leder i sin tur till att skattningen av antalskassaflödet påverkas i lägre grad.

5 Skattning av antalet rapporterade skador

Vi har tillgång till historisk data över backlognivån vid varje månadsslut, där backlogen omfattar både skador och övriga ärenden. Vidare finns information om antalet behandlade skador per skademånad i och utvecklingsmånad j (se avsnitt 2). Däremot saknas direkt information om antalet rapporterade skador per kalendermånad. Målet är att härleda en antalstriangel för rapporterade skador som möjliggör tillämpning av chain ladder. För detta introduceras variabeln R_t , vilken definieras som antalet rapporterade skador för kalendermånad t .

Vilket enligt Lindskog och Wüthrich (se [11, ss. 3–5]) kan härledas från sambandet

$$B_{t+1} = B_t + R_t - P_t,$$

där

- B_t och B_{t+1} är backlognivå vid månadssluten t respektive $t + 1$,
- P_t är antalet behandlade skador för kalendermånad t ,
- R_t är det härledda antalet rapporterade skador för kalendermånad t .

Genom att lösa ut R_t får vi

$$R_t = B_{t+1} - B_t + P_t.$$

Detta uttryck visar att antalet rapporterade skador motsvarar förändringen i backlognivån mellan två på varandra följande kalendermånader, justerat för behandlade skador under perioden. Eftersom faktisk rapporteringsdata saknas används denna modell för att härleda antalet rapporterade skador.

Syftet med skattningen av rapporterade skador är att möjliggöra tillämpning av Mack chain ladder på rapporterade skador istället för på behandlade skador. Utifrån avsnitt 4.1.1 framgår att det ej förefaller stabila utvecklingsmönster samt att skademånaderna inte är oberoende för de behandlade skadorna. Detta kan resultera i snedvridna skattningar av utvecklingsfaktorerna vid användning av Mack chain ladder metoden. Därför skattas de rapporterade skadorna i triangelform, i förhoppningen att antagandet om oberoende mellan skademånader i högre grad ska vara uppfyllt.

5.1 Skattning av rapporterade skador genom optimering

För att skatta antalet rapporterade skador i triangelformat (antalstriangel) formuleras ett optimeringsproblem där varje cell $r_{i,j} \geq 0$ representerar det inkrementella

antalet rapporterade skador för skademånad $i = 1, \dots, I$ och utvecklingsmånad $j = 1, \dots, J$. Endast de kombinationer av (i, j) som uppfyller villkoret $i+j-1 \leq T$, där T betecknar den sista observerade kalendermånad, inkluderas i optimeringen.

Målet är att finna en uppsättning rapporterade skador $r_{i,j}$ som, givet backlognivåer och behandlingskapacitet, resulterar i en modellbaserad approximation av behandlade skador som ligger så nära de observerade värdena $P_{i,j}$ som möjligt. Målfunktionen ges av

$$\sum_{\substack{i=1,\dots,I \\ j=1,\dots,J \\ i+j-1 \leq T}} \left(a_{i+j-1} \sum_{k=1}^{j-1} (r_{i,k} - P_{i,k}) + b_{i+j-1} r_{i,j} - P_{i,j} \right)^2,$$

där koefficienterna a_t och b_t beror på backlognivån och behandlade skador samt definieras enligt

$$a_t = \begin{cases} 1, & \text{om } B_t \leq P_t, \\ \frac{P_t}{B_t}, & \text{om } B_t > P_t, \end{cases}$$

$$b_t = \begin{cases} \frac{P_t - B_t}{R_t}, & \text{om } B_t \leq P_t, \\ 0, & \text{om } B_t > P_t. \end{cases}$$

Här betecknar B_t backlognivån, P_t det totala antalet behandlade skador och R_t det totala antalet rapporterade skador vid kalendermånad t . Parametrarna a_t och b_t styr hur mycket vikt som läggs på tidigare residualer respektive nya rapporteringar beroende på om backlognivån är under eller över kapaciteten (antalet behandlade skador). När backlognivån är lägre än eller lika med antalet behandlade skador ($B_t \leq P_t$) finns ingen restbacklog kvar, vilket motsvarar

$$a_t = 1,$$

$$b_t = \frac{P_t - B_t}{R_t}.$$

Om backlognivån överskrider kapaciteten ($B_t > P_t$) reduceras vikten på tidigare residualer proportionellt mot kapacitetsunderskottet ($a_t = \frac{P_t}{B_t}$) och sätter $b_t = 0$ för att markera att ingen tillgänglig kapacitet återstår för nya rapporteringar

$$a_t = \frac{P_t}{B_t},$$

$$b_t = 0.$$

Sammanfattningsvis justerar a_t vikten på tidigare residualer baserat på om och hur mycket backlognivån överskrider kapaciteten, medan b_t representerar den relativa andel av ny rapportering som kan antas bli behandlad.

Optimeringen utförs under följande restriktioner, för varje kalendermånad $t = 1, \dots, T$ kräver modellen att summan av de skattade rapporteringarna längs varje diagonal motsvarar det totala antalet observerade rapporterade skador R_t enligt

$$\sum_{(i,j):i+j-1=t} \hat{r}_{i,j} = R_t.$$

Samtidigt som alla $\hat{r}_{i,j}$ måste vara icke-negativa

$$\hat{r}_{i,j} \geq 0 \quad \text{för alla relevanta } (i,j).$$

För att underlätta numerisk lösning och möjliggöra viss flexibilitet införs upp mjukade restriktioner genom slackvariabler $s_t^+ \geq 0$ och $s_t^- \geq 0$. Dessa tillåter små avvikelser från restriktionerna enligt

$$\sum_{(i,j):i+j-1=t} r_{i,j} + s_t^- - s_t^+ = R_t.$$

Avvikelserna straffas i målfunktionen genom tillägg av en straffterm som väger summan av alla slackvariabler

$$\gamma \sum_{t=1}^T (s_t^+ + s_t^-),$$

där $\gamma > 0$ är en liten konstant som reglerar hur stora avvikelser som tillåts.

Den slutliga optimeringen kan därmed formuleras som

$$\min_{\mathcal{R}, s^+, s^-} \left[\sum_{\substack{i=1, \dots, I \\ j=1, \dots, J \\ i+j-1 \leq T}} \left(a_{i+j-1} \sum_{k=1}^{j-1} (r_{i,k} - P_{i,k}) + b_{i+j-1} r_{i,j} - P_{i,j} \right)^2 \right] + \gamma \sum_{t=1}^T (s_t^+ + s_t^-),$$

under de givna restriktionerna. Detta optimeringsproblem syftar till att finna en backloganpassad skattning av rapporterade skador $r_{i,j}$ baserat på observerad utveckling av behandlade skador över tid.[18]

Detta är ett kvadratisk optimeringsproblem för att skatta de okända rapporterade skadorna $r_{i,j}$ i antalstriangeln. Minimeringsproblemet har implementerats i R med hjälp av *CVXR* paketet tillsammans med användning av *ECOS* lösningsalgoritmen. Detta gör det möjligt att lösa konvexa optimeringsproblem på ett flexibelt sätt.

Målsättningen med optimeringen är att minimera två komponenter samtidigt. Den första är residualfelet, det vill säga skillnaden mellan en modellbaserad approximation av behandlade skador (baserad på rapporterade skador och backlognivåer) och de observerade behandlade skadorna.

Den andra delen är ett slackstraff, som är en liten straffterm som tillåter avvikelser (via slackvariabler) mellan de skattade diagonalsummorna och de härledda rapporterade skadorna R_t från observerad data. Slackvariablerna inför flexibilitet i modellen genom att tillåta små avvikelser från restriktionerna, vilket kan minska risken för instabilitet eller överanpassning vid numerisk lösning.

Bivillkoren utgörs av att summan av de skattade $r_{i,j}$ längs varje diagonal (kalendermånad) ska matcha det observerade antalet rapporterade skadorna per kalendermånad, justerat med slackvariablerna. Det andra bivillkoret är att alla skattade värden $r_{i,j}$ tvingas vara icke-negativa, eftersom det inte är meningsfullt med negativa antal rapporterade skador. Det slutliga resultatet blir en komplett skattad antalstriangel för rapporterade skador som kan användas för vidare analys.

5.2 Chain ladder baserat på skattade rapporterade skador

Utifrån optimeringsproblemet presenterat i avsnitt 5.1 erhöles en skattad antalstriangel. Optimeringen genomfördes genom en grid sökning av γ i intervallet $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. För varje värde beräknades både kvadratsumman av residualer (som speglar modellens passform) och summan av slackvariablerna (vilka tillåter avvikelser från diagonalkraven). Utifrån dessa mått identifierades det γ där en ytterligare minskning av slack innebar en oproportionerligt stor ökning av residualkvadratsumman. Mot bakgrund av denna avvägning valdes $\gamma = 10^{-2}$, då detta gav en marginell ökning i residualerna samtidigt som summan av slackvariablerna s_+ och s_- minskade avsevärt. Det innebär att modellen kunde uppfylla diagonalkraven i större utsträckning utan att försämma passformen nämnvärt.

En jämförelse mellan tabell 18 och tabell 19 visar att variationen mellan skademånader är mindre för rapporterade skador än för behandlade skador. Detta indikerar att rapporterade skador uppvisar mer stabila mönster över tid, vilket förbättrar förutsättningarna för att tillämpa chain ladder metoden på denna data. Dessutom ligger nivåerna för rapporterade skador genomgående högre, vilket är rimligt då backlog alltid är positiv (se avsnitt 2).

Med utgångspunkt i den skattade inkrementella antalstriangeln kan de kumulativa antalen beräknas, vilket möjliggör skattning av chain ladder utvecklingsfaktorer. Dessa presenteras i tabell 20. Utvecklingsfaktorerna jämförs sedan med motsvarande faktorer för behandlade skador, skattade med GLM radeffektsmodellen baserad på genomsnittlig backlognivå (tabell 21). Jämförelsen visar en systematisk och mycket nära överensstämmelse mellan metoderna, där skillnaderna är marginella.

Tabell 20: Skattade chain ladder utvecklingsfaktorer baserade på de skattade antalet rapporterade skadorna.

Utvecklingsfaktor	Värde	Utvecklingsfaktor	Värde	Utvecklingsfaktor	Värde
f_1	1.247671	f_2	1.046489	f_3	1.016994
f_4	1.009252	f_5	1.005232	f_6	1.003500
f_7	1.002433	f_8	1.001897	f_9	1.001543
f_{10}	1.001249	f_{11}	1.000997	f_{12}	1.000863
f_{13}	1.000636	f_{14}	1.000522	f_{15}	1.000485
f_{16}	1.000396	f_{17}	1.000301	f_{18}	1.000344
f_{19}	1.000254	f_{20}	1.000211	f_{21}	1.000231
f_{22}	1.000173	f_{23}	1.000153	f_{24}	1.000190
f_{25}	1.000160	f_{26}	1.000130	f_{27}	1.000098
f_{28}	1.000109	f_{29}	1.000080	f_{30}	1.000103
f_{31}	1.000089	f_{32}	1.000118	f_{33}	1.000079
f_{34}	1.000072	f_{35}	1.000077	f_{36}	1.000077
f_{37}	1.000065	f_{38}	1.000045	f_{39}	1.000049
f_{40}	1.000030	f_{41}	1.000060	f_{42}	1.000038
f_{43}	1.000062	f_{44}	1.000041	f_{45}	1.000063
f_{46}	1.000064	f_{47}	1.000023	f_{48}	1.000041

Tabell 21: Skattade utvecklingsfaktorer baserade på genomsnittlig backlognivå med GLM radeffektsmodellvarianten för antalet behandlade skadorna.

Utvecklingsfaktor	Värde	Utvecklingsfaktor	Värde	Utvecklingsfaktor	Värde
f_1	1.243242	f_2	1.046433	f_3	1.017667
f_4	1.009773	f_5	1.005605	f_6	1.003780
f_7	1.002660	f_8	1.002082	f_9	1.001712
f_{10}	1.001373	f_{11}	1.001124	f_{12}	1.000954
f_{13}	1.000713	f_{14}	1.000590	f_{15}	1.000539
f_{16}	1.000434	f_{17}	1.000362	f_{18}	1.000385
f_{19}	1.000282	f_{20}	1.000246	f_{21}	1.000260
f_{22}	1.000197	f_{23}	1.000187	f_{24}	1.000217
f_{25}	1.000189	f_{26}	1.000150	f_{27}	1.000125
f_{28}	1.000126	f_{29}	1.000098	f_{30}	1.000118
f_{31}	1.000110	f_{32}	1.000137	f_{33}	1.000097
f_{34}	1.000080	f_{35}	1.000088	f_{36}	1.000092
f_{37}	1.000077	f_{38}	1.000058	f_{39}	1.000061
f_{40}	1.000040	f_{41}	1.000075	f_{42}	1.000040
f_{43}	1.000077	f_{44}	1.000063	f_{45}	1.000072
f_{46}	1.000070	f_{47}	1.000052	f_{48}	1.000067

De fetmarkerade värdena i tabell 22 representerar skattade kumulativa antalet rapporterade skador för framtida perioder utifrån chain ladder metoden. Dessa används vid reservsättning för att skatta det återstående antalet skador som ännu inte rapporterats för respektive skademånad och utvecklingsmånad. Genom att jämföra rapporterade (tabell 22) och behandlade skador (tabell 23 – tabell 28) framgår att antalet rapporterade skador genomgående är fler än antalet behandlade skador. Skillnaden mellan dessa kan tolkas som backlog, det vill säga antalet skador som ännu inte hunnit hanteras av skadereglerarna. Det är därför naturligt att de två antalen skiljer sig åt, särskilt med tanke på att backlognivån historiskt aldrig varit noll.

Som framgår av tabell 22 används chain ladder metoden för att skatta det kumulativa antalet rapporterade skador för senare, ännu okända skademånader och utvecklingsmånader. Enligt avsnitt 2 kräver metoden att vissa grundläggande antaganden är uppfyllda. Två centrala antaganden är att skadeutvecklingen följer stabila utvecklingsmönster och att den är oberoende mellan olika skademånader. Med ett stabilt utvecklingsmönster avses att kvoten mellan det kumulativa antalet skador i två på varandra följande utvecklingsmånader är ungefär konstant för varje skademånad. Detta innebär att utvecklingsfaktorn f_j kan antas vara densamma oavsett skademånad. Oberoende mellan skademånader innebär att skadeutvecklingen i en viss skademånad inte påverkas av andra skademånader.

En visuell granskning av tabell 22 visar att de kumulativa antalen rapporterade skador utvecklas jämnt och konsekvent över skademånaderna, vilket stödjer antagandet om stabila utvecklingsmönster. Det finns inte heller några tydliga mönster

som tyder på att ett högt antal rapporterade skador i en skademånad påverkar efterföljande skademånader, vilket talar för att även antagandet om oberoende kan anses vara uppfyllt. Även om oberoende i strikt mening inte helt kan fastställas enbart genom visuell bedömning, observeras inga systematiska avvikelser eller tecken på beroende mellan skademånader. Sammanfattningsvis bedöms chain ladder metoden vara lämplig att tillämpa på de rapporterade skadorna, då dess centrala antaganden i huvudsak framstår som uppfyllda utifrån en visuell bedömning av data.

Tabell 22: För chain ladder metoden visas det kumulativa antalet rapporterade skador för de 15 sista skademånaderna och de 12 första utvecklingsmånaderna samt ultimo. De fetstilta cellerna avser framtida skattade värden med chain ladder metoden, medan övriga är kumulativa rapporterade skador beräknade utifrån de inkrementellt skattade rapporterade skadorna från optimeringen.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	14384	17376	17976	18210	18356	18436	18484	18517	18538	18562	18570	18591	18711
36	14705	17678	18270	18514	18676	18748	18805	18843	18858	18877	18892	18909	19038
37	14115	17101	17647	17891	18025	18108	18159	18187	18212	18226	18251	18266	18391
38	12309	14966	15443	15613	15722	15789	15820	15849	15883	15903	15922	15940	16044
39	13458	16309	16786	17010	17140	17221	17271	17301	17351	17378	17408	17425	17540
40	13765	16137	16616	16793	16899	16966	17016	17055	17097	17131	17152	17170	17282
41	13574	16503	17003	17217	17327	17416	17476	17521	17567	17594	17616	17634	17749
42	14294	17274	17797	18015	18168	18267	18335	18388	18423	18451	18474	18493	18614
43	14996	18080	18687	18984	19181	19294	19382	19429	19466	19496	19520	19540	19668
44	13367	16441	17098	17404	17569	17682	17744	17787	17821	17848	17871	17888	18006
45	14047	17313	18212	18517	18723	18821	18887	18933	18969	18998	19022	19041	19165
46	14989	19043	19935	20299	20487	20594	20666	20716	20756	20788	20814	20834	20971
47	15332	19293	20170	20513	20703	20811	20884	20934	20974	21007	21033	21054	21192
48	14731	18625	19491	19822	20005	20110	20181	20230	20268	20299	20325	20345	20478
49	14226	17749	18575	18890	19065	19165	19232	19279	19315	19345	19369	19388	19515

6 Diskussion

Precis som anges i avsnitt 4.1.1 – 4.1.4 antas att responsvariabeln är gammafördelad. Detta antagande motiveras av att de utvecklingsfaktorer som modelleras i denna tillämpning utgörs av decimaltal mellan ett och två, därmed är både icke-negativa och kontinuerliga. Utvecklingsfaktorerna speglar förändringen i antalet behandlade skador mellan två efterföljande utvecklingsmånader. Detta resulterar i att de inte lämpar sig för diskreta fördelningar.

Gammafördelningen är särskilt lämplig för att modellera kontinuerliga och positivt snedfördelade försäkringsrelaterade variabler, vilka ofta uppvisar just dessa egenskaper enligt [5, ss. 120–128]. Detta gör gammafördelningen till ett naturligt val i aktuariella tillämpningar av detta slag.

Poissonfördelningen är däremot en diskret sannolikhetsfördelning och används främst för att modellera antal händelser uttryckta i heltal. Eftersom den endast kan generera heltalsvärden vid prediktion, är den olämplig när responsvariabeln är kontinuerlig, såsom vid modellering av utvecklingsfaktorer.

Även normalfördelningen övervägdes som alternativ, då den liksom gammafördelningen kan hantera kontinuerlig data. Den tillåter dock även negativa värden, vilket inte är förenligt med utvecklingsfaktorer, som definitionsmässigt inte kan anta negativa värden. Eftersom modellresultaten för gamma- och normalfördelning var relativt likvärdiga, valdes i denna tillämpning att anta en gammafördelad responsvariabel. Detta val motiveras dels av gammafördelningens bättre överensstämmelse med datans egenskaper, dels med anledning av att Länsförsäkringar använder denna fördelning vid modellering av utvecklingsfaktorer. [5, ss. 120–128]

Trots att de prediktiva jämförelsemåtten enligt avsnitt 4.4 visar att samtliga modellvarianter uppvisar mycket god prediktiv förmåga, är det viktigt att betona att även små skillnader i modellernas precision kan få betydande konsekvenser i praktiken. Eftersom modellerna används för att beräkna reserver av mycket stora belopp kan en till synes marginell avvikelse, exempelvis en halv procent, resultera i en stor penningmässig skillnad. Det är därför särskilt positivt att MAPE ligger under 0.3 procent för radeffektsmodellen och under 0.15 procent för diagonaleffektsmodellen i den slutgiltigt valda modellvarianten. Detta indikerar att de genomsnittliga procentuella felen är mycket små. Denna låga nivå av prediktionsfel är av stor vikt eftersom felen tenderar att amplifieras i takt med att reservsättningsmetoderna appliceras på större portföljer. Detta innebär att modellernas höga noggrannhet inte bara är statistiskt tillfredsställande utan också praktiskt betydelsefull för att avsätta korrekta reserver. Det gör modellvarianten väl lämpad för användning inom reservsättning i praktiken.

En möjlig förklaring till att GLM modellvarianterna uppvisar bättre prediktiv förmåga än random forest i denna tillämpning, vilket framgår av avsnitt 4.4, är den begränsade datamängden samt det relativt lilla antalet förklarande variabler.

Random forest är visserligen kraftfull när det gäller att fånga komplexa och icke-linjära samband, men för att modellen verkligen ska komma till sin rätt krävs ofta större datamaterial och ett större antal förklarande variabler.

I detta fall, där sambanden är tydliga och antalet förklarande variabler begränsat, har enklare modeller som GLM en fördel. Den linjära ansatsen är tillräcklig för att fånga strukturen i data och samtidigt minska risken för överanpassning. Ytterligare en fördel med GLM är dess höga grad av tolkbarhet, varje parameter kan direkt kopplas till effekten av en specifik förklarande variabel, vilket gör modellen enkel att förstå och kommunicera. Detta står i kontrast till random forest, där den prediktiva processen bygger på medelvärdet av ett stort antal olika träd, vilket gör det svårt att isolera och tolka enskilda förklarande variablers inverkan på utfallet. Den höga flexibilitet som random forest erbjuder kan då snarare bli en nackdel, eftersom modellen riskerar att identifiera mönster som inte är överförbara till ny data. Det är därför inte oväntat att random forest inte överträffar GLM i detta sammanhang. Däremot finns det anledning att i framtiden undersöka hur mer avancerade modeller presterar. Detta gäller speciellt för tillämpningar med större datamaterial och med fler förklarande variabler, särskilt om dessa introducerar mer komplexitet och icke-linjära samband.

I avsnitt 5 har antalet rapporterade skador skattats med hjälp av ett kvadratisk optimeringsproblem, vilket därefter använts som underlag för tillämpning av chain ladder metoden. En central fråga är om de skattade rapporterade skadorna är lämpliga att använda för att skatta utvecklingsfaktorer, särskilt i jämförelse med alternativa metoder såsom GLM modellvarianter tillämpade på behandlade skador.

I avsnitt 5.2 presenterades utifrån en visuell granskning att antagandena för Mack chain ladder modellen i huvudsak är uppfyllda för det skattade antalet rapporterade skador. Från avsnitt 4 visades däremot att motsvarande antaganden inte är uppfyllda för det behandlade antalet skadorna. Detta beror främst på att behandlade skador påverkas av kapacitetsbegränsningar och varierande backlognivåer, vilket leder till beroenden mellan skademånader och icke-stationära utvecklingsmönster. Dessa beroenden strider mot både oberoende- och stabilitetsantagandena i Mack chain ladder modellen, och gör det därför problematiskt att direkt tillämpa chain ladder på behandlade skador.

Genom att istället fokusera på rapporterade skador, vars utveckling i högre grad påverkas av försäkringstagarnas rapporteringsbeteende snarare än av intern kapacitet, kan dessa problem undvikas. Den skattade antalstriangeln för rapporterade skador uppvisar ett jämnt och systematiskt mönster utan tydliga tecken på strukturella förändringar eller beroenden mellan skademånader. Detta talar för att skadeutvecklingen är tillräckligt stabil och uppfyller oberoendeantagandet, vilket talar för att chain ladder kan ge tillförlitliga utvecklingsfaktorer. Samtidigt är det viktigt att betona att denna bedömning baseras på en visuell granskning. En strikt

statistisk verifiering av oberoende förutsätter ytterligare tester, vilket skulle kunna utgöra ett möjligt utvecklingsområde för framtida arbete.

Att basera reservsättningen på rapporterade skador innebär två tydliga fördelar. För det första möjliggör det att chain ladder metoden kan tillämpas utan att modellering krävs av komplexa beroenden relaterade till kapacitet och backlog. För det andra erbjuder metoden en hög grad av tolkbarhet och enkelhet, vilket underlättar både förståelse och implementering. Denna ansats förutsätter dock att den initiala skattningen av rapporterade skador är rimlig. Samtidigt bör chain ladder snarare betraktas som ett komplement än som en ersättning till mer avancerade modeller baserade på antal behandlade skador. Utifrån avsnitt 4.4 framgår att GLM modellvarianterna är mycket väl lämpade för att prediktera utvecklingsfaktorer, vilket bekräftas av modellernas höga träffsäkerhet enligt de prediktiva jämförelsemåtten. Med anledning av de mycket låga prediktiva felen motiveras att GLM modellvarianterna är väl lämpade för att modellera utvecklingsfaktorer baserat på antalet behandlade skador. Samtidigt kan även användningen av chain ladder metoden på det skattade antalet rapporterade skador ses som ett lämpligt alternativ. Detta särskilt då metodens antaganden i huvudsak bedöms vara uppfyllda och möjliggör användning av den Mack chain ladder metoden för att skatta utvecklingsfaktorerna.

Det är dock viktigt att notera att backlog och antalet behandlade skador har olika definitioner och därmed inte är direkt jämförbara. En viktig skillnad är att backlogen även omfattar nollskador, vilket inte gäller för de behandlade skadorna. Denna skillnad påverkar härledningen av de rapporterade skadorna och kan därmed påverka noggrannheten i den skattade triangeln för det inkrementella antalet rapporterade skador. Även om detta kan påverka själva antalsnivåerna, förväntas det inte påverka utvecklingsmönstrets stabilitet. Detta eftersom andelen nollskador relativt faktiska skador antas vara ungefär konstant över tid. Därmed bedöms skillnaden inte påverka lämpligheten i att tillämpa Mack chain ladder metoden på rapporterade skador. Dock skulle definitionerna av backlog och behandlade skador behöva överensstämma för att detta ska vara praktiskt användbart, då skillnader i definitionerna annars medför en betydande risk för feltolkningar eller missvisande analyser.

7 Slutsatser

Analysen visar att det är olämpligt att tillämpa chain ladder metoden direkt på data för antalet behandlade skador när backlognivåerna varierar. Stora avvikelser från backloggenomsnittet kan leda till betydande snedvridningar i skattningen av utvecklingsfaktorerna, vilket i sin tur leder till missvisande beräkningar av antalsreserver. Däremot bedöms det vara lämpligt att använda chain ladder på den skattade antalstriangeln för rapporterade skador, eftersom antagandena bakom metoden i detta fall i huvudsak anses vara uppfyllda. En visuell granskning tyder på att både antagandet om oberoende mellan skademånader och antagandet om stabila utvecklingsmönster kan anses uppfyllt för rapporterade skador. För antalet behandlade skador är däremot dessa antaganden inte uppfyllda, vilket gör användning av chain ladder metoden mindre lämplig.

Utifrån olika jämförelsemått, både avseende prediktiv förmåga och modellernas anpassning till data, framgår att samtliga GLM modellvarianter presterar likartat. Samtliga dessa olika modellvarianter är därmed lämpliga för att prediktera utvecklingsfaktorer. Random forest modellen presterade generellt något sämre än GLM modellvarianterna enligt jämförelsemåtten. Eftersom även små avvikelser kan förstärkas när reserverna omfattar mycket stora belopp är det avgörande att välja den modell som ger högsta möjliga träffsäkerhet.

Jämförelsen av de olika modellvarianterna utgår ifrån jämförelsemått för både modellvarianternas anpassning till data samt dess prediktiva förmåga. Resultatet visar att det, utifrån jämförelsemåtten för anpassningen till data, står mellan de två kombinerade modellvarianterna *Kombinerad + Δ Backlog* och *GLM Kombinerad*. Den förstnämnda presterar bäst i utvecklingsmånad 2 till 5, medan den sistnämnda uppvisar något bättre resultat i utvecklingsmånad 1. Utvecklingsmånad 1 står för en stor andel av det totala antalet behandlade skadorna. Det inkrementella antalet skador är avsevärt högre i denna utvecklingsmånad än i övriga utvecklingsmånader. Därför kan det argumenteras för att utvecklingsmånad 1 bör tilldelas större vikt vid modellvalet. De prediktiva måtten visar resultaten att samtliga GLM modellvarianter ger träffsäkra prediktioner. Bland dessa uppnår dock modellvarianten *GLM Kombinerad* konsekvent den bästa prestandan för de prediktiva måtten, vilket gör den till det mest fördelaktiga valet. Den uppvisar genomgående bäst resultat för samtliga prediktiva mått vid utvärdering av modellvarianternas förmåga att prediktera utvecklingsfaktorer längs både diagonaler och rader. Specifikt presterar den bäst för radeffekten och delar förstaplatsen med *GLM Diagonaleffekt* vad gäller att prediktera utvecklingsfaktorer längs diagonalen. Med anledning av att syftet med modellerna är att kunna prediktera utvecklingsfaktorer läggs extra vikt vid detta. Tillsammans med att *GLM Kombinerad* bäst passar data för den första utvecklingsmånaden, vilken är den som har störst påverkan på det totala antalet behandlade skador, kommer denna modellvariant att väljas som den

slutgiltiga. Modellvarianter som inkluderar förändringsvariabler i backlog ger inte någon ytterligare förbättrad prediktiv förmåga i denna tillämpning. Vidare uppvisar random forest ingen fördel jämfört med GLM. Sammantaget talar detta för att GLM modellvarianter fortsatt bör föredras för prediktion av utvecklingsfaktorer, åtminstone baserat på tillgängliga data.

Vidare visar analysen att den framtagna metoden för att prediktera framtida antalskassaflöde, baserat på backlognivån vid bokslutstidpunkten, kan utgöra ett värdefullt verktyg i bokslutsarbetet. Kassaflödet beräknas med hjälp av en uppsättning tidigare anpassade GLM radeffektsmodeller, en för varje utvecklingsmånad. Metoden är enkel att tillämpa och skulle innebära en effektivisering i bokslutsprocessen, eftersom den möjliggör repeterbara och automatiserade prognoser. Radeffektsmodellen har dessutom visat låg avvikelse gentemot observerade utvecklingsfaktorer vid test på historisk data. Detta stärker förtroendet för modellens träffsäkerhet och dess användbarhet i situationer där snabba och tillförlitliga uppskattningar är avgörande.

Referenser

- [1] Agresti, A. (2013) *Categorical Data Analysis* (Tredje upplagan), Wiley.
- [2] Barry, A.T., et.al. (1991) "Valuation actuary symposium proceedings", *Society of Actuaries*, volym 9, ss. 409–446, <https://www.soa.org/globalassets/assets/Library/proceedings/valuation-actuary-symposium-proceedings/1985-99/1991/january/vasp91111.pdf>.
- [3] Bornhuetter, R. L. och Furguson, R. E. (1993) "The actuary and IBNR", *Proceedings of the Casualty Actuarial Society*, volym 59, ss. 181–195, https://www.casact.org/sites/default/files/database/proceed_proceed72_72181.pdf.
- [4] de Myttenaere, A., et.al. (2016) *Mean absolute percentage error for regression models*, HAL Archives Ouvertes, <https://hal.science/hal-01312590v2/document>.
- [5] De Jong, P. och Heller, G. Z. (2008) *Generalized Linear Models for Insurance Data*, Cambridge University Press.
- [6] Engler, N. och Lindskog, F. (2024) "Mack's estimator motivated by large exposure asymptotics in a compound Poisson setting", *ASTIN Bulletin*, 54, s. 310–326, <https://doi.org/10.1017/asb.2024.11>.
- [7] Hardin, J. W. och Hilbe, J. M. (2012) *Generalized Linear Models and Extensions* (Tredje upplagan), Stata Press.
- [8] Hodson, T.O. (2022) "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not", *Geoscientific Model Development*, volym 15, ss. 5481–5487, <https://gmd.copernicus.org/articles/15/5481/2022/gmd-15-5481-2022.pdf>.
- [9] James, G., et.al. (2021). *An Introduction to Statistical Learning* (Andra upplagan). Springer. <https://www.casact.org/sites/default/files/2022-12/James-G.-et-al.-2nd-edition-Springer-2021.pdf>
- [10] Kutner, M.H., et.al. (2005) *Applied Linear Statistical Models* (Femte upplagan), McGraw Hill/Irwin, https://users.stat.ufl.edu/~winner/sta4211/ALSM_5Ed_Kutner.pdf.
- [11] Lindskog, F. och Wüthrich, M. V. (2024) "Claims processing and costs under capacity constraints", arXiv:2409.09091.

- [12] Mack, T. (1993) "Distribution-free calculation of the standard error of Chain Ladder reserve estimates", *ASTIN Bulletin*, volym 23, ss.213–225, <https://doi.org/10.2143/ast.23.2.2005092>.
- [13] McCullagh, P. och Nelder, J. A. (1989) *Generalized Linear Models* (Andra upplagan), Chapman and Hall.
- [14] Mukaka, M. M. (2012). "A guide to appropriate use of correlation coefficient in medical research", *Malawi Medical Journal*, volym 24(3), ss. 69—71. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3576830/pdf/MMJ2403-0069.pdf>
- [15] Ohlsson, E. och Johansson, B. (2010) *Non-life insurance pricing with generalized linear models* (Första upplagan), Springer.
- [16] Ohlsson, E. (u.å) "Reservsättning med GLM", ss. 96–106.
- [17] Sundberg, R. (2022) *Lineära stokastiska modeller*, Stockholms Universitet.
- [18] Privat kommunikation med Filip Lindskog

Appendix

Tabell 23: För GLM diagonaleffektsmodellen presenteras det kumulativa antalet behandlade skador, där de okända cellerna har skattats utifrån utvecklingsfaktorer som estimerats av modellen. Tabellen omfattar de 15 senaste skademånaderna, de första 12 utvecklingsmånaderna samt ultimo. Fetstil markerar skattade värden, övriga är observerade.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	13327	16100	16667	16885	17031	17108	17155	17187	17210	17232	17244	17263	17384
36	13635	16387	16933	17179	17333	17404	17461	17497	17513	17532	17548	17563	17698
37	13081	15804	16345	16578	16708	16790	16840	16869	16892	16909	16930	16946	17080
38	11249	13822	14269	14436	14544	14609	14641	14672	14702	14720	14738	14755	14869
39	13069	15705	16165	16386	16514	16592	16640	16675	16716	16743	16770	16788	16915
40	12735	14971	15433	15608	15710	15776	15823	15857	15897	15928	15954	15972	16093
41	12813	15572	16052	16253	16362	16444	16495	16540	16581	16613	16636	16655	16780
42	13525	16260	16740	16952	17092	17174	17240	17288	17331	17361	17385	17404	17536
43	13832	16531	17089	17361	17523	17632	17711	17760	17797	17828	17852	17872	18007
44	11788	14405	14975	15226	15384	15486	15555	15597	15629	15656	15678	15695	15814
45	12023	14709	15424	15715	15902	15987	16047	16090	16123	16151	16173	16191	16314
46	12309	15480	16309	16639	16791	16886	16950	16995	17030	17059	17083	17102	17231
47	12057	15669	16464	16727	16891	16986	17050	17096	17131	17161	17184	17204	17334
48	13494	17021	17665	17981	18158	18260	18329	18378	18416	18447	18473	18493	18633
49	12886	15823	16573	16869	17035	17131	17196	17241	17277	17307	17331	17350	17481

Tabell 24: För GLM radeffektsmodellen presenteras det kumulativa antalet behandlade skador, där okända celler har skattats utifrån utvecklingsfaktorer som estimerats av modellen. Tabellen omfattar de första 12 utvecklingsmånaderna samt ultimo för de 15 senaste skademånaderna. Fetstil markerar skattade värden, övriga är observerade.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	13327	16100	16667	16885	17031	17108	17155	17187	17210	17232	17244	17263	17388
36	13635	16387	16933	17179	17333	17404	17461	17497	17513	17532	17548	17563	17697
37	13081	15804	16345	16578	16708	16790	16840	16869	16892	16909	16930	16946	17077
38	11249	13822	14269	14436	14544	14609	14641	14672	14702	14720	14738	14755	14866
39	13069	15705	16165	16386	16514	16592	16640	16675	16716	16743	16770	16789	16915
40	12735	14971	15433	15608	15710	15776	15823	15857	15897	15928	15950	15968	16089
41	12813	15572	16052	16253	16362	16444	16495	16540	16581	16609	16632	16651	16777
42	13525	16260	16740	16952	17092	17174	17240	17288	17324	17353	17377	17397	17528
43	13832	16531	17089	17361	17523	17632	17711	17758	17795	17826	17850	17870	18005
44	11788	14405	14975	15226	15384	15486	15544	15585	15618	15645	15666	15684	15802
45	12023	14709	15424	15715	15902	15991	16051	16094	16128	16155	16177	16195	16318
46	12309	15480	16309	16639	16802	16897	16961	17006	17041	17070	17094	17113	17242
47	12057	15669	16464	16758	16923	17018	17082	17127	17163	17192	17216	17235	17366
48	13494	17021	17828	18147	18325	18428	18498	18547	18585	18617	18643	18664	18805
49	12886	15919	16616	16899	17060	17154	17219	17265	17301	17331	17354	17374	17505

Tabell 25: För den kombinerade GLM modellvarianten presenteras det kumulativa antalet behandlade skador, där okända celler har skattats utifrån utvecklingsfaktorer som estimerats av modellen. Tabellen omfattar de första 12 utvecklingsmånaderna samt ultimo för de 15 senaste skademånaderna. Fetstil markerar skattade värden, övriga är observerade.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	13327	16100	16667	16885	17031	17108	17155	17187	17210	17232	17244	17263	17388
36	13635	16387	16933	17179	17333	17404	17461	17497	17513	17532	17548	17563	17697
37	13081	15804	16345	16578	16708	16790	16840	16869	16892	16909	16930	16946	17077
38	11249	13822	14269	14436	14544	14609	14641	14672	14702	14720	14738	14755	14866
39	13069	15705	16165	16386	16514	16592	16640	16675	16716	16743	16770	16789	16915
40	12735	14971	15433	15608	15710	15776	15823	15857	15897	15928	15950	15968	16089
41	12813	15572	16052	16253	16362	16444	16495	16540	16581	16609	16632	16651	16777
42	13525	16260	16740	16952	17092	17174	17240	17288	17324	17353	17377	17397	17528
43	13832	16531	17089	17361	17523	17632	17711	17758	17795	17826	17850	17870	18005
44	11788	14405	14975	15226	15384	15486	15544	15585	15618	15645	15666	15684	15802
45	12023	14709	15424	15715	15902	15987	16047	16090	16123	16151	16173	16191	16314
46	12309	15480	16309	16639	16791	16886	16950	16995	17030	17059	17083	17102	17231
47	12057	15669	16464	16727	16891	16986	17050	17096	17131	17161	17184	17204	17334
48	13494	17021	17665	17981	18158	18260	18329	18378	18416	18447	18473	18493	18633
49	12886	15823	16517	16798	16958	17051	17116	17161	17197	17226	17250	17269	17400

Tabell 26: För den kombinerade GLM modellvarianten med backlogförändring (antal) presenteras det kumulativa antalet behandlade skador, där okända celler har skattats utifrån utvecklingsfaktorer som estimerats av modellen. Tabellen omfattar de första 12 utvecklingsmånaderna samt ultimo för de 15 senaste skademånaderna. Fetstil markerar skattade värden, övriga är observerade.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	13327	16100	16667	16885	17031	17108	17155	17187	17210	17232	17244	17263	17388
36	13635	16387	16933	17179	17333	17404	17461	17497	17513	17532	17548	17563	17697
37	13081	15804	16345	16578	16708	16790	16840	16869	16892	16909	16930	16946	17077
38	11249	13822	14269	14436	14544	14609	14641	14672	14702	14720	14738	14755	14866
39	13069	15705	16165	16386	16514	16592	16640	16675	16716	16743	16770	16789	16915
40	12735	14971	15433	15608	15710	15776	15823	15857	15897	15928	15950	15968	16089
41	12813	15572	16052	16253	16362	16444	16495	16540	16581	16609	16632	16651	16777
42	13525	16260	16740	16952	17092	17174	17240	17288	17324	17353	17377	17397	17528
43	13832	16531	17089	17361	17523	17632	17711	17758	17795	17826	17850	17870	18005
44	11788	14405	14975	15226	15384	15486	15544	15585	15618	15645	15666	15684	15802
45	12023	14709	15424	15715	15902	16011	16072	16114	16148	16176	16198	16216	16338
46	12309	15480	16309	16639	16842	16936	17000	17045	17081	17110	17133	17153	17282
47	12057	15669	16464	16789	16954	17049	17114	17159	17195	17225	17248	17268	17398
48	13494	17021	17774	18092	18270	18372	18442	18491	18529	18561	18587	18608	18748
49	12886	16230	16941	17230	17394	17490	17556	17603	17639	17670	17694	17714	17848

Tabell 27: För den kombinerade GLM modellvarianten med procentuell backlogförändring presenteras det kumulativa antalet behandlade skador, där okända celler har skattats utifrån utvecklingsfaktorer som estimerats av modellen. Tabellen omfattar de första 12 utvecklingsmånaderna samt ultimo för de 15 senaste skademånaderna. Fetstil markerar skattade värden, övriga är observerade.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	13327	16100	16667	16885	17031	17108	17155	17187	17210	17232	17244	17263	17388
36	13635	16387	16933	17179	17333	17404	17461	17497	17513	17532	17548	17563	17697
37	13081	15804	16345	16578	16708	16790	16840	16869	16892	16909	16930	16946	17077
38	11249	13822	14269	14436	14544	14609	14641	14672	14702	14720	14738	14755	14866
39	13069	15705	16165	16386	16514	16592	16640	16675	16716	16743	16770	16789	16915
40	12735	14971	15433	15608	15710	15776	15823	15857	15897	15928	15950	15968	16089
41	12813	15572	16052	16253	16362	16444	16495	16540	16581	16609	16632	16651	16777
42	13525	16260	16740	16952	17092	17174	17240	17288	17324	17353	17377	17397	17528
43	13832	16531	17089	17361	17523	17632	17711	17758	17795	17826	17850	17870	18005
44	11788	14405	14975	15226	15384	15486	15544	15585	15618	15645	15666	15684	15802
45	12023	14709	15424	15715	15902	15993	16054	16096	16130	16157	16180	16198	16320
46	12309	15480	16309	16639	16806	16900	16964	17009	17045	17074	17097	17117	17246
47	12057	15669	16464	16748	16912	17007	17072	17117	17153	17182	17206	17225	17355
48	13494	17021	17707	18024	18201	18303	18372	18421	18460	18491	18517	18537	18677
49	12886	15843	16537	16818	16979	17072	17137	17182	17218	17248	17271	17291	17421

Tabell 28: För den kombinerade modellvarianten med backlogförändring baserad på random forest presenteras det kumulativa antalet behandlade skador, där okända celler har skattats utifrån utvecklingsfaktorer som estimerats av modellen. Tabellen omfattar de första 12 utvecklingsmånaderna samt ultimo för de 15 senaste skademånaderna. Fetstil markerar skattade värden, övriga är observerade.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	Ultimo
35	13327	16100	16667	16885	17031	17108	17155	17187	17210	17232	17244	17263	17388
36	13635	16387	16933	17179	17333	17404	17461	17497	17513	17532	17548	17563	17697
37	13081	15804	16345	16578	16708	16790	16840	16869	16892	16909	16930	16946	17077
38	11249	13822	14269	14436	14544	14609	14641	14672	14702	14720	14738	14755	14866
39	13069	15705	16165	16386	16514	16592	16640	16675	16716	16743	16770	16789	16915
40	12735	14971	15433	15608	15710	15776	15823	15857	15897	15928	15950	15968	16089
41	12813	15572	16052	16253	16362	16444	16495	16540	16581	16609	16632	16651	16777
42	13525	16260	16740	16952	17092	17174	17240	17288	17324	17353	17377	17397	17528
43	13832	16531	17089	17361	17523	17632	17711	17758	17795	17826	17850	17870	18005
44	11788	14405	14975	15226	15384	15486	15544	15585	15618	15645	15666	15684	15802
45	12023	14709	15424	15715	15902	16011	16072	16115	16148	16176	16198	16216	16339
46	12309	15480	16309	16639	16835	16929	16993	17038	17074	17103	17127	17146	17275
47	12057	15669	16464	16744	16909	17004	17068	17113	17149	17178	17202	17221	17351
48	13494	17021	17765	18083	18261	18363	18433	18482	18520	18552	18577	18598	18739
49	12886	16061	16859	17167	17363	17459	17525	17572	17608	17639	17663	17683	17816

Tabell 29: Utvecklingsfaktorer estimerade med diagonaleffektsmodellen för skademånaderna 25 till 49 och utvecklingsmånaderna 1 till 15. Gröna celler representerar utvecklingsfaktorer som har estimerats med GLM modellen med enbart diagonaleffekt. Blåa celler visar utvecklingsfaktorer som har beräknats med chain ladder metoden. Utvecklingsfaktorer under den färgmarkerade diagonalen är skattade med chain ladder, medan utvecklingsfaktorer ovanför diagonalen motsvarar observerade värden.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25	1.388972	1.041382	1.013991	1.009390	1.004433	1.003578	1.002199	1.001928	1.001626	1.001656	1.001058	1.000991	1.000627	1.000363	1.000495
26	1.204007	1.030873	1.015178	1.007111	1.004635	1.002955	1.001940	1.001865	1.001611	1.001322	1.001142	1.001141	1.000392	1.000641	1.000783
27	1.176971	1.038287	1.011986	1.009328	1.004853	1.003010	1.002045	1.002205	1.002266	1.001606	1.001309	1.001307	1.000914	1.000815	1.000749
28	1.216009	1.026856	1.014611	1.008280	1.005213	1.003339	1.002478	1.002084	1.002185	1.001442	1.001369	1.001262	1.000630	1.000490	1.000735
29	1.167180	1.038053	1.014406	1.007957	1.004812	1.003505	1.002869	1.002270	1.001831	1.001580	1.001113	1.000803	1.000617	1.000463	1.000555
30	1.247110	1.036656	1.012224	1.007426	1.004759	1.003846	1.003193	1.002476	1.001799	1.001972	1.001054	1.000772	1.000702	1.000421	1.000421
31	1.223999	1.030814	1.014993	1.008560	1.005301	1.004084	1.002944	1.002421	1.001268	1.001025	1.001295	1.000602	1.000601	1.000481	1.000360
32	1.178993	1.032793	1.015826	1.008429	1.005399	1.003106	1.003289	1.002218	1.001668	1.001185	1.001151	1.000703	1.000479	1.000383	1.000542
33	1.203139	1.030922	1.014029	1.010055	1.004396	1.004594	1.002305	1.001509	1.001399	1.000716	1.000716	1.000644	1.000643	1.000536	1.000928
34	1.191362	1.035293	1.014963	1.009119	1.005301	1.003625	1.002418	1.001787	1.001427	1.000920	1.000653	1.000622	1.000533	1.000740	1.000355
35	1.208152	1.035199	1.013092	1.008646	1.004486	1.002753	1.001891	1.001309	1.001277	1.000698	1.001123	1.000637	1.000636	1.000636	1.000539
36	1.201788	1.033339	1.014555	1.008925	1.004106	1.003277	1.002068	1.000927	1.001076	1.000895	1.000835	1.000894	1.000893	1.000590	1.000539
37	1.208225	1.034236	1.014214	1.007860	1.004917	1.002992	1.001678	1.001396	1.000991	1.001238	1.000927	1.001143	1.000713	1.000590	1.000539
38	1.228740	1.032408	1.011662	1.007504	1.004461	1.002149	1.002145	1.002069	1.001175	1.001280	1.001101	1.000954	1.000713	1.000590	1.000539
39	1.201722	1.029254	1.013661	1.007856	1.004721	1.002870	1.002107	1.002448	1.001597	1.001625	1.001124	1.000954	1.000713	1.000590	1.000539
40	1.175583	1.030864	1.011325	1.006538	1.004197	1.002985	1.002150	1.002574	1.001942	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
41	1.215356	1.030815	1.012518	1.006729	1.005021	1.003087	1.002697	1.002499	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
42	1.202151	1.029576	1.012629	1.008242	1.004837	1.003839	1.002762	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
43	1.195135	1.033744	1.015923	1.009314	1.006212	1.004511	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
44	1.222010	1.039634	1.016703	1.010379	1.006633	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
45	1.223407	1.048597	1.018863	1.011888	1.005360	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
46	1.257620	1.053544	1.020214	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
47	1.299653	1.050695	1.015995	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
48	1.261411	1.037845	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
49	1.227281	1.047384	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539

Tabell 30: Utvecklingsfaktorer estimerade med radeffektsmodellen för skademånaderna 25 till 49 och utvecklingsmånaderna 1 till 15. Lila celler representerar utvecklingsfaktorer som har estimerats med GLM modellvarianten med enbart radeffekt. Blåa celler visar utvecklingsfaktorer som har beräknats med chain ladder metoden. Samtliga utvecklingsfaktorer under den färgmarkerade diagonalen är också skattade enligt chain ladder metoden, medan samtliga utvecklingsfaktorer ovanför representerar observerade utvecklingsfaktorer.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25	1.388972	1.041382	1.013991	1.009390	1.004433	1.003578	1.002199	1.001928	1.001626	1.001656	1.001058	1.000991	1.000627	1.000363	1.000495
26	1.204007	1.030873	1.015178	1.007111	1.004635	1.002955	1.001940	1.001865	1.001611	1.001322	1.001142	1.001141	1.000392	1.000641	1.000783
27	1.176971	1.038287	1.011986	1.009328	1.004853	1.003010	1.002045	1.002205	1.002266	1.001606	1.001309	1.001307	1.000914	1.000815	1.000749
28	1.216009	1.026856	1.014611	1.008280	1.005213	1.003339	1.002478	1.002084	1.002185	1.001442	1.001369	1.001262	1.000630	1.000490	1.000735
29	1.167180	1.038053	1.014406	1.007957	1.004812	1.003505	1.002869	1.002270	1.001831	1.001580	1.001113	1.000803	1.000617	1.000463	1.000555
30	1.247110	1.036656	1.012224	1.007426	1.004759	1.003846	1.003193	1.002476	1.001799	1.001972	1.001054	1.000772	1.000702	1.000421	1.000421
31	1.223999	1.030814	1.014993	1.008560	1.005301	1.004084	1.002944	1.002421	1.001268	1.001025	1.001295	1.000602	1.000601	1.000481	1.000360
32	1.178993	1.032793	1.015826	1.008429	1.005399	1.003106	1.003289	1.002218	1.001668	1.001185	1.001151	1.000703	1.000479	1.000383	1.000542
33	1.203139	1.030922	1.014029	1.010055	1.004396	1.004594	1.002305	1.001509	1.001399	1.000716	1.000716	1.000644	1.000643	1.000536	1.000928
34	1.191362	1.035293	1.014963	1.009119	1.005301	1.003625	1.002418	1.001787	1.001427	1.000920	1.000653	1.000622	1.000533	1.000740	1.000355
35	1.208152	1.035199	1.013092	1.008646	1.004486	1.002753	1.001891	1.001309	1.001277	1.000698	1.001123	1.000637	1.000636	1.000636	1.000539
36	1.201788	1.033339	1.014555	1.008925	1.004106	1.003277	1.002068	1.000927	1.001076	1.000895	1.000835	1.000894	1.000893	1.000590	1.000539
37	1.208225	1.034236	1.014214	1.007860	1.004917	1.002992	1.001678	1.001396	1.000991	1.001238	1.000927	1.001143	1.000713	1.000590	1.000539
38	1.228740	1.032408	1.011662	1.007504	1.004461	1.002149	1.002145	1.002069	1.001175	1.001280	1.001101	1.000954	1.000713	1.000590	1.000539
39	1.201722	1.029254	1.013661	1.007856	1.004721	1.002870	1.002107	1.002448	1.001597	1.001625	1.001124	1.000954	1.000713	1.000590	1.000539
40	1.175583	1.030864	1.011325	1.006538	1.004197	1.002985	1.002150	1.002574	1.001942	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
41	1.215356	1.030815	1.012518	1.006729	1.005021	1.003087	1.002697	1.002499	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
42	1.202151	1.029576	1.012629	1.008242	1.004837	1.003839	1.002762	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
43	1.195135	1.033744	1.015923	1.009314	1.006212	1.004511	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
44	1.222010	1.039634	1.016703	1.010379	1.006633	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
45	1.223407	1.048597	1.018863	1.011888	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
46	1.257620	1.053544	1.020214	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
47	1.299653	1.050695	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
48	1.261411	1.047384	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
49	1.235335	1.043810	1.017022	1.009528	1.005516	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539

Tabell 31: Utvecklingsfaktorer estimerade med den kombinerade GLM modell-varianten för skademånaderna 25 till 49 och utvecklingsmånaderna 1 till 15. Lila celler representerar utvecklingsfaktorer estimerade med radeffektsmodellen, medan gröna celler representerar utvecklingsfaktorer estimerade med diagonaleffektsmodellen. Blåa celler visar utvecklingsfaktorer som har beräknats med chain ladder metoden. Samtliga utvecklingsfaktorer under den färgmarkerade diagonalen som inte är färgade har också skattats enligt chain ladder metoden, medan samtliga utvecklingsfaktorer ovanför representerar observerade utvecklingsfaktorer.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25	1.388972	1.041382	1.013991	1.009390	1.004433	1.003578	1.002199	1.001928	1.001626	1.001656	1.001058	1.000991	1.000627	1.000363	1.000495
26	1.204007	1.030873	1.015178	1.007111	1.004635	1.002955	1.001940	1.001865	1.001611	1.001322	1.001142	1.001141	1.000392	1.000641	1.000783
27	1.176971	1.038287	1.011986	1.009328	1.004853	1.003010	1.002045	1.002205	1.002266	1.001606	1.001309	1.001307	1.000914	1.000815	1.000749
28	1.216009	1.026856	1.014611	1.008280	1.005213	1.003339	1.002478	1.002084	1.002185	1.001442	1.001369	1.001262	1.000630	1.000490	1.000735
29	1.167180	1.038053	1.014406	1.007957	1.004812	1.003505	1.002869	1.002270	1.001831	1.001580	1.001113	1.000803	1.000617	1.000463	1.000555
30	1.247110	1.036656	1.012224	1.007426	1.004759	1.003846	1.003193	1.002476	1.001799	1.001972	1.001054	1.000772	1.000702	1.000421	1.000421
31	1.223999	1.030814	1.014993	1.008560	1.005301	1.004084	1.002944	1.002421	1.001268	1.001025	1.001295	1.000602	1.000601	1.000481	1.000360
32	1.178993	1.032793	1.015826	1.008429	1.005399	1.003106	1.003289	1.002218	1.001668	1.001185	1.001151	1.000703	1.000479	1.000383	1.000542
33	1.203139	1.030922	1.014029	1.010055	1.004396	1.004594	1.002305	1.001509	1.001399	1.000716	1.000716	1.000644	1.000643	1.000536	1.000928
34	1.191362	1.035293	1.014963	1.009119	1.005301	1.003625	1.002418	1.001787	1.001427	1.000920	1.000653	1.000622	1.000533	1.000740	1.000355
35	1.208152	1.035199	1.013092	1.008646	1.004486	1.002753	1.001891	1.001309	1.001277	1.000698	1.001123	1.000637	1.000636	1.000636	1.000539
36	1.201788	1.033339	1.014555	1.008925	1.004106	1.003277	1.002068	1.000927	1.001076	1.000895	1.000835	1.000894	1.000893	1.000590	1.000539
37	1.208225	1.034236	1.014214	1.007860	1.004917	1.002992	1.001678	1.001396	1.000991	1.001238	1.000927	1.001143	1.000713	1.000590	1.000539
38	1.228740	1.032408	1.011662	1.007504	1.004461	1.002149	1.002145	1.002069	1.001175	1.001280	1.001101	1.000954	1.000713	1.000590	1.000539
39	1.201722	1.029254	1.013661	1.007856	1.004721	1.002870	1.002107	1.002448	1.001597	1.001625	1.001124	1.000954	1.000713	1.000590	1.000539
40	1.175583	1.030864	1.011325	1.006538	1.004197	1.002985	1.002150	1.002574	1.001942	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
41	1.215356	1.030815	1.012518	1.006729	1.005021	1.003087	1.002697	1.002499	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
42	1.202151	1.029576	1.012629	1.008242	1.004837	1.003839	1.002762	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
43	1.195135	1.033744	1.015923	1.009314	1.006212	1.004511	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
44	1.222010	1.039634	1.016703	1.010379	1.006633	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
45	1.223407	1.048597	1.018863	1.011888	1.005360	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
46	1.257620	1.053544	1.020214	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
47	1.299653	1.050695	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
48	1.261411	1.037845	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
49	1.227281	1.043810	1.017022	1.009528	1.005516	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539

Tabell 32: Utvecklingsfaktorer estimerade med den kombinerade GLM modellvarianten, där förändring i backlog (antal) ingår som förklarande variabel i diagonaleffektsmodellen. Utvecklingsfaktorerna presenteras för skademånaderna 25 till 49 och utvecklingsmånaderna 1 till 15. Lila celler representerar utvecklingsfaktorer estimerade med radeffektsmodellen, medan gröna celler representerar utvecklingsfaktorer estimerade med diagonaleffektsmodellen. Blåa celler visar utvecklingsfaktorer som har beräknats med chain ladder metoden. Samtliga utvecklingsfaktorer under den färgmarkerade diagonalen och som inte är färgade har också skattats enligt chain ladder metoden, medan samtliga utvecklingsfaktorer ovanför representerar observerade utvecklingsfaktorer.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25	1.388972	1.041382	1.013991	1.009390	1.004433	1.003578	1.002199	1.001928	1.001626	1.001656	1.001058	1.000991	1.000627	1.000363	1.000495
26	1.204007	1.030873	1.015178	1.007111	1.004635	1.002955	1.001940	1.001865	1.001611	1.001322	1.001142	1.001141	1.000392	1.000641	1.000783
27	1.176971	1.038287	1.011986	1.009328	1.004853	1.003010	1.002045	1.002205	1.002266	1.001606	1.001309	1.001307	1.000914	1.000815	1.000749
28	1.216009	1.026856	1.014611	1.008280	1.005213	1.003339	1.002478	1.002084	1.002185	1.001442	1.001369	1.001262	1.000630	1.000490	1.000735
29	1.167180	1.038053	1.014406	1.007957	1.004812	1.003505	1.002869	1.002270	1.001831	1.001580	1.001113	1.000803	1.000617	1.000463	1.000555
30	1.247110	1.036656	1.012224	1.007426	1.004759	1.003846	1.003193	1.002476	1.001799	1.001972	1.001054	1.000772	1.000702	1.000421	1.000421
31	1.223999	1.030814	1.014993	1.008560	1.005301	1.004084	1.002944	1.002421	1.001268	1.001025	1.001295	1.000602	1.000601	1.000481	1.000360
32	1.178993	1.032793	1.015826	1.008429	1.005399	1.003106	1.003289	1.002218	1.001668	1.001185	1.001151	1.000703	1.000479	1.000383	1.000542
33	1.203139	1.030922	1.014029	1.010055	1.004396	1.004594	1.002305	1.001509	1.001399	1.000716	1.000716	1.000644	1.000643	1.000536	1.000928
34	1.191362	1.035293	1.014963	1.009119	1.005301	1.003625	1.002418	1.001787	1.001427	1.000920	1.000653	1.000622	1.000533	1.000740	1.000355
35	1.208152	1.035199	1.013092	1.008646	1.004486	1.002753	1.001891	1.001309	1.001277	1.000698	1.001123	1.000637	1.000636	1.000636	1.000539
36	1.201788	1.033339	1.014555	1.008925	1.004106	1.003277	1.002068	1.000927	1.001076	1.000895	1.000835	1.000894	1.000893	1.000590	1.000539
37	1.208225	1.034236	1.014214	1.007860	1.004917	1.002992	1.001678	1.001396	1.000991	1.001238	1.000927	1.001143	1.000713	1.000590	1.000539
38	1.228740	1.032408	1.011662	1.007504	1.004461	1.002149	1.002145	1.002069	1.001175	1.001280	1.001101	1.000954	1.000713	1.000590	1.000539
39	1.201722	1.029254	1.013661	1.007856	1.004721	1.002870	1.002107	1.002148	1.001597	1.001625	1.001124	1.000954	1.000713	1.000590	1.000539
40	1.175583	1.030864	1.011325	1.006538	1.004197	1.002985	1.002150	1.002574	1.001942	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
41	1.215356	1.030815	1.012518	1.006729	1.005021	1.003087	1.002697	1.002499	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
42	1.202151	1.029576	1.012629	1.008242	1.004837	1.003839	1.002762	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
43	1.195135	1.033744	1.015923	1.008314	1.006212	1.004511	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
44	1.222010	1.039634	1.016703	1.010379	1.006633	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
45	1.223407	1.048597	1.018863	1.011888	1.006882	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
46	1.257620	1.053544	1.020214	1.012175	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
47	1.299653	1.050695	1.019773	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
48	1.261411	1.044206	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
49	1.256673	1.043810	1.017022	1.009528	1.005516	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539

Tabell 33: Utvecklingsfaktorer estimerade med den kombinerade GLM modellvarianten, där förändring i backlog (procent) ingår som förklarande variabel i diagonaleffektsmodellen. Utvecklingsfaktorerna presenteras för skademånaderna 25 till 49 och utvecklingsmånaderna 1 till 15. Lila celler representerar utvecklingsfaktorer estimerade med radeffektsmodellen, medan gröna celler representerar utvecklingsfaktorer estimerade med diagonaleffektsmodellen. Blåa celler visar utvecklingsfaktorer som har beräknats med chain ladder metoden. Samtliga utvecklingsfaktorer under den färgmarkerade diagonalen och som inte är färgade har också skattats enligt chain ladder metoden, medan samtliga utvecklingsfaktorer ovanför representerar observerade utvecklingsfaktorer.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25	1.388972	1.041382	1.013991	1.009390	1.004433	1.003578	1.002199	1.001928	1.001626	1.001656	1.001058	1.000991	1.000627	1.000363	1.000495
26	1.204007	1.030873	1.015178	1.007111	1.004635	1.002955	1.001940	1.001865	1.001611	1.001322	1.001142	1.001141	1.000392	1.000641	1.000783
27	1.176971	1.038287	1.011986	1.009328	1.004853	1.003010	1.002045	1.002205	1.002266	1.001606	1.001309	1.001307	1.000914	1.000815	1.000749
28	1.216009	1.026856	1.014611	1.008280	1.005213	1.003339	1.002478	1.002084	1.002185	1.001442	1.001369	1.001262	1.000630	1.000490	1.000735
29	1.167180	1.038053	1.014406	1.007957	1.004812	1.003505	1.002869	1.002270	1.001831	1.001580	1.001113	1.000803	1.000617	1.000463	1.000555
30	1.247110	1.036656	1.012224	1.007426	1.004759	1.003846	1.003193	1.002476	1.001799	1.001972	1.001054	1.000772	1.000702	1.000421	1.000421
31	1.223999	1.030814	1.014993	1.008560	1.005301	1.004084	1.002944	1.002421	1.001268	1.001025	1.001295	1.000602	1.000601	1.000481	1.000360
32	1.178993	1.032793	1.015826	1.008429	1.005399	1.003106	1.003289	1.002218	1.001668	1.001185	1.001151	1.000703	1.000479	1.000383	1.000542
33	1.203139	1.030922	1.014029	1.010055	1.004396	1.004594	1.002305	1.001509	1.001399	1.000716	1.000716	1.000644	1.000643	1.000536	1.000928
34	1.191362	1.035293	1.014963	1.009119	1.005301	1.003625	1.002418	1.001787	1.001427	1.000920	1.000653	1.000622	1.000533	1.000740	1.000355
35	1.208152	1.035199	1.013092	1.008646	1.004486	1.002753	1.001891	1.001309	1.001277	1.000698	1.001123	1.000637	1.000636	1.000636	1.000539
36	1.201788	1.033339	1.014555	1.008925	1.004106	1.003277	1.002068	1.000927	1.001076	1.000895	1.000835	1.000894	1.000893	1.000590	1.000539
37	1.208225	1.034236	1.014214	1.007860	1.004917	1.002992	1.001678	1.001396	1.000991	1.001238	1.000927	1.001143	1.000713	1.000590	1.000539
38	1.228740	1.032408	1.011662	1.007504	1.004461	1.002149	1.002145	1.002069	1.001175	1.001280	1.001101	1.000954	1.000713	1.000590	1.000539
39	1.201722	1.029254	1.013661	1.007856	1.004721	1.002870	1.002107	1.002448	1.001597	1.001625	1.001124	1.000954	1.000713	1.000590	1.000539
40	1.175583	1.030864	1.011325	1.006538	1.004197	1.002985	1.002150	1.002574	1.001942	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
41	1.215356	1.030815	1.012518	1.006729	1.005021	1.003087	1.002697	1.002499	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
42	1.202151	1.029576	1.012629	1.008242	1.004837	1.003839	1.002762	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
43	1.195135	1.033744	1.015923	1.009314	1.006212	1.004511	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
44	1.222010	1.039634	1.016703	1.010379	1.006633	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
45	1.223407	1.048597	1.018863	1.011888	1.005758	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
46	1.257620	1.053544	1.020214	1.010037	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
47	1.299653	1.050695	1.017250	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
48	1.261411	1.040238	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
49	1.228021	1.043810	1.017022	1.009528	1.005516	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539

Tabell 34: Utvecklingsfaktorer estimerade med den kombinerade random forest modellvarianten, där förändring i backlog (antal) ingår som förklarande variabel i diagonaleffektsmodellen. Utvecklingsfaktorerna presenteras för skademånaderna 25 till 49 och utvecklingsmånaderna 1 till 15. Lila celler representerar utvecklingsfaktorer estimerade med radeffektsmodellen, medan gröna celler representerar utvecklingsfaktorer estimerade med diagonaleffektsmodellen. Blåa celler visar utvecklingsfaktorer som har beräknats med chain ladder metoden. Samtliga utvecklingsfaktorer under den färgmarkerade diagonalen och som inte är färgade har också skattats enligt chain ladder metoden, medan samtliga utvecklingsfaktorer ovanför representerar observerade utvecklingsfaktorer.

Skademånad	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
25	1.388972	1.041382	1.013991	1.009390	1.004433	1.003578	1.002199	1.001928	1.001626	1.001656	1.001058	1.000991	1.000627	1.000363	1.000495
26	1.204007	1.030873	1.015178	1.007111	1.004635	1.002955	1.001940	1.001865	1.001611	1.001322	1.001142	1.001141	1.000392	1.000641	1.000783
27	1.176971	1.038287	1.011986	1.009328	1.004853	1.003010	1.002045	1.002205	1.002266	1.001606	1.001309	1.001307	1.000914	1.000815	1.000749
28	1.216009	1.026856	1.014611	1.008280	1.005213	1.003339	1.002478	1.002084	1.002185	1.001442	1.001369	1.001262	1.000630	1.000490	1.000735
29	1.167180	1.038053	1.014406	1.007957	1.004812	1.003505	1.002869	1.002270	1.001831	1.001580	1.001113	1.000803	1.000617	1.000463	1.000555
30	1.247110	1.036656	1.012224	1.007426	1.004759	1.003846	1.003193	1.002476	1.001799	1.001972	1.001054	1.000772	1.000702	1.000421	1.000421
31	1.223999	1.030814	1.014993	1.008560	1.005301	1.004084	1.002944	1.002421	1.001268	1.001025	1.001295	1.000602	1.000601	1.000481	1.000360
32	1.178993	1.032793	1.015826	1.008429	1.005399	1.003106	1.003289	1.002218	1.001668	1.001185	1.001151	1.000703	1.000479	1.000383	1.000542
33	1.203139	1.030922	1.014029	1.010055	1.004396	1.004594	1.002305	1.001509	1.001399	1.000716	1.000716	1.000644	1.000643	1.000536	1.000928
34	1.191362	1.035293	1.014963	1.009119	1.005301	1.003625	1.002418	1.001787	1.001427	1.000920	1.000653	1.000622	1.000533	1.000740	1.000355
35	1.208152	1.035199	1.013092	1.008646	1.004486	1.002753	1.001891	1.001309	1.001277	1.000698	1.001123	1.000637	1.000636	1.000636	1.000539
36	1.201788	1.033339	1.014555	1.008925	1.004106	1.003277	1.002068	1.000927	1.001076	1.000895	1.000835	1.000894	1.000893	1.000590	1.000539
37	1.208225	1.034236	1.014214	1.007860	1.004917	1.002992	1.001678	1.001396	1.000991	1.001238	1.000927	1.001143	1.000713	1.000590	1.000539
38	1.228740	1.032408	1.011662	1.007504	1.004461	1.002149	1.002145	1.002069	1.001175	1.001280	1.001101	1.000954	1.000713	1.000590	1.000539
39	1.201722	1.029254	1.013661	1.007856	1.004721	1.002870	1.002107	1.002448	1.001597	1.001625	1.001124	1.000954	1.000713	1.000590	1.000539
40	1.175583	1.030864	1.011325	1.006538	1.004197	1.002985	1.002150	1.002574	1.001942	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
41	1.215356	1.030815	1.012518	1.006729	1.005021	1.003087	1.002697	1.002499	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
42	1.202151	1.029576	1.012629	1.008242	1.004837	1.003839	1.002762	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
43	1.195135	1.033744	1.015923	1.009314	1.006212	1.004511	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
44	1.222010	1.039634	1.016703	1.010379	1.006633	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
45	1.223407	1.048597	1.018863	1.011888	1.006899	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
46	1.257620	1.053544	1.020214	1.011805	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
47	1.299653	1.050605	1.017142	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
48	1.261411	1.043401	1.017882	1.009827	1.005611	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539
49	1.243276	1.048563	1.017946	1.011321	1.005507	1.003780	1.002660	1.002082	1.001712	1.001373	1.001124	1.000954	1.000713	1.000590	1.000539