Skadeprosenten – hvor tilfeldig er den?

av cand.scient. **Jon Holtan**, aktuar i UNI Storebrand AS

Jon Holtan

Artikkelen analyserer skadeprosentens tilfeldige variasjon innenfor en vilkårlig skadeforsikringsportefølje. Det legges særlig vekt på metodisk enkelhet slik at analyseresultatene både er enkle å forstå og praktiske å forholde seg til - også for ikke-aktuarer. Følgende sentrale spørsmål blir belyst: *I hvilken grad kan avviket mellom en faktisk observert skadeprosent og en målsatt/budsjettert skadeprosent forklares av statistisk tilfeldighet?* Artikkelen presenterer herunder en konkret sammenheng mellom den tilfeldige variasjonen og seks velkjente beregningsfaktorer. Praktiske anvendelser blir illustrert via konkrete eksempler.

1. Introduksjon

En klassisk problemstilling innenfor et skadeforsikringsselskap knytter seg til evalueringen av en påløpt skadeprosent. Mer presist dreier problemstillingen seg om i hvilken grad avviket mellom en påløpt skadeprosent og en målsatt/budsjettert skadeprosent er tilfeldig, og herunder hvorvidt avviket f.eks. gir grunnlag for premie- og/eller vilkårsendringer. Det synes klart at det er erstatningssidens *tilfeldige variasjon* som genererer hele problemstillingen. I de fleste forsikringsselskap er det derfor skadeaktuarene som har ansvaret for å formidle/kommunisere omfanget av denne variasjonen, og da i særdeleshet overfor selskapenes økonomi- og markedsavdelinger. I det følgende skal vi presentere en enkel matematisk metode for å avklare hvilke generelle faktorer som i hovedsak påvirker den tilfeldige variasjonen til en skadeprosent, og herunder avklare hvor mye hver av disse faktorene betyr i denne sammenheng. Artikkelen legger særlig vekt på metodisk enkelhet slik at analyseresultatene er lett forståelige – og dermed også bevisstgjørende – for de fleste i forsikringsnæringen. På denne bakgrunn bør resultatene på ingen måte tolkes som eksakte sannheter, men kun som resultatmessige indikasjoner. Analytisk presisjon er med andre ord ofret på forenklingens alter. Det understrekes at med begrepet skadeprosent menes *brutto* skadeprosent gjennom hele artikkelen.

2. Risikomodell

Variansen og standardavviket til en skadeprosent er de generelle størrelsene som forteller noe om skadeprosentens tilfeldige variasjon. For å løse vår innledningsvis skisserte problemstilling, er det derfor i første omgang behov for å finne matematiske uttrykk for disse størrelsene. Anta på denne bakgrunn følgende risikomodell:

Vi betrakter en portefølje med k forsikringstakere. Porteføljen observeres fra tidspunkt 0 til tidspunkt t, dvs. i tiden $(0, t)$. Fordi vi i denne sammenheng kun er interessert i å analysere den tilfeldige variasjonen som er knyttet til porteføljens skadeprosent, og ikke i å analysere de risikomessige forskjellene mellom porteføljens ulike forsikringstakere, kan vi anvende en gjennomsnittsbetraktning der alle forsikringstakere i porteføljen antas å være risikohomogene; dvs. at alle forsikringstakere antas å ha like store sannsynligheter for skadeinntreffelse og like store sannsynligheter for ulike skadebeløpsstørrelser gitt skadeinntreffelse. Mer spesifikt antar vi følgende standardmodell:

Alle skadeinntreffelser antas å være innbyrdes uavhengige og Poissonfordelt med tidsmessig konstant inntreffelsessannsynlighet θ.

De enkeltvise erstatningsbeløpene antas å være innbyrdes uavhengige og identisk sannsynlighetsfordelte med forventning μ og varians σ^2, samt uavhengige av sannsynligheten for skadeinntreffelse.

3. Forventning, varians og standardavvik

La $X(t)$ og $P(t)$ betegne henholdsvis den aktuelle porteføljens totalt inntrufne erstatningsbeløp og totalt opptjente premie i løpet av tiden $(0, t)$. La $S(t)$ være porteføljens skadeprosent for perioden $(0, t)$. Herav har vi at: $S(t) = X(t)/P(t)$.

Under modellantagelsene i foregående avsnitt vil $X(t)$ være en såkalt sammensatt Poissonprosess med forventning $\mathrm{E}X(t) = k\theta t\mu$ og varians $\mathrm{Var}\ X(t) = k\theta t\ (\mu^2 + \sigma^2)$.

Anta videre at $P(t) = kpt$, der p = porteføljens gjennomsnittspremie pr. forsikringstaker. Dette er en rimelig antagelse både på bakgrunn av modellantagelsene og fordi premien alltid tjenes opp proporsjonalt med tiden t.

Herav finner vi følgende uttrykk for skadeprosentens *forventning*, *varians* og *standardavvik*:

$$\mathrm{E}\ S(t) = \frac{\mu\theta}{p},$$

$$\mathrm{Var}\ S(t) = \left(\frac{\theta}{p^2 kt}\right)(\mu^2 + \sigma^2),$$

$$\mathrm{Std}\ S(t) = \sqrt{\mathrm{Var}\ S(t)}\ .$$

Med vanlige ord har vi grovt sagt at:

Forventningen til skadeprosenten =

$$\frac{\text{Gjennomsnittserstatning} \bullet \text{Skadefrekvens}}{\text{Gjennomsnittspremie}}$$

Variansen til skadeprosenten =

$$\frac{\text{Skadefrekvens} \bullet (\text{Gjennomsnittserstatning}^2 + \text{Variansen till erstatningsbelopene})}{\text{Gjennomsnittspremie}^2 \bullet \text{Portefoljestorrelse} \bullet \text{Periodelengde}}$$

Standardavviket til skadeprosenten =
Kvadratroten av variansen til skadeprosenten

Vi ser at variansen og standardavviket til skadeprosenten øker med økende skadefrekvens, gjennomsnittserstatning og erstatningsbeløpenes varians. Derimot vil en økning i gjennomsnittspremien, porteføljestørrelsen og observert periodelengde gi motsatt effekt.

4. Konfidensintervall

Som tidligere nevnt er standardavviket til skadeprosenten et mål for størrelsen på skadeprosentens tilfeldige variasjon. Isolert sett er imidlertid informasjonsverdien til denne målstørrelsen av relativt begrenset verdi. Skadeprosentens *konfidensintervall* er derimot et mer informativt mål for den tilfeldige variasjonen.

Et konfidensintervall krever imidlertid antagelser om skadeprosentens sannsynlighetsfordeling. Siden skadeprosenten $S(t)$ er en direkte funksjon av det stokastiske erstatningsbeløpet $X(t)$ og av premiekonstanten $P(t)$, medfører dette igjen at sannsynlighetsfordelingen til $S(t)$ i sin helhet er styrt av sannsynlighetsfordelingen til $X(t)$. En svært enkel antagelse når det gjelder sannsynlighetsfordelingen til $X(t)$ knytter seg til normalfordelingsapproksima- sjonen innenfor det statistiske sentralgrenseteoremet; eller mer presist: Hvis mange nok skader inntreffer for den observerte porteføljen i løpet av tiden $(0, t)$, kan man på bakgrunn av modellantagelsene i avsnitt 2 anta at sannsynlighetsfordelingen til summen av enkelterstatningsbeløpene er tilnærmet normalfordelt *uansett hvilken sannsynlighetsfordeling de enkelte erstatningsbeløpene følger.* Siden porteføljens totale erstatningsbeløp $X(t)$ altså antas som tilnærmet normalfordelt og skadeprosenten $S(t)$ bare er en funksjon av $X(t)$ og av premiekonstanten $P(t)$, medfører dette igjen at også $S(t)$ kan antas som tilnærmet normalfordelt med forventning $\mathrm{E}S(t) = \mu\theta/p$ og varians $\mathrm{Var}\, S(t) = (\theta/p^2kt)\,(\mu^2 + \sigma^2)$.

Presisjonen til normalfordelingsapproksimasjonen synker ikke bare når antall skader synker, men synker også jo mer skjevfordelt (langhalet) enkelterstatningsfordelingen er – noe som er ganske vanlig innenfor skadeforsikring. For å unngå denne modellsvakheten kunne vi ha gjort mer realistiske antagelser om totalerstatningsfordelingen. Den såkalte NP-approksimasjonen (NP = Normal Power) tar f.eks. spesielt hensyn til denne skjevheten, og regnes herunder som en praktisk god approksimasjon til totalerstatningsfordelingen. Videre kunne vi ha antatt konkrete skjevfordelte enkelterstatningsfordelinger for de ulike skadeproduktene, f.eks. at erstatningsbeløpene fulgte ulike lognormal- eller gammafordelinger.

Antagelser av denne kategori er vanlige innenfor aktuariell modellbygging. I forhold til denne artikkelens allmenne målsetning, ville imidlertid slike antagelser virket kompliserende, og vi ville tapt mye av analysens tilsiktede enkelhet. Bare vi er klar over denne grove modellforenklingen, og herunder hvordan den gjør seg utslag i øket resultatusikkerhet, bør vi i denne sammenheng likevel kunne anta at $S(t)$ er tilnærmet normalfordelt.

Anta nå at den underliggende forventede skadeprosenten $\mathrm{E}S(t)$ er ukjent. På bakgrunn av foranstående forutsetninger kan konfidensintervallet til den ukjente forventede skadeprosenten utledes som følger:

Vi har generelt at den standardiserte variabelen

$$\frac{S(t) - \mathrm{E}S(t)}{\mathrm{Std}\, S(t)}$$

er tilnærmet standard normalfordelt med forventning 0 og varians 1. Siden vi herav har at

$$\Pr[S(t) - z_{\alpha/2}\mathrm{Std}\, S(t) < \mathrm{E}S(t) < S(t) + z_{\alpha/2}\mathrm{Std}\, S(t)] = 1 - \alpha,$$

der $z_{\alpha/2} = \alpha/2$-fraktilen i den standardiserte normalfordelingen, vil $100(1-\alpha)$ % konfidensintervallet til den ukjente forventede skadeprosenten være lik

$$S(t) \pm z_{\alpha/2}\,\mathrm{Std}\, S(t).$$

Dette intervallet angir med andre ord et område rundt den observerte skadeprosenten som med en gitt sannsynlighet vil inneholde den ukjente forventede skadeprosenten.

5. Hypotesetesting

Praktisk anvendelse av foranstående konfidensintervall er særlig aktuelt innenfor følgende problemstilling: For en bestemt forsikringsportefølje observeres en bestemt skadeprosent i løpet av en bestemt tidsperiode. Denne skadeprosenten skal vurderes opp mot en på forhånd målsatt eller budsjettert skadeprosent. Følgende sentrale spørsmål melder seg: I hvilken grad kan avviket mellom den faktisk observerte skadeprosenten og den målsatte/budsjetterte skadeprosenten forklares av *statistisk tilfeldighet*? En annen måte å stille spørsmålet på er: I hvilken grad kan avviket forklares av at den målsatte/budsjetterte skadeprosenten er *systematisk* forskjellig fra den ukjente forventede skadeprosenten?

Å ha et fornuftig svar på dette spørsmålet er av stor praktisk betydning fordi man da bl.a. kan unngå unødvendige premie- og/eller vilkårsendringer på bakgrunn av den observerte skadeprosenten. Svaret er åpenbart knyttet til *statistisk hypotesetesting*. To typer feil knytter seg til spørsmålet om å endre et premienivå på basis av en observert skadeprosent:

Feil I: Justering av premienivået når premienivået er *riktig* i forhold til målsatt skadeprosent.

Feil II: Opprettholde premienivået når premienivået er *feil* i forhold til målsatt skadeprosent.

Hele poenget med statistisk hypotesetesting er i denne sammenheng å kontrollere sannsynligheten for å gjøre henholdsvis feil I og feil II. Siden de to feilene står i et motsetningsforhold til hverandre (dvs. en reduksjon av sannsynligheten for den ene feilen, gir en økning av sannsynligheten for den andre feilen, og omvendt), synes det herunder klart at man blir nødt til å velge hvilken av de to feilene man mest ønsker å unngå, for deretter å sørge for at sannsynligheten for denne feilen er liten. Det synes herunder å være mest fornuftig å primært unngå feil I. Dette pga. konservativstankegangen samt behovet for å unngå unødige premieendringer i relasjon til en optimal kundebevaring. Herav kan vi f.eks. bestemme oss for at sannsynligheten for å gjøre feil I, dvs. å justere premienivået når premienivået likevel er riktig i forhold til den målsatte skadeprosenten, ikke skal overstige 10 %. Innenfor hypotesetesting sier man iså-fall at vurderingen av den aktuelle premieendringen har et signifikansnivå på 10 %.

Hvis vi lar $m(t)$ være den forutbestemt målsatte/budsjetterte skadeprosenten og $S(t)$ den observerte skadeprosenten, vil valgene av null-hypotese () og alternativ hypotese () i dette tilfellet være:

$H_0 : ES(t) = m(t) \Rightarrow$ Den ukjente forventede skadeprosenten er ikke systematisk forskjelling fra målsatt/budsjettert skadeprosent.

$H_1 : ES(t) \neq m(t) \Rightarrow$ Den ukjente forventede skadeprosenten er systematisk forskjellig fra målsatt/budsjettert skadeprosent.

H_0 sitt to-sidige 10 %-forkastningsområde er således:

$$\text{R: } m(t) < S(t) - z_{0.05}\,\text{Std}\,S(t) \text{ eller}$$
$$m(t) > S(t) + z_{0.05}\,\text{Std}\,S(t)\,.$$

Med andre ord: blir ikke forkastet såfremt $m(t)$ ligger innenfor 90 %-konfidensintervallet rundt den observerte skadeprosenten. Eller mer praktisk forståelig i forhold til vår skisserte problemstilling: Det finnes ikke noe statistisk signifikant grunnlag for å foreta særskilte premie- og/eller vilkårsendringer basert på den observerte skadeprosenten såfremt den skadeprosenten man styrer mot ligger *innenfor* konfidensintervallet til den observerte skadeprosenten. Hvis den ønskede skadeprosenten derimot ligger *utenfor* dette konfidensintervallet, foreligger det et statistisk signifikant grunnlag for å foreta premie- og/eller vilkårsendringer.

Oppsummert kan vi si at siden vår problemstilling (dvs. i hvilken grad avviket mellom observert og målsatt/budsjettert skadeprosent er tilfeldig) vanskelig kan ha et ja/nei-svar, må svaret istedet basere seg på sannsynlighetsutsagn. Et konfidensintervall representerer nettopp et slikt utsagn.

6. Praktiske eksempler

De matematiske uttrykkene for skadeprosentens standardavvik (avsnitt 3) og skadeprosentens konfidensintervall (avsnitt 4) gjelder generelt for alle typer skadeforsikringsporteføljer. Praktisk anvendelse av uttrykkene krever imidlertid verdifastsettelse av de ulike parameterne. Fire av parameterne kan kalles *karakteristiske* idet de karakteriserer den enkelte portefølje. Dette gjelder skadefrekvens, gjennomsnittserstatning, gjennomsnittspremie og erstatningsbeløpenes varians. Disse størrelsene er velkjente, og det er lett å fastsette deres parameterverdier basert på statistiske erfaringer.

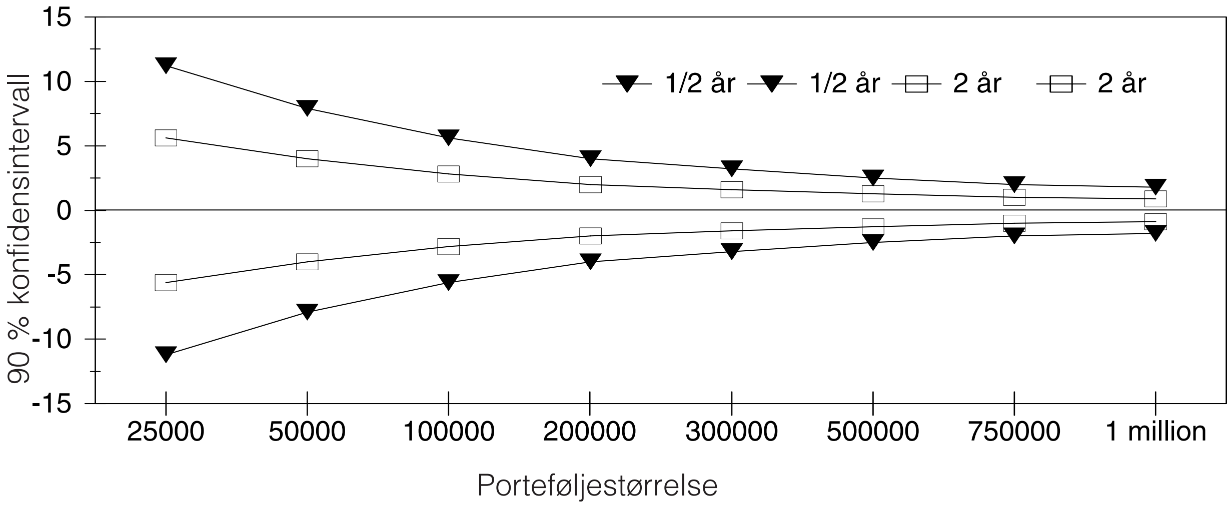
La oss på denne bakgrunn studere f.eks. personbilforsikring. Etter å ha fastsatt konkrete verdier for de fire nevnte parameterne, viser figur 1 skadeprosentens 90 %-konfidensintervall som funksjon av de to øvrige parameterne observasjonstid og porteføljestørrelse. Merk at konfidensintervallet baserer seg på parameterverdier som gjelder for det norske forsikringsmarkedet.

Figur 1 viser at bredden på konfidensintervallet er synkende med økende porteføljestørrelse og med økende observasjonstid, og omvendt. Dette gjelder generelt for alle skadeforsikringsporteføljer, jfr. uttrykket for standardavviket i avsnitt 3. Eksempelvis ser vi at med en observasjonstid på 1/2 år og med en porteføljestørrelse på 200 000 forsikringer, er 90 %-konfidensintervallet rundt den observerte skadeprosenten på omtrent +/- 4 %-poeng. Det betyr at hvis den målsatte/budsjetterte skadeprosenten ligger utenfor +/- 4 %-poeng fra den observerte skadeprosenten, kan vi med 90 % sannsynlighet si at avviket *ikke* skyldes ren tilfeldighet.

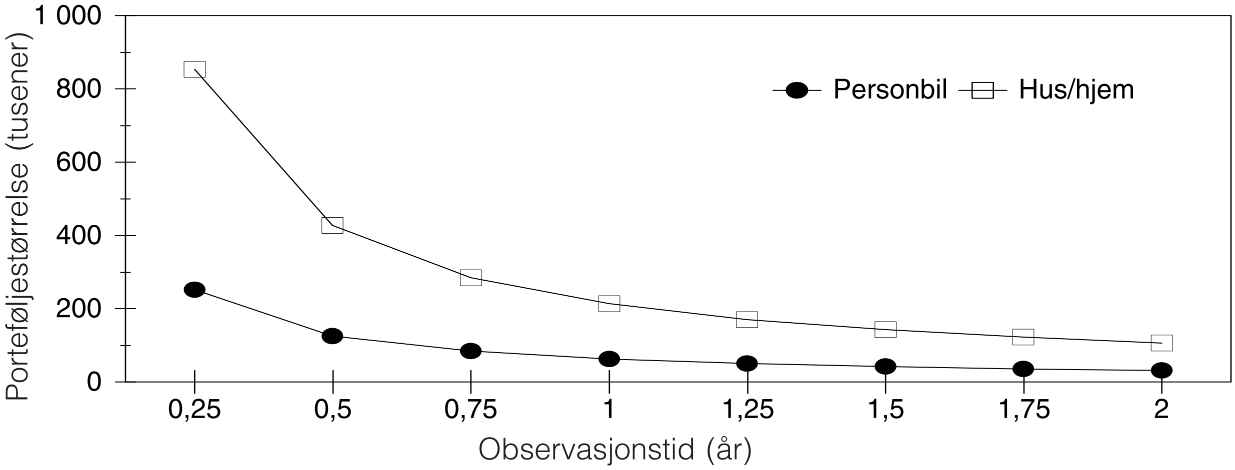
Tilsvarende kan vi studere spesifiserte delporteføljer (som er karakterisert av andre karakteristiske parametere) innenfor personbilforsikring, og dessuten andre typer forsikringsprodukter. Det er bare fantasien som setter grenser for hvordan vi kan presentere ulike analyser av skadeprosentens konfidensintervall. Tredimensjonal grafikk – som bl.a. er tilgjengelig innenfor programvaresystemet *Mathematica* – egner seg ypperlig til slike formål.

Vi skal i denne sammenheng begrense oss til å studere hvordan personbilforsikring opptrer i forhold til hus/hjem-forsikring: Anta at vi forutsetter at skadeprosentens konfidensintervall *skal* være lik +/- 5 %-poeng. Hvordan vil denne forutsetningen gi seg utslag i krav om observasjonstid og porteføljestørrelser? Figur 2 viser hvordan!

Figur 1: Skadeprosentens konfidensintervall for personbilforsikring som funksjon av observasjonstid og porteføljestørrelse



Figur 2: Porteføljestørrelse vs. observasjonstid under forutsetning om et konfidensinterval på +/- 10 %



Figur 2 viser – ikke overraskende – at kravet til observasjonstid er omvendt proporsjonalt med kravet til porteføljestørrelse: Jo lengre observasjonstid, jo lavere krav til porteføljestørrelse for å oppnå et konfidensintervall på +/- 5 %-poeng, og omvendt. Dernest ser vi at hus/hjem-forsikring krever både større forsikringsportefølje og tildels lengre observasjonstid for å oppnå samme tilfeldige variasjon i skadeprosent som personbilforsikring. Eller sagt på en annen måte: Personbilforsikring representerer et mindre risikofylt forretningsområde enn hus/hjem-forsikring gitt like store porteføljestørrelser. Merk at dette resultatet gjelder særskilt for det norske forsikringsmarkedet. Andre sammensetninger av karakteristiske parameterverdier kan åpenbart gi andre resultater for det svenske og danske markedet.

Nettopp i sammensetningen av parameterverdiene finner vi årsaken til risikoforskjellen mellom personbil og hus/hjem som eksisterer innenfor det norske markedet. Ved å studere hvordan parameterverdiene for hver av produktene påvirker standardavviket i avsnitt 3, finner vi at den lave risikoen til personbilforsikring i hovedsak er forårsaket av en kombinasjon av høyere gjennomsnittspremie og lavere variasjon i enkelterstatningsbeløpene i forhold til hus/hjem-forsikring.

Generelt kan det antydes – kanskje noe overraskende – at skadeforsikringsprodukter med høy skadefrekvens er *mindre* risikofylte enn de med lav skadefrekvens. Dette fordi høy skadefrekvens ofte er kombinert med relativt høy gjennomsnittspremie og begrenset variasjon i erstatningsbeløpene, noe som gir et forholdsvis begrenset standardavvik for skadeprosenten. Produkter med lav skadefrekvens preges derimot ofte av motsatte verdier for gjennomsnittspremie og beløpsvariasjon, noe som genererer høyere standardavvik.

7. Avsluttende kommentarer

I et skadeforsikringsselskap er det mange mennesker som har til oppgave å vurdere eller mene noe om påløpne skadeprosenter, enten det gjelder totalt sett for et produkt eller det gjelder en delportefølje av et produkt. I denne sammenheng er det helt nødvendig å ha en kvalifisert oppfatning av skadeprosentens tilfeldige variasjon, og herunder hvordan denne variasjonen varierer mellom de ulike forsikringsproduktene. Eller sagt på en annen måte: Å ha kontroll over den tilfeldige porteføljevariasjonen er minst like viktig innenfor en skadeforsikringportefølje som innenfor en kapitalforvaltningsportefølje. I denne artikkelen fremsettes en enkel sammenheng som antyder hvordan seks velkjente beregningsfaktorer påvirker den tilfeldige variasjonen til skadeprosenten. Generelt har vi at:

- Alt annet likt vil den tilfeldige variasjonen *øke* med økende skadefrekvens, gjennomsnittserstatning og variansen til de enkelte erstatningsbeløpene.
- Alt annet likt vil den tilfeldige variasjonen *synke* med økende gjennomsnittspremie, porteføljestørrelse og observert periodelengde.

En velkjent og åpenbar konsekvens av denne sammenhengen er at jo større markedsandel et forsikringsselskap har innenfor et forretningssegment, desto lavere tilfeldig variasjon er knyttet til selskapets virksomhet innenfor dette segmentet. Store selskaper bruker derfor en mindre andel av forsikringspremien til å dekke den tilfeldige variasjonen (risikofluktuasjonen) i forhold til sine mindre konkurrenter. Gitt samme krav til fluktuasjonssikkerhet, har derfor store selskaper en profittfordel i forhold til mindre selskaper ved at de i prinsippet kan:

- systematisk sørge for relativt større overskudd ved å forsikre riskene til *samme priser* som sine mindre konkurrenter, eller
- systematisk øke sine markedsandeler ytterligere ved å forsikre riskene til *lavere priser* enn sine mindre konkurrenter.

Denne automatiske profittfordelen er en av flere årsaker til at det er så vanskelig for mindre selskaper å øke sine markedsandeler på bekostning av større selskaper uten betydelig fri egenkapital i ryggen.