

Prof. Dr. Dietmar Pfeifer
Universität Hamburg
Fachbereich Mathematik
Institut für Mathematische Stochastik

Diplomarbeit

Stochastische Modelle für Schadenabwicklungsschemata unter Berücksichtigung von Reservenbildung

eingereicht von:

Irina Kaiser

10. Sem. Dipl. Mathematik

Matr. Nr.: 4737817

Ruststr. 5

21073 Hamburg

Tel.: 040 / 702 56 69

Hamburg, den 5. Juli 2000

Inhaltsverzeichnis

1	Einführung.....	3
1.1	Ursachen lang andauernder Schadenabwicklung.....	3
1.2	Das Abwicklungsdreieck und das Schadenreservierungsproblem.....	3
1.3	Datenarten.....	5
2	Verfahren.....	6
2.1	Chain-Ladder.....	7
2.1.1	Das Modell.....	7
2.1.2	Numerische Berechnung.....	8
2.2	Robuste Faktoren.....	10
2.2.1	Chain-Ladder und der KQ-Schätzer.....	10
2.2.2	Begriff der Robustheit.....	11
2.2.3	M-Schätzung.....	11
2.2.4	Lösung der Gleichungen (5) und (6).....	13
2.2.5	Numerische Berechnung.....	14
2.3	Stochastische Faktoren.....	17
2.3.1	Das Modell.....	17
2.3.2	Endschaden im Log-Normal-Modell.....	20
2.3.3	Endschaden im Log-Gamma-Modell.....	24
2.3.4	Endschaden im Log-inversen gaußschen Modell.....	30
2.3.5	Numerische Berechnung.....	34
2.3.6	Kolmogorov-Test.....	38
2.4	Die Kleinste-Quadrate-Regression.....	41
2.4.1	Modellierung für die KQ – Methode und Modellannahmen.....	41
2.4.2	Schätzen im linearen Regressionsmodell.....	42
2.4.3	Schätzeigenschaften der KQ-Methode.....	43
2.4.4	Angepasste Werte, Residuen und Standardabweichung.....	44
2.4.5	Das Bestimmtheitsmaß.....	44
2.4.6	Signifikanztest für die Regressionskoeffizienten.....	45
2.4.7	Signifikanz des Gesamtzusammenhangs.....	45
2.4.8	Numerische Berechnung.....	46
3	Zusammenfassung.....	68
3.1	Vergleich der Verfahren.....	68
3.2	Prognosefehlermaße.....	69
3.3	Numerische Ergebnisse.....	70
4	Literaturverzeichnis.....	73
	Anhang A: Hilfssätze.....	75

1 Einführung

1.1 Ursachen lang andauernder Schadenabwicklung

Ein versicherter Schaden durchläuft bis seine Regulierung mehrere Stadien: die Verursachung des Schadens, den Eintritt des Schadenereignisses, die Entdeckung des Schadens und seine Meldung beim Versicherer. Während in der Sachversicherung die Zeitspanne zwischen dem Schadeneintritt und der Schadenregulierung in der Regel noch in der Versicherungsperiode stattfindet, ist dies in der Haftpflichtversicherung¹ oft nicht der Fall. Es können Jahre oder sogar Jahrzehnte vergehen, bis die endgültige Schadenhöhe feststeht. Der Versicherer muss also damit rechnen, dass er noch lange Zeit nach Ablauf des Versicherungsvertrages in Anspruch genommen werden kann.

Solche Schäden, deren Abwicklung so lange dauert, werden Spät- oder IBNR-Schäden² („incurred but not reported“) genannt. Dies sind die Schäden, die bereits eingetreten, dem Versicherer jedoch noch nicht gemeldet sind.

Als Sonderfall der IBNR-Schäden sind IBNER-Schäden („incurred but not enough reported“)³. Der Versicherer kennt zwar den Eintritt des Schadenereignisses, kennt aber nicht das vollständige Ausmaß des Schadens.

Für den Versicherer ist es wesentlich, die Höhe der noch ausstehenden Zahlungen zu schätzen. Dieser Betrag, der unter anderem wichtig für die Prämienkalkulation ist, wird als Spätschadenreserve oder kurz Reserve bezeichnet.

1.2 Das Abwicklungsdreieck und das Schadenreservierungsproblem

Das Ziel ist also die Prognose der zukünftigen Auszahlungen für bereits entstandene Schäden. Dem Versicherungsunternehmen liegen die Daten vor, die für diese Prognose benutzt werden können. Es sind nämlich zu jedem Schaden der Zeitpunkt der Entstehung sowie die Zeitpunkte und die Höhen der bereits geleisteten Zahlungen bekannt. In dieser Arbeit werden nur die Verfahren untersucht, die nicht die Zahlungsabwicklung eines einzelnen Schadens

¹ Die Haftpflichtversicherung gehört wegen der langen Regulierungsdauer der Großschäden zum sogenannten „long tail business“.

² Die lange Regulierungsdauer der großen Personenschäden in der Kfz-Haftpflichtversicherung ergibt sich daraus, dass oft lebenslang Heil- und Pflegekosten erstattet und Rentenzahlungen geleistet werden müssen.

³ In der Praxis wird meistens zwischen den beiden Schadenarten nicht unterschieden; sie werden dann zusammen betrachtet.

untersuchen, sondern die Zahlungsabwicklung für die Summe einer Gruppe von Schäden, die in dem gleichen Jahr eingetreten sind.

Die Zahlungen unterscheiden sich nach zwei zeitlichen Merkmalen:

- Anfalljahr: das Jahr, in dem der Schaden eingetreten ist, und
- Abwicklungsjahr: gibt den Abstand zwischen dem Anfalljahr und dem Jahr, in dem die Auszahlung stattgefunden hat, an.

Durch das Anfalljahr (Accident Year) und das Abwicklungsjahr (Development Year) wird das Kalenderjahr (Payment Year) bestimmt. Das ist das Jahr, in dem die Auszahlung geleistet wurde.

Die Daten liegen meist in folgender Dreiecksform vor:

Das Abwicklungsdreieck

AnfJahr	AbwJahr					
	1	2	...	$n+1-i$...	n
1	$P_{1,1}$	$P_{1,2}$...	$P_{1,n+1-i}$...	$P_{1,n}$
2	$P_{2,1}$	$P_{2,2}$...	$P_{2,n+1-i}$...	
⋮	⋮	⋮	...	⋮		
i	$P_{i,1}$	$P_{i,2}$...	$P_{i,n+1-i}$		
⋮	⋮	⋮				
$n-1$	$P_{n-1,1}$	$P_{n-1,2}$				
n	$P_{n,1}$					

$P_{i,j}$ ist also die Auszahlung im Abwicklungsjahr j für den Schaden, der im Anfalljahr i entstanden ist. Das erste Abwicklungsjahr entspricht dem Anfalljahr. n ist die Anzahl der Anfall- und Abwicklungsjahre. Es wird angenommen, dass nach n Jahren der Schaden vollständig reguliert ist. Damit ist die vollständige Abwicklung für das erste Anfalljahr bekannt, und die Beträge $P_{i,n+1}, P_{i,n+2}, \dots, i = 1, \dots, n$ sind alle gleich null.

Der Betrag

$$U_i = P_{i,1} + P_{i,2} + \dots + P_{i,n}$$

ist der Gesamtschaden für das Anfalljahr i , der für die Zwecke der Prämienkalkulation und für die interne Rechnung benötigt wird.

Von U_i ist aber für $i = 2, \dots, n$ nur der Betrag

$$P_{i,1} + P_{i,2} + \dots + P_{i,n+1-i}$$

bekannt. Ziel der mathematischen Verfahren ist, den noch unbekanntem Teil

$$R_i = P_{i,n+2-i} + \dots + P_{i,n}$$

zu schätzen. R_i wird als Spätschadenreserve oder kurz Reserve für das Anfalljahr i bezeichnet.

Für einige Verfahren werden die Daten in kumulierter Form verwendet, d. h. an der Stelle (i, k) des Dreiecks steht nicht der Zuwachs $P_{i,k}$, sondern der kumulierte Schadenstand

$$C_{i,k} = P_{i,1} + P_{i,2} + \dots + P_{i,k} .$$

Aus einem kumulierten Abwicklungsdreieck werden die Zuwächse mittels $P_{i,k} = C_{i,k} - C_{i,k-1}$ ermittelt, wobei $C_{i,0} = 0$.

1.3 Datenarten

Ein Abwicklungsdreieck kann aus zwei verschiedenen Arten von Daten – bezahlt oder angefallen – gebildet werden. Beide Typen von Dreiecken der bezahlten und der angefallenen Schäden werden in dieser Arbeit untersucht.

$C_{i,k}$ stellt entweder alle Zahlungen bis einschließlich des Abwicklungsjahres k dar, oder $C_{i,k}$ beinhaltet zusätzlich die zu diesem Zeitpunkt bestehenden Einzelfallreserven. Im ersten Fall hat man bezahlte Zahlungen und im zweiten angefallene Zahlungen.

Die Einzelfallreserven sind schon Schätzwerte, die z. B. vom Sachbearbeiter aufgrund der Erfahrungen für den bestimmten Fall festgesetzt werden können. Meist fallen die Reserven, berechnete auf der Basis bezahlter Schäden, höher aus. Die Frage ist, für welche der beiden Reserven, der Versicherer sich entscheiden soll. Da die bezahlte Schäden keine Schätzungen beinhalten, erscheinen sie auf den ersten Blick zuverlässiger. Sind aber die Schätzungen für die Einzelfallreserven von erfahrenen Sachbearbeiter festgesetzt, so lässt sich die Höhe des Endschadens pro Anfalljahr viel früher erkennen, was wiederum für angefallene Schäden spricht. Es ist also die Erfahrung des Versicherers gefragt, welche von den beiden Reserven er letztendlich verwendet.¹

¹ L. Halliwell hat in seinem Artikel „Conjoint Prediction of Paid and Incurred Losses“ in PCAS, 1998(?) ein Verfahren vorgeschlagen, das die Reserve simultan mit beiden Datenarten schätzt. Dieses Verfahren wird in dieser Arbeit nicht behandelt. In der Praxis werden die Reserven immer für beide Datenarten geschätzt und anschließend wird die Entscheidung getroffen.

2 Verfahren

Die Schätzung der Reserve ist in den Branchen mit langer Abwicklungsdauer von entscheidender Bedeutung für die Rechnungslegung und die Prämienkalkulation. Eine Reihe von mathematischen und weniger mathematischen Schätzverfahren wurde dafür entwickelt. Alle diese Verfahren versuchen, die Erfahrungen früherer Anfalljahre auf spätere Anfalljahre zu übertragen. Daher muss für alle diese Verfahren, damit sie richtig funktionieren, vorausgesetzt werden, dass die betrachteten Jahre keine Trend- und Strukturbrüche enthalten. In dieser Arbeit werden sechs unterschiedliche Verfahren zur Schätzung der Reserve bzw.¹ des Endschadens auf vorgegebene Dreiecke der bezahlten bzw. angefallenen Schäden angewandt.

Als einführendes Verfahren wird das Chain-Ladder-Verfahren betrachtet. Das ist das einfachste und älteste Verfahren zur Schätzung der Schadenreserve und ist das Verfahren, das in der Praxis immer noch am häufigsten angewendet wird. Das Verfahren der robusten Faktoren wird als eine Modifizierung des Chain-Ladder-Verfahrens betrachtet und ist besonders gut für Daten mit Ausreißern geeignet.

Drei weitere Verfahren legen unterschiedliche Verteilungsannahmen der Abwicklungsfaktoren zugrunde. Es werden die Log-Normal-, Log-Gamma- und Log-inverse Gaußverteilung behandelt. Aufgrund der Modellannahmen kann der Endschaden geschätzt werden. Die Verteilungsannahmen ermöglichen auch die Durchführung von Simulationen, mit deren Hilfe man den Endschaden und die Standardabweichung des Endschadens ermitteln kann.

Als letztes Verfahren wird die Kleinste-Quadrate-Regression, die z.B. aus dem Bereich der Ökonometrie bekannt ist, dargestellt.

Zum Schluss werden alle Verfahren mittels Prognosefehlermaßen miteinander verglichen.

¹ Es reicht die Schätzung einer der beiden Größen, des Endschadens oder der Reserve, da aufgrund der einen die andere immer berechnet werden kann.

2.1 Chain-Ladder

2.1.1 Das Modell

Es werden folgende Annahmen getroffen:

- Jedes Anfalljahr besitzt sein eigenes Anfalljahr-Niveau.
- Die Aufteilung des Endschadens auf die Abwicklungsjahre ist im Schnitt für alle Anfalljahre gleich.

Der Endschaden für das Anfalljahr i beträgt

$$C_{i,n} = P_{i,1} + \dots + P_{i,n}.$$

Der additiv zerlegte Endschaden kann nach dem Modell multiplikativ dargestellt werden in der Form

$$C_{i,n} = C_{i,1} \cdot F_{i,2} \cdot F_{i,3} \cdot \dots \cdot F_{i,n},$$

wobei

$$F_{i,k} = \frac{C_{i,k}}{C_{i,k-1}}$$

die multiplikative Zunahme des Schadens von Abwicklungsjahr $k-1$ zu Abwicklungsjahr k ist.

Nach der zweiten Annahme ist der Erwartungswert der Zufallsvariablen $F_{i,k}$ unabhängig vom Anfalljahr i :

$$E(F_{i,k}) = f_k, \quad i = 1, \dots, n, \quad k = 2, \dots, n.$$

Der Schätzer für die f_k werden mittels des $C_{i,k-1}$ -gewichteten arithmetischen Mittels ermittelt durch:

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k+1} C_{i,k-1} F_{i,k}}{\sum_{i=1}^{n-k+1} C_{i,k-1}} = \frac{\sum_{i=1}^{n-k+1} C_{i,k}}{\sum_{i=1}^{n-k+1} C_{i,k-1}}$$

$$2 \leq k \leq n.$$

Die \hat{f}_k geben damit die durchschnittliche Steigerung des Schadenstands von Abwicklungsjahr $k-1$ auf Abwicklungsjahr k und werden als Chain-Ladder-Faktoren oder Abwicklungsfaktoren bezeichnet.

Mit den ermittelten Faktoren lässt sich der Endschaden für die Anfalljahre $i = 2, \dots, n$ leicht berechnen:

$$\hat{C}_{i,n} = C_{i,n+1-i} \cdot \hat{f}_{n+2-i} \cdot \dots \cdot \hat{f}_n, \quad 2 \leq i \leq n.$$

Daraus ergibt sich für die Reserve

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n+1-i} = C_{i,n+1-i} \cdot (\hat{f}_{n+2-i} \cdot \dots \cdot \hat{f}_n - 1), \quad 2 \leq i \leq n.$$

Das heißt, für die Prognose von R_i wird nur der aktuelle Schaden $C_{i,n+1-i}$ des Anfalljahres i gebraucht, während die früheren Stände $C_{i,1}, \dots, C_{i,n-i}$ nicht berücksichtigt werden.

2.1.2 Numerische Berechnung

Mit dem Dreieck der angefallenen kumulierten Schadenstände¹:

116.324.464	145.770.755	166.839.732	186.371.739	195.585.093	199.626.757	204.064.977
88.735.292	123.263.458	142.524.171	150.114.806	160.020.806	167.280.477	
121.026.936	145.665.387	154.914.921	169.865.543	182.263.482		
146.309.408	182.574.022	208.900.961	235.167.579			
148.605.503	185.581.590	203.552.546				
214.033.228	262.103.554					
297.779.107						

berechnet man die Faktoren:

$$\hat{f}_2 = 1,251395$$

$$\hat{f}_3 = 1,119916$$

$$\hat{f}_4 = 1,101518$$

$$\hat{f}_5 = 1,062244$$

$$\hat{f}_6 = 1,031781$$

$$\hat{f}_7 = 1,022233$$

Analog mit dem Dreieck der bezahlten kumulierten Schadenstände:

52.686.881	80.060.344	100.996.975	125.423.264	147.503.538	168.977.139	179.534.285
43.411.656	69.003.908	83.824.923	98.865.737	117.262.052	130.972.680	
67.445.366	103.616.206	113.809.052	128.143.987	145.327.589		
89.936.759	129.135.967	150.368.251	176.446.759			
81.854.058	132.066.969	144.120.579				
132.685.938	192.820.072					
163.718.217						

berechnet man die Faktoren:

$$\hat{f}_2 = 1,509983$$

$$\hat{f}_3 = 1,154191$$

$$\hat{f}_4 = 1,177908$$

$$\hat{f}_5 = 1,163606$$

$$\hat{f}_6 = 1,132888$$

$$\hat{f}_7 = 1,062477$$

¹ Für alle Verfahren werden zur numerischen Berechnung immer die hier definierten 7x7 Dreiecke der bezahlten und der angefallenen Schäden verwendet.

Daraus ergibt sich für die Reserven für die Anfalljahre $i = 2, \dots, 7$ entsprechend für die bezahlten und angefallenen Schäden:

Anfalljahr	Reserve_ang	Reserve_bez
2	3.719.078	8.182.750
3	9.973.396	29.598.495
4	28.306.992	70.683.649
5	47.653.105	93.645.484
6	100.148.912	174.338.720
7	217.244.607	307.010.720
Gesamt:	407.046.090	683.459.819

Die Reserve, berechnet auf Basis der bezahlten Schäden, ist beträchtlich höher als die Reserve für die angefallenen Schäden.

2.2 Robuste Faktoren

2.2.1 Chain-Ladder und der KQ-Schätzer

Es lässt sich leicht erkennen, dass der CL-Schätzer nichts anderes als der KQ-Schätzer der gewöhnlichen linearen Regressionsgleichung der Form:

$$C_{i,j} = f_j C_{i,j-1} + u_{i,j}, \quad 2 \leq j \leq n, 1 \leq i \leq n, \quad (1)$$

ist, wobei $u_{i,j}$ ein stochastischer Störterm mit folgenden Eigenschaften ist:

$$E[u_{i,j}] = 0$$

$$\text{Var}[u_{i,j}] = s_j^2 C_{i,j-1}, \quad \text{für alle } i = 1, \dots, n.$$

Also ist die Varianz des Störterms nicht konstant über alle Beobachtungen $C_{i,j-1}$, $i = 1, \dots, n$.

Es handelt sich also um eine gewichtete lineare Regression. Nach der Methode der kleinsten Quadrate wird daher der Regressionskoeffizient f_j durch Minimieren von:

$$Q(f_j) = \sum_{i=1}^{n+1-j} (C_{i,j} - f_j C_{i,j-1})^2 / C_{i,j-1} \quad (2)$$

bestimmt.

Tatsächlich führt die Lösung dieser Minimierungsaufgabe ebenfalls zum Chain-Ladder-Schätzer:

$$\hat{f}_j = \frac{\sum_{i=1}^{n+1-j} C_{i,j}}{\sum_{i=1}^{n+1-j} C_{i,j-1}} \quad (3)$$

Beweis:

$$\begin{aligned} \frac{d(Q(f_j))}{d(f_j)} &= 2 \sum_{i=1}^{n+1-j} (C_{i,j} - f_j C_{i,j-1}) \left(-\frac{C_{i,j-1}}{C_{i,j-1}^2} \right) = 0 \\ &\Rightarrow \sum_{i=1}^{n+1-j} C_{i,j} = f_j \sum_{i=1}^{n+1-j} C_{i,j-1} \\ &\Rightarrow \hat{f}_j = \frac{\sum_{i=1}^{n+1-j} C_{i,j}}{\sum_{i=1}^{n+1-j} C_{i,j-1}} \end{aligned}$$

Q.E.D.

2.2.2 Begriff der Robustheit

Dieser Schätzer reagiert sehr empfindlich auf Ausreißer in den Daten. Eine Schätzung wird dadurch in die Richtung eines oder mehrerer Ausreißer „verzerrt“, so dass sie nicht mehr die Struktur der Masse der Beobachtungswerte wiedergibt. Wenn man aus der Stichprobe die Ausreißer entfernen würde, könnte sich ein völlig unterschiedliches Bild ergeben. Ein robustes Verfahren zielt dagegen darauf ab, atypische Beobachtungswerte bei einer Parameterschätzung in dem Sinne unter Kontrolle zu halten, dass die Ergebnisse durch sie nicht beliebig verzerrt werden können. Robuste Schätzer besitzen in solcher Situationen im allgemeinen eine höhere Effizienz, was eine größere Zuverlässigkeit der Schätzung bedeutet. Die Aufgabe dieses Kapitels ist die Robustifizierung der Chain-Ladder-Schätzung.

2.2.3 M-Schätzung

Um einen robusten M-Schätzer¹ für das gegebene Regressionsmodell (1) zu entwickeln, wird die obige Minimierungsaufgabe (2) verändert, indem man anstelle der quadratischen Abweichungen alternative Funktionen \mathbf{r} der standardisierten Residuen betrachtet, die weniger empfindlich gegenüber Ausreißern sind. Die modifizierte KQ-Funktion nimmt damit die Form

$$Q^*(f_j, \mathbf{s}_j) = \sum_{i=1}^{n+1-j} \mathbf{r} \left(\frac{C_{i,j} - f_j C_{i,j-1}}{\mathbf{s}_j} \right) \mathbf{s}_j \quad (4)$$

an, aus der zwei robuste Schätzer für f_j und \mathbf{s}_j gewonnen werden können. Der Faktor \mathbf{s}_j wurde aus technischen Gründen hinzugefügt.² Sofern die Funktion \mathbf{r} differenzierbar ist, erhält man mit $\mathbf{r}'(x) = \mathbf{y}(x)$ und $\mathbf{c}(x) = x \cdot \mathbf{y}(x) - \mathbf{r}(x)$ durch Differenzieren von (4) nach f_j und \mathbf{s}_j die zwei Gleichungen:

$$\sum_{i=1}^{n+1-j} \mathbf{y} \left(\frac{C_{i,j} - f_j C_{i,j-1}}{\mathbf{s}_j} \right) C_{i,j-1} = 0 \quad (5)$$

$$\sum_{i=1}^{n+1-j} \mathbf{c} \left(\frac{C_{i,j} - f_j C_{i,j-1}}{\mathbf{s}_j} \right) = 0 \quad (6)$$

Die simultane Lösung beider Gleichungen liefert robuste Schätzer für f_j und $\mathbf{s}_j, j = 2, \dots, n$.

¹ Der Begriff der M-Schätzung ist eine Abkürzung für verallgemeinerte Maximum-Likelihood-Schätzung, vgl. H. Eckey, R. Kosfeld, C. Dreger, Ökonometrie (1995), Gabler Verlag, S. 224 f

² Siehe P. Huber, Robust Statistics (1981), Wiley series in probability and mathematical statistics, S. 176.

Für $\mathbf{r}(x) = \frac{x^2}{2C_{i,j-1}}$ ist $\mathbf{y}(x) = \frac{x}{C_{i,j-1}}$, so dass durch (5) der KQ-Schätzer \hat{f}_j definiert ist, der

mit dem Maximum-Likelihood-Schätzer übereinstimmt, wenn die Störgrößen einer Normalverteilung folgen. Eine solche Wahl der \mathbf{y} -Funktion erfolgt jedoch im Rahmen einer robusten Schätzung für die „guten“ Daten. Für einen robusten M-Schätzer ist daher eine andere, geeignetere Form der \mathbf{y} -Funktion zu wählen.

Um den Einfluss von Ausreißern auf die Schätzung zu begrenzen, hat Kremer die Verwendung folgender Funktion für \mathbf{r} vorgeschlagen:¹

$$\begin{aligned}\mathbf{r}(x) &= c|x|, & \text{für } |x| \geq c \\ &= x^2/2 + c^2/2, & \text{für } |x| \leq c\end{aligned}\quad (7)$$

Daraus ergibt sich für \mathbf{y} :

$$\begin{aligned}\mathbf{y}(x) &= -c, & \text{für } x \leq -c \\ &= x, & \text{für } |x| \leq c \\ &= c, & \text{für } x \geq c\end{aligned}\quad (8)$$

und für \mathbf{c} :

$$\mathbf{c}(x) = \frac{1}{2}[\mathbf{y}(x)^2 - c^2] \quad (9)$$

wobei c das 95% -Quantil einer Standardnormalverteilung ist, also:

$$c = 1,645. \quad (10)$$

Damit liefern die Gleichungen (5) bis (10) zwei robuste Schätzer \hat{f}_j und $\hat{\mathbf{s}}_j$ für f_j bzw. \mathbf{s}_j .

\hat{f}_j , $j=2,\dots,n$ sind damit die robusten Abwicklungsfaktoren, die auf das Dreieck der kumulierten Schäden angewendet werden können.

¹ E. Kremer, „Robust Lagfactors“ (1997), Blätter der deutschen Gesellschaft für Versicherungsmathematik.

2.2.4 Lösung der Gleichungen (5) und (6)

Die Gleichungen (5) und (6) können nicht analytisch gelöst werden. Es bieten sich numerische Iterationsalgorithmen an, wie zum Beispiel das Newton-Verfahren¹, das auf die Gleichung (5) angewandt folgende iterative Gleichung für \hat{f}_j liefert:

$$\hat{f}_j^{(m+1)} = \hat{f}_j^{(m)} + \frac{\sum_{i=1}^{n+1-j} C_{i,j-1} \mathcal{Y} \left(\frac{C_{i,j} - \hat{f}_j^{(m)} C_{i,j-1}}{S_j} \right) S_j}{\sum_{i=1}^{n+1-j} C_{i,j-1}^2 \mathcal{Y}' \left(\frac{C_{i,j} - \hat{f}_j^{(m)} C_{i,j-1}}{S_j} \right)} \quad (11)$$

wobei²

$$S_j = 1,483 \cdot MAD_j$$

ein robuster Skalenschätzer für s_j ist und MAD_j den Median der absoluten Abweichungen nach einer Adjustierung darstellt:

$$MAD_j = \text{median} \left\{ \left| C_{i,j} - \hat{f}_j^{(0)} C_{i,j-1} \right|, i = 1, \dots, n+1-j \right\}.$$

$\hat{f}_j^{(0)}$ ist der Startwert für die Iteration (11), bestimmt durch die Gleichung:

$$\hat{f}_j^{(0)} = \text{median} \left\{ \frac{C_{i,j}}{C_{i,j-1}}, i = 1, \dots, n+1-j \right\}.$$

Die Iteration (11) liefert eine gute Lösung schon nach einem Schritt, was einen robusten Ein-Schritt-Schätzer für f_j bestimmt:

$$\hat{f}_j^R = \hat{f}_j^{(0)} + \frac{\sum_{i=1}^{n+1-j} C_{i,j-1} \mathcal{Y} \left(\frac{C_{i,j} - \hat{f}_j^{(0)} C_{i,j-1}}{S_j} \right) S_j}{\sum_{i=1}^{n+1-j} C_{i,j-1}^2 \mathcal{Y}' \left(\frac{C_{i,j} - \hat{f}_j^{(0)} C_{i,j-1}}{S_j} \right)}. \quad (12)$$

¹ Die Nullstelle einer differenzierbarer Funktion $f(x)$ lässt sich durch die Iterationsvorschrift:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}, \quad n = 0, 1, 2, \dots \text{ bestimmen.}$$

² Die Konstante 1.483 wurde bei E. Kremer (1997): a. a. O. vorgeschlagen, um die Verzerrung zu reduzieren.

$$y'(x) = 0, \quad \text{für } |x| > c$$
$$= 1, \quad \text{für } |x| \leq c,$$

entsprechend der Definition (8).

2.2.5 Numerische Berechnung

Mit dem Abwicklungsdreieck der angefallenen Schäden ergeben sich folgende Reserven für die Anfalljahre $i = 2, \dots, 7$:

Anf.Jahr	Reserve_CL	Reserve_Robust
2	3.719.078	3.719.078
3	9.973.396	9.747.806
4	28.306.992	27.293.837
5	47.653.105	47.599.172
6	100.148.912	103.239.618
7	217.244.607	216.598.503

Solange die Daten gutartig sind, sollten sich die Reserven nach beiden Verfahren kaum unterscheiden, was die obige Tabelle auch bestätigt.

Für die Abwicklungsfaktoren ergeben sich folgende Werte:

Abw. Periode	CL-Faktoren	Robuste Faktoren
1 - 2	1,251395	1,239252
2 - 3	1,119916	1,129714
3 - 4	1,101518	1,105533
4 - 5	1,062244	1,059402
5 - 6	1,031781	1,030570
6 - 7	1,022233	1,022233

Analog mit dem Dreieck der bezahlten Schäden, sind die Reserven für Anfalljahre $i = 2, \dots, 7$:

Anf.Jahr	Reserve_CL	Reserve_Robust
2	8.182.750	8.182.750
3	29.598.495	29.844.692
4	70.683.649	71.910.210
5	93.645.484	94.398.327
6	174.338.720	172.369.430
7	307.010.720	299.940.730

Für die Faktoren haben sich folgende Werte ergeben:

Abw. Periode	CL-Faktoren	Robuste Faktoren
1 - 2	1,509983	1,495325
2 - 3	1,154191	1,144377
3 - 4	1,177908	1,175802
4 - 5	1,163606	1,167738
5 - 6	1,132888	1,134483
6 - 7	1,062477	1,062477

Da beide Dreiecke keine Ausreißer-Werte erhielten, kann man an den Ergebnissen die Vorteile der robusten Faktoren nicht erkennen. Um die Wirkung der robusten Faktoren zu zeigen, führen wir in das Dreieck der bezahlten Schäden einen Ausreißer ein und beobachten die Reaktion beider Verfahren.

In Dreieck der bezahlten Schäden wurde anstelle von 80.060.344 im Anfalljahr 1 und Abwicklungsjahr 2 der viel höhere Betrag von 1.000.000.000 im Vergleich zu allen anderen Werten eingefügt. Die Verfahren wurden mit diesen neuen Daten noch einmal durchgeführt.

Für die Reserven ergeben sich damit folgende Werte:

Anf.Jahr	Reserve_CL	Reserve_Robust
2	8.182.750	8.182.750
3	29.598.495	29.844.692
4	70.683.649	71.910.210
5	93.645.484	94.398.327
6	-61.230.050	41.498.233
7	224.606.454	138.561.986

Mit folgenden Faktoren für die Berechnung:

Abw. Periode	CL-Faktoren	Robuste Faktoren
1 - 2	3,475580	1,519353
2 - 3	0,413663	0,734272
3 - 4	1,177908	1,175802
4 - 5	1,163606	1,167738
5 - 6	1,132888	1,134483
6 - 7	1,062477	1,062477

Während die Faktoren von Chain-Ladder auf die Veränderung extrem reagieren, verhalten sich die robusten Faktoren eher entsprechend dem allgemeinen Trend der Maße der Beobachtungen und werden vom Ausreißer nur wenig beeinflusst.

Die Auswirkungen des Ausreißers sind besonders an der Reserve für das Anfalljahr 6 deutlich zu erkennen, welche sogar negativ ist.

Es empfiehlt sich also in der Praxis beide Verfahren auf die Daten anzuwenden, weil dadurch die Unregelmäßigkeiten in der Struktur der Daten erkannt werden können, ähnliche Ergebnisse lassen auf gutartige Struktur der Daten schließen.

2.3 Stochastische Faktoren

Dieses Verfahren basiert auf der Annahme einer Wahrscheinlichkeitsverteilung der Abwicklungsfaktoren.

Mit dieser Annahme kann dann die Verteilung für den Endschaden ermittelt werden. Mit Hilfe von Simulationen erhält man daraus den empirischen Erwartungswert als Schätzer für den Endschaden und weitere Eigenschaften wie die empirische Varianz.

Wir werden hier drei Verteilungsannahmen untersuchen: Log-Normal-, Log-Gamma- und Log-inverse Gaußverteilung.

Mit Hilfe von Simulationen werden anschließend die p -Quantile für den Endschaden S über alle Anfalljahre ermittelt. p gibt dann die Wahrscheinlichkeit an, dass der Endschaden die Größe S nicht überschreitet.

Die Verteilungsannahmen werden anschließend mit dem Kolmogorov-Smirnov-Test überprüft, um das beste Verteilungsmodell für gegebene Dreiecke zu wählen.

Zur numerischen Berechnung werden das Tabellenkalkulationsprogramm-Excel und eine von der Firma *Averill M. Law & Associates* entwickelte statistische Software „ExpertFit“¹ verwendet.

2.3.1 Das Modell

Gegeben sei das Dreieck der kumulierten bezahlten oder angefallenen Schäden:

AnfJahr	AbwJahr					
	1	2	...	$n+1-i$...	n
1	$C_{1,1}$	$C_{1,2}$...	$C_{1,n+1-i}$...	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$...	$C_{2,n+1-i}$...	
⋮	⋮	⋮	...	⋮		
i	$C_{i,1}$	$C_{i,2}$...	$C_{i,n+1-i}$		
⋮	⋮	⋮				
$n-1$	$C_{n-1,1}$	$C_{n-1,2}$				
n	$C_{n,1}$					

Wir nehmen an, dass mit dem Abwicklungsjahr n die Schadenabwicklung abgeschlossen ist.

¹ Diese Software wurde mir für die Berechnung von der Firma *AON Jauch & Hübener, Hamburg* zur Verfügung gestellt.

Durch Dividieren der Spalte k durch die Elemente der Spalte $k-1$ erhalten wir das Dreieck der beobachteten Abwicklungsfaktoren:

AnfJahr	AbwJahr			
	1->2	2->3	...	$n-1 \rightarrow n$
1	$d_{1,2}$	$d_{1,3}$...	$d_{1,n}$
2	$d_{2,2}$	$d_{2,3}$...	
⋮	⋮	⋮		
$n-1$	$d_{n-1,2}$			

Unter dem Modell der stochastischen Abwicklungsfaktoren wird das Abwicklungsdreieck folgendermaßen vollendet:

AnfJahr	AbwJahr			
	1	2	...	n
1	I_1	$I_1 D_{1,2}$...	$I_1 \prod_{k=2}^n D_{1,k}$
2	I_2	$I_2 D_{2,2}$...	$I_2 \prod_{k=2}^n D_{2,k}$
⋮	⋮	⋮	...	
n	I_n	$I_n D_{n,2}$...	$I_n \prod_{k=2}^n D_{n,k}$

wobei $U_i = I_i \prod_{k=2}^n D_{i,k}$ der Endschaten vom Anfalljahr i ist. Der gesamte Schaden über alle

Anfalljahre ist dann $S_n = \sum_{i=1}^n U_i$, und die gesamte Reserve über alle Anfalljahre beträgt:

$R = S_n - \sum_{k=1}^n C_{k,n+1-k} \cdot I_i$ ist eine Zufallsvariable, die den Anfangsschaten für das Anfalljahr

i , $i = 1, \dots, n$ darstellt.

$D_{i,j}$ ist eine Zufallsvariable, die dem Schadenzuwachs für das Anfalljahr i vom Abwicklungsjahr $j-1$ zum Abwicklungsjahr j darstellt. I_i und $\{D_{i,j}, j = 2, \dots, n\}$ sind stochastisch unabhängig.

Seien $D_{i,j}$ $i=1,\dots,n$, j fest, unabhängig identisch verteilt mit der Verteilungsfunktion $F(d, \mathbf{q}_j)$ und unabhängig für $j=2,\dots,n$. Die Annahme der Unabhängigkeit von $D_{i,j}$ und $D_{i,k}$ mit $k \neq j$, ermöglicht eine ziemlich einfache Berechnung der bedingten Verteilung vom Endschaten $U_i: P^{U_i|I_i}$.

Lemma:

Seien I_i und $D_{i,j}$ $i=1,\dots,n$, $j=2,\dots,n$ zwei reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$.

Es gelte für festes i : I_i und $\{D_{i,j}, j=2, \dots, n\}$ sind stochastisch unabhängig.

Dann gilt:

a) $E(U_i | I_i) = I_i \cdot \left\{ \prod_{k=2}^n D_{i,k} \right\}$, $i=1,\dots,n$ P- f.s.

b) $E(U_i | I_i = C_{i,1}) = C_{i,1} \cdot \left\{ \prod_{k=2}^n D_{i,k} \right\}$, $i=1,\dots,n$ P- f.s.

Beweis:¹

a) $E(U_i | I_i) = E(I_i \cdot \left\{ \prod_{k=2}^n D_{i,k} \right\} | I_i)$ Satz 1, c)

$$I_i \cdot E(\left\{ \prod_{k=2}^n D_{i,k} \right\} | I_i) \xrightarrow{\text{Satz 1, e)}} I_i \cdot E(\left\{ \prod_{k=2}^n D_{i,k} \right\})$$

b) Folgt aus Ersetzungslemma (2).

Q.E.D.

Für jede Spalte wird der Parameter \mathbf{q}_j geschätzt durch $\hat{\mathbf{q}}_j(d_{1,j}, \dots, d_{n+1-j,j})$, $j=2,\dots,n$ als Funktion der beobachteten Abwicklungsfaktoren.

Vergleicht man die durch das Modell angepassten Werte mit den beobachteten, wird so eine Aussage über die Güte der Anpassung möglich.

¹ Die Hilfssätze, die hier erwähnt sind, findet man im Anhang A.

Der gesamte Schaden $S_n = \sum_{k=1}^n U_k$ ist auch eine Zufallsvariable, deren bedingter

Erwartungswert $E(S_n | (I_1, \dots, I_n))$ gleich $\sum_{k=1}^n E(U_k | I_k)$ und nach Lemma a) gleich

$\sum_{k=1}^n I_k \cdot E(\{\prod_{l=2}^n D_{k,l}\})$ ist. Die Eigenschaften von U_i für jedes Anfalljahr werden mit Hilfe von Simulationen ermittelt.

2.3.2 Endscha-den im Log-Normal-Modell.

Ist eine Zufallsvariable X normalverteilt mit Erwartungswert m und Varianz s^2 , dann ist $Y = e^X$ Log-normalverteilt mit Dichte:

$$f(y) = \frac{1}{\sqrt{2\pi s y}} \exp\left(\frac{-(\ln y - m)^2}{2s^2}\right), \quad y > 0. \quad (13)$$

Für eine Log-normalverteilte Zufallsvariable ist das Moment r -ter Ordnung:

$$E(Y^r) = E((e^X)^r) = M_X(r) = \exp(rm + \frac{1}{2}r^2s^2)^1 \quad (14)$$

mit $M_X(r)$ als momenterzeugender Funktion einer normalverteilten Zufallsvariablen.

Die Log-Normalverteilung hat folgende nützliche Eigenschaft: Sind Y_1, \dots, Y_n unabhängige

Log-normalverteilte Variablen, so dass $Y_i = e^{X_i}$, so ist $P_n = \prod_{j=1}^n (a_j Y_j)$ für $a_j \geq 0$ wieder Log-

normalverteilt mit Parametern: ²

$$m = \sum_{j=1}^n (m_j + \ln(a_j)) \quad \text{und} \quad s^2 = \sum_{j=1}^n s_j^2 \quad (15)$$

¹ Daraus folgt für den Erwartungswert $E(Y) = M_X(1) = \exp\left(m + \frac{1}{2}s^2\right)$ und für die Varianz

$$Var(Y) = E(Y^2) - E(Y)^2 = M_X(2) - E(Y)^2 = \exp\left(2m + s^2\right)\left(\exp(s^2) - 1\right)$$

² folgt aus der bekannten Eigenschaft einer normalverteilten Variable: Die Summe unabhängiger normalverteilten Variablen ist wieder normalverteilt und der Exponent einer normalverteilten Variable ist Log-normalverteilt.

Wir nehmen an, dass alle Spalten von Abwicklungsfaktoren $D_{i,j}$ $i = 1, \dots, n$ unabhängig Log-normalverteilt mit Parametern \mathbf{m}_j und \mathbf{s}_j^2 sind. Aus (15) folgt, dass der Endschaden

$U_i = I_i \prod_{j=2}^n D_{ij}$ Log-normalverteilt ist mit den Parametern: $\mathbf{m}_{(i)} = \ln(I_i) + \sum_{j=2}^n \mathbf{m}_j$ und

$\mathbf{s}^2 = \sum_{j=2}^n \mathbf{s}_j^2$. Nach (13) hat dann die Verteilungsfunktion des Endschadens folgende Gestalt:

$$g_i(u_i) = \frac{1}{\sqrt{2\mathbf{p}\mathbf{s}u_i}} \exp\left(-\frac{(\ln u_i - \mathbf{m}_{(i)})^2}{2\mathbf{s}^2}\right), \quad u_i > 0. \quad (16)$$

Aus (14) folgt für den bedingten Erwartungswert des Endschadens für Anfalljahr i :

$$E(U_i | I_i) = I_i \exp\left(\sum_{j=2}^n \mathbf{m}_j + \frac{1}{2} \sum_{j=2}^n \mathbf{s}_j^2\right).$$

Unter unseren Modellannahmen sind dann $d_{i,j}$ $i = 1, \dots, n+1-j$, $j = 2, \dots, n$ Stichproben von Log-normalverteilten Variablen mit Parametern \mathbf{m}_j und \mathbf{s}_j^2 mit dem Maximum-Likelihood-Schätzer für \mathbf{m}_j :

$$\hat{\mathbf{m}}_j = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} \ln(d_{ij}) \quad (17)$$

und für \mathbf{s}_j^2 :

$$\hat{\mathbf{s}}_j^2 = \frac{\sum_{i=1}^{n-j+1} (\ln(d_{ij}) - \hat{\mathbf{m}}_j)^2}{n-j+1} \quad (18)$$

Der Schätzer $\hat{\mathbf{m}}_j$ ist unverzerrt, und der Schätzer $\hat{\mathbf{s}}_j^2$ ist asymptotisch unverzerrt.

Satz 1: Der erwartungstreue und varianzminimale Schätzer von $E(U_i | I_i)$ ist

$$\hat{\mathbf{m}}_{(i)} = I_i \exp\left(\sum_{j=2}^n \hat{\mathbf{m}}_j\right) \prod_{j=2}^n {}_0F_1\left(\frac{(n-j)}{2}; \frac{(n-j)}{4(n-j+1)} SS_j\right)$$

wobei $SS_j = \sum_{i=1}^{n-j+1} (\ln(d_{ij}) - \hat{\mathbf{m}}_j)^2$ und ${}_0F_1(\mathbf{h}; z) = \sum_{t=0}^{\infty} \frac{z^t}{t!} \frac{\Gamma(\mathbf{h})}{\Gamma(\mathbf{h}+t)}$ (eine verallgemeinerte

Hypergeometrische Funktion) sind.

Beweis:

Diese Schätzer ist varianzminimal aufgrund der Eigenschaften eines ML-Schätzers.¹

Es bleibt die Erwartungstreue zu zeigen, d. h.

$$E[\hat{\mathbf{m}}_{(i)}] = \mathbf{m}_{(i)} = I_i \exp\left(\sum_{j=2}^n \mathbf{m}_j + \frac{1}{2} \sum_{j=2}^n \mathbf{s}_j^2\right). \quad (19)$$

Es gilt: $\hat{\mathbf{m}}_j \sim N\left(\mathbf{m}_j, \frac{\mathbf{s}_j^2}{n-j+1}\right) \Rightarrow e^{\hat{\mathbf{m}}_j} \sim LN$ mit $E[e^{\hat{\mathbf{m}}_j}] = e^{\mathbf{m}_j + \frac{\mathbf{s}_j^2}{2(n-j+1)}}$, und daraus folgt:

$$E\left[\prod_{j=2}^n e^{\hat{\mathbf{m}}_j}\right] = \prod_{j=2}^n E[e^{\hat{\mathbf{m}}_j}] = \prod_{j=2}^n e^{\mathbf{m}_j + \frac{\mathbf{s}_j^2}{2(n-j+1)}} = e^{\sum_{j=2}^n \left(\mathbf{m}_j + \frac{\mathbf{s}_j^2}{2(n-j+1)}\right)}.$$

Sei $h\left(\sum_j SS_j\right) = \prod_{j=2}^n {}_0F_1\left(\frac{(n-j)}{2}; \frac{(n-j)}{4(n-j+1)} SS_j\right)$, dann gilt

$$E[\hat{\mathbf{m}}_{(i)}] = E\left[I_i e^{\sum_{j=2}^n \hat{\mathbf{m}}_j} h\left(\sum_j SS_j\right)\right] = I_i e^{\sum_{j=2}^n \mathbf{m}_j + \frac{1}{2} \sum_{j=2}^n \frac{\mathbf{s}_j^2}{2(n-j+1)}} E\left[h\left(\sum_j SS_j\right)\right];$$

und nach (19) bleibt nur noch zu zeigen, dass

$$E\left[h\left(\sum_j SS_j\right)\right] = e^{\sum_j \frac{\mathbf{s}_j^2(n-j)}{2(n-j+1)}}; \quad (20)$$

$$\begin{aligned} E\left[h\left(\sum_j SS_j\right)\right] &= E\left[\prod_{j=2}^n \sum_{t=0}^{\infty} \frac{SS_j^t (n-j)^t}{4^t (n-j+1)^t t!} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(t + \frac{n-j}{2}\right)}\right] = \\ &= \prod_{j=2}^n \sum_{t=0}^{\infty} \frac{E[SS_j^t] (n-j)^t}{4^t (n-j+1)^t t!} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(t + \frac{n-j}{2}\right)} \end{aligned} \quad (21)$$

Wegen $\frac{SS_j}{\mathbf{s}_j^2} \sim \chi^2(n-j)$ folgt

¹ Nach den Aussagen der statistischen Theorie ist der Maximum-Likelihood-Schätzer ein effizienter Schätzer, jedoch nicht notwendigerweise ein erwartungstreu.

$$\begin{aligned}
 E \left[\left(\frac{SS_j}{\mathbf{s}_j^2} \right)^k \right] &= \int_0^\infty \frac{\left(\frac{SS_j}{\mathbf{s}_j^2} \right)^k \left(\frac{SS_j}{\mathbf{s}_j^2} \right)^{\frac{n-j}{2}-1} e^{-\frac{SS_j}{2\mathbf{s}_j^2}}}{2^{\frac{n-j}{2}} \Gamma\left(\frac{n-j}{2}\right)} d\left[\frac{SS_j}{\mathbf{s}_j^2}\right] = \\
 &= \frac{2^{k+\frac{n-j}{2}} \Gamma\left(k+\frac{n-j}{2}\right)}{2^{\frac{n-j}{2}} \Gamma\left(\frac{n-j}{2}\right)} \int_0^\infty \frac{\left(\frac{SS_j}{\mathbf{s}_j^2} \right)^{\frac{2k+n-j}{2}-1} e^{-\frac{SS_j}{2\mathbf{s}_j^2}}}{2^{\frac{k+n-j}{2}} \Gamma\left(\frac{2k+n-j}{2}\right)} d\left[\frac{SS_j}{\mathbf{s}_j^2}\right] = \\
 &= 2^k \frac{\Gamma\left(k+\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}\right)} \\
 \Rightarrow E[SS_j^k] &= 2^k \mathbf{s}_j^{2k} \frac{\Gamma\left(k+\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}\right)}.
 \end{aligned}$$

Einsetzen in (21) \Rightarrow

$$\begin{aligned}
 \prod_{j=2}^n \sum_{t=0}^{\infty} \frac{\mathbf{s}_j^{2t} 2^t (n-j)^t}{4^t (n-j+1)^t t!} \frac{\Gamma\left(\frac{n-j}{2}\right)}{\Gamma\left(t+\frac{n-j}{2}\right)} \frac{\Gamma\left(t+\frac{n-j}{2}\right)}{\Gamma\left(\frac{n-j}{2}\right)} &= \\
 \prod_{j=2}^n \sum_{t=0}^{\infty} \left(\frac{\mathbf{s}_j^2 (n-j)}{2(n-j+1)} \right)^t / t! &= \prod_{j=2}^n e^{\frac{\mathbf{s}_j^2 (n-j)}{2(n-j+1)}} = e^{\sum_{j=2}^n \frac{\mathbf{s}_j^2 (n-j)}{2(n-j+1)}}
 \end{aligned}$$

Q.E.D

In der Praxis ist die Summation bis zum fünften oder sechsten Term hinreichend genau.

Mit geschätzten $\hat{\mathbf{m}}_{(1)}, \dots, \hat{\mathbf{m}}_{(n)}$ lässt sich der gesamte Endschaden leicht ausrechnen:

$\hat{\mathbf{m}}^{LN} = \sum_{k=1}^n \hat{\mathbf{m}}_{(k)}$. Mit Hilfe der Simulation (s. unten), lassen sich weitere Größen wie die

empirische Varianz und Quantile von $\mathbf{m}_{(i)}$ schätzen.

Simulationsalgorithmus zum Log-Normal-Modell:

1. Generiere 2 rechteckverteilte Variablen: $U_1, U_2 \sim R(0,1)$
2. Mit Box-Muller Transformation erhalte 2 unabhängige standardnormalverteilte Variablen:

$$Z_1 = \sqrt{-2\ln(U_1)} \cos(2\pi U_2) \text{ und } Z_2 = \sqrt{-2\ln(U_1)} \sin(2\pi U_2)$$

3. Mit Transformation erhalte 2 Log-normalverteilte Variablen mit Parametern $\hat{\mathbf{m}}$ und $\hat{\mathbf{S}}^2$:

$$Y_1 = \exp(\hat{\mathbf{m}} + \hat{\mathbf{S}} Z_1) \text{ und } Y_2 = \exp(\hat{\mathbf{m}} + \hat{\mathbf{S}} Z_2), \text{ wobei}$$

$$\hat{\mathbf{m}} = \ln C_{k,1} + \sum_{j=2}^n \hat{\mathbf{m}}_j \text{ und } \hat{\mathbf{S}}^2 = \sum_{j=2}^n \hat{\mathbf{S}}_j^2$$

mit $\hat{\mathbf{m}}_j$ und $\hat{\mathbf{S}}_j^2$ aus (17) und (18) entsprechend.

4. Für jedes Anfalljahr wiederhole Schritte 1-3 5.000-mal und erhalte 10.000 Log-normalverteilte Zufallsvariablen $U_{k1}^*, U_{k2}^*, \dots, U_{k,10.000}^*$.
5. Mit diesen 10.000 Werten können Erwartungswert, Standardabweichung und Quantile vom Endscha-den für jedes Anfalljahr geschätzt werden.

2.3.3 Endscha-den im Log-Gamma-Modell

Dieses Modell setzt voraus, dass die Abwicklungsfaktoren einer Log-Gammaverteilung unterliegen.

Aufgrund der Eigenschaften dieser Verteilung gelten im Vergleich zum Log-Normal-Modell folgende Einschränkungen:

- Die Abwicklungsfaktoren sollen größer als 1, also nur auf kumulierte Daten anwendbar sein.
- Die Schätzung der Parameter geschieht hier simultan für das ganze Abwicklungsdreieck. Beim Log-Normal-Modell wurden die Parameter getrennt für jede Spalte berechnet.

Die Log-Gammaverteilung wird von der Gammaverteilung abgeleitet.

Die Dichtefunktion einer gammaverteilten Variablen mit Parametern \mathbf{a} und \mathbf{l} hat folgende Gestalt:

$$f(x) = \frac{x^{\mathbf{a}-1} e^{-I x} I^{\mathbf{a}}}{\Gamma(\mathbf{a})}, \quad x \geq 0 \quad (22)$$

mit momenterzeugender Funktion:

$$M_X(t) = E[e^{tX}] = \left[\frac{I}{I-t} \right]^{\mathbf{a}}, \quad t < I. \quad (23)$$

Die Gammaverteilung hat folgende wichtige Eigenschaften:

1. Sind X_1, X_2, \dots, X_n unabhängige Zufallsvariablen mit

$$X_i \sim \text{Gamma}(\mathbf{a}_i, I), \quad i=1, \dots, n, \text{ so ist } \sum_{i=1}^n X_i \sim \text{Gamma}(\mathbf{a}, I) \text{ mit } \mathbf{a} = \sum_{i=1}^n \mathbf{a}_i.$$

2. Ist $X \sim \text{Gamma}(\mathbf{a}, I)$ verteilt, so ist $Z = X + \mathbf{d}$ verschoben gammaverteilt mit Lageparameter \mathbf{d} und Dichtefunktion gemäß:

$$h(z) = \frac{(z-\mathbf{d})^{\mathbf{a}-1} e^{-I(z-\mathbf{d})} I^{\mathbf{a}}}{\Gamma(\mathbf{a})}, \quad z \geq \mathbf{d} \quad (24)$$

und momenterzeugender Funktion:

$$M_Z(t) = E[e^{tZ}] = e^{t\mathbf{d}} \left[\frac{I}{I-t} \right]^{\mathbf{a}}, \quad t < I \quad (25)$$

3. Ist $X \sim \text{Gamma}(\mathbf{a}, I)$ verteilt, so ist $Y = e^X$ nach dem Transformationssatz für Lebesgue-Dichten Log-gammaverteilt mit Dichte:

$$g(y) = f(\ln y) \frac{1}{y} = \frac{(\ln y)^{\mathbf{a}-1} y^{-I-1} I^{\mathbf{a}}}{\Gamma(\mathbf{a})}, \quad y \geq 1 \quad (26)$$

Für unser Abwicklungsdreieck nehmen wir an, dass jede Spalte $D_{i,j}$, $i=1, \dots, n$, $j=2, \dots, n$ von Abwicklungsfaktoren unabhängig und identisch Log-gammaverteilt sind mit Indexparameter \mathbf{a}_j und Skalenparameter I .

Aus (24) und (26) folgt dann, dass der Endschaden vom Anfalljahr i : $U_i = I_i \prod_{j=2}^n D_{ij}$ Log-gammaverteilt ist mit der Dichte:

$$g_i(u_i) = \frac{(\ln u_i - \ln I_i)^{\mathbf{a}-1} u_i^{-I-1} I_i^{\mathbf{a}}}{\Gamma(\mathbf{a})}, \quad u_i \geq I_i \quad (27)$$

wobei $\mathbf{a} = \sum_{j=2}^n \mathbf{a}_j$.

Beweis:

$\ln U_i = \ln I_i + \sum \ln D_{ij} = \mathbf{d} + X \sim$ verschoben gammaverteilt nach (3), dann folgt mit (2) und (1) die Behauptung.

Q.E.D.

Aus (25) und (27) gilt für den Erwartungswert des Endschadens vom Anfalljahr i :

$$E[U_i|I_i] = I_i \left[\frac{\mathbf{I}}{\mathbf{I}-1} \right]^{\mathbf{a}}, \quad \mathbf{I} > 1 \quad (28)$$

Bei $\mathbf{I} \leq 1$ ist der Erwartungswert der Verteilung nicht definiert.

Beweis:

$$E[U_i|I_i] = E[e^{X_i}] = E[e^{t(\ln I_i + \sum \ln D_{ij})}] = E[e^{t(\mathbf{d} + X)}] = e^{\mathbf{d}} \left[\frac{\mathbf{I}}{\mathbf{I}-1} \right]^{\mathbf{a}} = I_i \left[\frac{\mathbf{I}}{\mathbf{I}-1} \right]^{\mathbf{a}}, \quad \mathbf{I} > 1.$$

Q.E.D.

Die Schätzung der Parameter $\mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{I}$ erfolgt numerisch. Aus (26) erhält man die Likelihood-Funktion:

$$L(\mathbf{a}_{\{j\}}, \mathbf{I}; d_{\{i,j\}}) = \prod_{j=2}^n \prod_{i=1}^{n-j+1} \frac{(\ln d_{i,j})^{\mathbf{a}_j-1} d_{i,j}^{-\mathbf{I}-1} \mathbf{I}^{\mathbf{a}_j}}{\Gamma(\mathbf{a}_j)}$$

und dann die Log-Likelihood-Funktion:

$$\ln L(\mathbf{a}_{\{j\}}, \mathbf{I}; d_{\{i,j\}}) = \sum_{j=2}^n \sum_{i=1}^{n-j+1} [-\mathbf{I} \ln d_{i,j} - \ln d_{i,j} + (\mathbf{a}_j - 1) \ln(\ln d_{i,j}) + \mathbf{a}_j \ln \mathbf{I} - \ln \Gamma(\mathbf{a}_j)]$$

Maximieren dieser Gleichung, indem man die Ableitungen nach \mathbf{a}_j und \mathbf{I} Null setzt, liefert n Gleichungen mit n Unbekannten:

$$\hat{\mathbf{I}} = \frac{\sum_{j=2}^n (n-j+1) \mathbf{a}_j}{d_{..}} \quad (29)$$

$$\mathbf{y}(\hat{\mathbf{a}}_j) = \ln \hat{\mathbf{I}} + \bar{d}_{.j}, \quad j = 2, \dots, n \quad (30)$$

wobei $\mathbf{y}(x) = \frac{d \ln \Gamma(x)}{dx}$ die Digamma-Funktion bezeichnet, und $\bar{d}_{.j} = \frac{\sum_{i=1}^{n-j+1} \ln(\ln d_{i,j})}{(n-j+1)}$ und

$$d_{..} = \sum_{j=2}^n \sum_{i=1}^{n-j+1} \ln d_{i,j}.$$

Beweis:

$$\frac{d(\ln L)}{d\mathbf{a}_j} = \sum_{i=1}^{n-j+1} \ln(\ln d_{i,j}) + (n-j+1)\ln \mathbf{I} - (n-j+1)\mathbf{y}(\mathbf{a}_j)$$

$$\Rightarrow \mathbf{y}(\hat{\mathbf{a}}_j) = \frac{\sum_{i=1}^{n-j+1} \ln(\ln d_{i,j})}{n-j+1} + \ln \hat{\mathbf{I}}$$

$$\frac{d(\ln L)}{d\mathbf{I}} = \sum_{j=2}^n \mathbf{a}_j \frac{n-j+1}{\mathbf{I}} - \sum_{i=1}^{n-j+1} \sum_{j=2}^n \ln d_{i,j}$$

$$\Rightarrow \hat{\mathbf{I}} = \frac{\sum_{j=2}^n (n-j+1)\hat{\mathbf{a}}_j}{d_{..}}$$

Q.E.D.

Die exakten Werte der Digamma-Funktion erhält man z.B. aus *Abramowitz und Stegun (1965), Handbook of Mathematical Functions* oder man berechnet die Werte approximativ durch:

$$\begin{aligned} \mathbf{y}(x) \sim \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \frac{1}{240x^8} - \frac{1}{132x^{10}} + \frac{691}{32.760x^{12}} - \\ - \frac{1}{12x^{14}} + \frac{255}{28.936x^{16}} + \dots \end{aligned} \quad (31)$$

Das Newton-Verfahren, zum Lösen von (29) und (30):

Sei

$$\mathbf{F}(\hat{\mathbf{a}}) = (f_2(\hat{\mathbf{a}}), f_3(\hat{\mathbf{a}}), \dots, f_n(\hat{\mathbf{a}}))^T$$

ein $(n-1)$ -dimensionaler Vektor mit:

$$f_k(\hat{\mathbf{a}}) = f_k(\hat{\mathbf{a}}_2, \hat{\mathbf{a}}_3, \dots, \hat{\mathbf{a}}_n) = \mathbf{y}(\hat{\mathbf{a}}_k) - \ln \left[\frac{\sum_{j=2}^n (n-j+1)\hat{\mathbf{a}}_j}{d_{..}} \right] - \bar{d}_k$$

für $k = 2, \dots, n$.

Die Matrix der Ableitungen ist gegeben durch:

$$\mathbf{J}(\hat{\mathbf{a}}) = \mathbf{y}'(\hat{\mathbf{a}}) - \frac{1}{\sum_{j=2}^n (n-j+1)\hat{\mathbf{a}}_j} \mathbf{R}$$

mit

$$\mathbf{y}'(\hat{\mathbf{a}}) = \begin{pmatrix} \mathbf{y}'(\hat{\mathbf{a}}_2) & 0 & \cdots & 0 \\ 0 & \mathbf{y}'(\hat{\mathbf{a}}_3) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{y}'(\hat{\mathbf{a}}_n) \end{pmatrix}$$

und

$$\mathbf{R} = \begin{pmatrix} n-1 & n-1 & \cdots & n-1 \\ n-2 & n-2 & \cdots & n-2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Die Ableitungen $\mathbf{y}'(x) = \frac{d\mathbf{y}(x)}{dx}$ erhält man durch einfaches Differenzieren von (31).

Die m -te Iteration erzeugt folgenden Vektor der Parameter:

$$(\hat{\mathbf{a}}_{2,m}, \hat{\mathbf{a}}_{3,m}, \dots, \hat{\mathbf{a}}_{n,m})^T = (\hat{\mathbf{a}}_{2,m-1}, \hat{\mathbf{a}}_{3,m-1}, \dots, \hat{\mathbf{a}}_{n,m-1})^T - \mathbf{F}(\hat{\mathbf{a}}_{m-1}) \cdot \mathbf{J}^{-1}(\hat{\mathbf{a}}_{m-1}) \quad (32)$$

Die Startwerte für die Iteration (32) werden nach der Momentenmethode durch

$$\hat{\mathbf{a}}_{j0} = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} \ln d_{i,j}$$

bestimmt.

Nach hinreichend vielen Iterationsschritten erhält man die Parameter $\hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_n$, und durch

Einsetzen dieser Parameter in (29) kann man den Parameter $\hat{\mathbf{I}}$ ausrechnen.

Den endgültigen Endschaten für das Gesamtdreieck erhält man durch:

$$\hat{\mathbf{m}}^{LG} = \left[\frac{\hat{\mathbf{I}}}{\hat{\mathbf{I}} - 1} \right]^{\hat{\mathbf{a}}} \sum_{i=1}^n I_i \quad \text{wobei} \quad \hat{\mathbf{a}} = \sum_{j=2}^n \hat{\mathbf{a}}_j.$$

Da dieser Schätzer verzerrt ist, verwendet man besser das Stichprobenmittel aus der Simulation. Sind $S_1^*, S_2^*, \dots, S_{10.000}^*$ mit der Simulation (s. unten) erzeugte Beobachtungen, so

folgt für den Gesamtschaten:

$$\hat{\mathbf{m}}^{LG} = \frac{1}{10.000} \sum_{k=1}^{10.000} S_k^* .$$

Simulationsalgorithmus im Log-Gamma Fall:

1. Erzeuge gammaverteilte Zufallsvariable X mit Parametern \hat{a} und $l=1$ (s. unten).
Unterschiedliche Algorithmen sollen verwendet werden für $\hat{a} < 1$ und $\hat{a} \geq 1$ (s. unten).
2. Durch Transformation erhalte Log-gammaverteilte Variable Y mit Parametern \hat{a}, \hat{l} und Lageparameter d :

$$Y = \exp\left(\frac{1}{\hat{l}} X + d\right).$$

3. Für jedes Anfalljahr $k = 1, \dots, n$ wiederhole Schritte 1-3 10.000 mal und erhalte 10.000 Log-gammaverteilte Zufallsvariablen $U_{k1}^*, U_{k2}^*, \dots, U_{k,10.000}^*$, wobei

$$d = \ln C_{k,1} \quad \text{und} \quad \hat{a} = \sum_{j=2}^n \hat{a}_j$$

mit \hat{l} und \hat{a}_j aus den Ergebnissen des Newton-Verfahrens. Damit ist der Gesamtenschaden: $S_k^* = \sum_{j=1}^n U_{j,k}^*, k = 1, \dots, 10.000$.

4. Mit diesen 10.000 Werten können Erwartungswert, Standardabweichung und Quantile geschätzt werden.

Simulationsalgorithmus einer gammaverteilten Variablen X :¹

Fall 1: $\hat{a} \geq 1$.

1. Setze: $a = \frac{1}{\sqrt{2a-1}}$; $b = a - \ln 4$; $q = a + \frac{1}{a}$; $q = 4,5$; $d = 1 + \ln q$.
2. Erzeuge 2 rechteckverteilte Variablen: $U_1, U_2 \sim R(0,1)$.
3. Definiere $V = a \cdot \ln[U_1/(1-U_1)]$; $Y = a \cdot e^V$; $Z = U_1^2 U_2$; $W = b + qV - Y$.
4. Ist $W + d - qZ \geq 0$, setze $\mathbf{X}=\mathbf{Y}$; Abbruch.
Sonst gehe zu 5.
5. Ist $W \geq \ln Z$, setze $\mathbf{X}=\mathbf{Y}$.
Sonst gehe zu 2.

¹ Der Algorithmus wurde aus A. Law, W. Kelton (1991), Simulation Modelling und Analysis, McGraw-Hill, S. 487 f, entnommen.

Fall 2: $0 < \hat{a} < 1$.

1. Setze: $b = \frac{e + \mathbf{a}}{e}$.
2. Generiere eine rechteckverteilte Variable $U_1 \sim R(0,1)$, setze $P = bU_1$.
Ist $P > 1$, gehe zu 4,
sonst weiter mit 3.
3. Setze $Y = P^{1/\mathbf{a}}$ und generiere $U_2 \sim R(0,1)$.
Ist $U_2 \leq e^{-Y}$, setze $\mathbf{X} = \mathbf{Y}$.
Sonst gehe zu 2.
4. Setze $Y = -\ln[(b - P)/\mathbf{a}]$ und generiere $U_2 \sim R(0,1)$.
Ist $U_2 \leq Y^{\mathbf{a}-1}$, setze $\mathbf{X} = \mathbf{Y}$.
Sonst gehe zu 2.

2.3.4 Endscha-den im Log-inversen gauschen Modell

Dieses Modell setzt voraus, dass die Abwicklungsfaktoren einer Log-inversen gauschen (Log-IG-) Verteilung unterliegen.

Aufgrund der Eigenschaften dieser Verteilung gelten analog zum Log-Gamma-Modell folgende Einschrnkungen:

- Die Abwicklungsfaktoren sollen groer als 1 sein.
- Die Schtzung der Parameter erfolgt simultan.

Die Log-IG-Verteilung wird von der inversen Gauverteilung (IG) induziert.

Die Dichtefunktion einer IG-verteilten Variablen mit Parametern \mathbf{m} und \mathbf{b} hat folgende Gestalt:

$$f(x) = \mathbf{m} \sqrt{\frac{\mathbf{b}}{2\mathbf{p}}} x^{-1,5} \exp\left(\frac{-\mathbf{b}(x - \mathbf{m})^2}{2x}\right), \quad x > 0 \quad (33)$$

mit der zugehrigen momenterzeugenden Funktion gem:

$$M_X(t) = E[e^{tX}] = \exp[\mathbf{m}t - \mathbf{m}t \sqrt{1 - \frac{2}{\mathbf{b}}t}], \quad t \leq \frac{\mathbf{b}}{2}. \quad (34)$$

Die IG-Verteilungsfunktion hat folgende wichtige Eigenschaften:

1. Ist $X \sim IG(\mathbf{m}, \mathbf{b})$ verteilt, so ist $Z = X + \mathbf{d}$ verschoben IG-verteilt mit Lageparameter $\mathbf{d} \geq 0$ und Dichtefunktion gemäß:

$$h(z) = \mathbf{m} \sqrt{\frac{\mathbf{b}}{2\mathbf{p}}} (z - \mathbf{d})^{-1,5} \exp\left(\frac{-\mathbf{b}(z - \mathbf{d} - \mathbf{m})^2}{2(z - \mathbf{d})}\right), \quad z > \mathbf{d} \quad (35)$$

und momentenerzeugender Funktion:

$$M_Z(t) = E[e^{tZ}] = \exp[\mathbf{m}\mathbf{b} + t\mathbf{d} - \mathbf{m}\mathbf{b} \sqrt{(1 - \frac{2}{\mathbf{b}}t)}], \quad t \leq \frac{\mathbf{b}}{2}. \quad (36)$$

2. Sind X_1, X_2, \dots, X_n unabhängige Zufallsvariablen mit

$$X_i \sim IG(\mathbf{m}_i, \mathbf{b}), \quad i = 1, \dots, n, \text{ so ist } T_n = \sum_{i=1}^n (X_i + \mathbf{d}_i) \sim IG(\mathbf{m}, \mathbf{b}) \text{ mit } \mathbf{m} = \sum_{i=1}^n \mathbf{m}_i$$

$$\text{und Lageparameter } \mathbf{d} = \sum_{i=1}^n \mathbf{d}_i.$$

3. Ist $X \sim IG(\mathbf{m}, \mathbf{b})$, so ist $Y = e^X$ Log-IG-verteilt mit Dichte:

$$g(y) = f(\ln y) \frac{1}{y} = \mathbf{m} \sqrt{\frac{\mathbf{b}}{2\mathbf{p}}} (\ln y)^{-1,5} y^{-(1+0,5\mathbf{b})} \exp\left(\mathbf{m}\mathbf{b} - \frac{\mathbf{b}\mathbf{m}^2}{2\ln y}\right), \quad y > 1 \quad (37)$$

Für unser Abwicklungsdreieck nehmen wir an, dass alle Spalten $D_{i,j}$, $i = 1, \dots, n$ von Abwicklungsfaktoren unabhängig und identisch Log-IG-verteilt sind mit Parametern \mathbf{m}_j und \mathbf{b} .

Aus (35) und (37) folgt dann, dass der Endschaten vom Anfalljahr i : $U_i = I_i \prod_{j=2}^n D_{ij}$

Log-IG-verteilt ist mit Lageparameter $\ln I_i$ und Dichte

$$g_i(u_i) = \mathbf{m} \sqrt{\frac{\mathbf{b}}{2\mathbf{p}}} I_i^{0,5\mathbf{b}} (\ln u_i - \ln I_i)^{-1,5} u_i^{-(1+0,5\mathbf{b})} \exp\left(\mathbf{m}\mathbf{b} - \frac{\mathbf{b}\mathbf{m}^2}{2(\ln u_i - \ln I_i)}\right), \quad u_i > I_i \quad (38)$$

mit $\mathbf{m} = \sum_{j=2}^n \mathbf{m}_j$.

Mit (36) und (38) gilt für den Erwartungswert des Endschatens vom Anfalljahr i :

$$E[U_i | I_i] = I_i \exp\left[\mathbf{b} \left(1 - \left(1 - \frac{2}{\mathbf{b}}\right)^{0,5}\right) \sum_{k=1}^n \mathbf{m}_k\right], \quad \mathbf{b} \geq 2 \quad (39)$$

Bei $\mathbf{b} < 2$ ist der Erwartungswert der Verteilung nicht definiert.

Wie im Log-Gamma-Modell erfolgt die Schätzung der Parameter $\mathbf{m}_2, \dots, \mathbf{m}_n, \mathbf{b}$ numerisch. Aus (37) erhält man die Log-Likelihood-Funktion:

$$\ln L(\mathbf{m}_{\{j\}}, \mathbf{b}; d_{\{i,j\}}) = \sum_{j=2}^n \sum_{i=1}^{n-j+1} \left[\ln \mathbf{m}_j + \frac{1}{2} \ln \mathbf{b} - \frac{1}{2} \mathbf{b} \frac{(\ln d_{i,j} - \mathbf{m}_j)^2}{\ln d_{i,j}} \right]$$

Maximieren dieser Gleichung, in dem man die Ableitungen nach \mathbf{m}_j und \mathbf{b} zur 0 setzt, liefert n Gleichungen mit n Unbekannten:

$$\frac{1}{\hat{\mathbf{b}}} = \frac{\sum_{j=2}^n \sum_{i=1}^{n-j+1} [(\ln d_{ij} - \hat{\mathbf{m}}_j)^2 / \ln d_{ij}]}{0,5n(n-1)^2} \quad (40)$$

$$\hat{\mathbf{m}}_j^2 \sum_{i=1}^{n-j+1} \frac{1}{\ln d_{ij}} - (n-j+1)\hat{\mathbf{m}}_j = \frac{n-j+1}{\hat{\mathbf{b}}}, \quad j = 2, \dots, n \quad (41)$$

Beweis:

$$\frac{d(\text{Log}L)}{d\mathbf{m}_j} = (n-j+1) \frac{1}{\mathbf{m}_j} + \frac{2}{\mathbf{b}} \sum_{i=1}^{n-j+1} \frac{(\ln d_{i,j} - \mathbf{m}_j)}{\ln d_{i,j}}$$

$$\Rightarrow \mathbf{m}_j^2 \sum_i \frac{1}{\ln d_{i,j}} - \mathbf{m}_j(n-j+1) = \frac{(n-j+1)}{\mathbf{b}}$$

$$\frac{d(\text{Log}L)}{d\mathbf{b}} = \sum_j \sum_i \frac{1}{2\mathbf{b}} - \frac{1}{2} \sum_{j=2}^n \sum_{i=1}^{n-j+1} \frac{(\ln d_{i,j} - \mathbf{m}_j)^2}{\ln d_{i,j}}$$

$$\Rightarrow \frac{1}{\mathbf{b}} = \frac{\sum_j \sum_i \frac{(\ln d_{i,j} - \mathbf{m}_j)^2}{\ln d_{i,j}}}{0,5n(n-1)^2}$$

Q.E.D.

Diese n Gleichungen können folgendermaßen iterativ gelöst werden:

Setze als Startwert für $\mathbf{m} = (\mathbf{m}_2, \dots, \mathbf{m}_n)$: $\hat{\mathbf{m}}_{j0} = \frac{1}{(n-j+1)} \sum_{i=1}^{n-j+1} \ln d_{ij}$, $j = 2, \dots, n$.

Führe für $m = 1, \dots, N$ (mit N - Anzahl der Iterationen) folgende Iterationsschritte durch:

1. Mit (40) berechne $\hat{\mathbf{b}}_m$ als Funktion von $(\hat{\mathbf{m}}_{2,m-1}, \dots, \hat{\mathbf{m}}_{n,m-1})$.
2. Setze $\hat{\mathbf{b}}_m$ in (41) ein und berechne $\hat{\mathbf{m}}_{j,m}$ gemäß:

$$\hat{\mathbf{m}}_{j,m} = (n-j+1 + [(n-j+1)(n-j+1 + 4\hat{\mathbf{b}}_m^{-1} \sum_{i=1}^{n-j+1} (\ln d_{i,j})^{-1})]^{0,5}) (2 \sum_{i=1}^{n-j+1} (\ln d_{i,j})^{-1})^{-1}.$$

Den endgültigen Endschaten für das Gesamtdreieck erhält man durch:

$$\hat{m}^{LIG} = \sum_{i=1}^n I_i \cdot \exp[\hat{m}\hat{b}(1-(1-\frac{2}{\hat{b}})^{0,5})] \text{ wobei } \hat{m} = \sum_{j=2}^n \hat{m}_j.$$

Da dieser Schätzer verzerrt ist, empfiehlt sich das Stichprobenmittel aus der Simulation. Sind $S_1^*, S_2^*, \dots, S_{10.000}^*$ mit der Simulation (s. unten) erzeugte Beobachtungen, so folgt für den Gesamtschaten:

$$\hat{m}^{LIG} = \frac{1}{10.000} \sum_{k=1}^{10.000} S_k^* .$$

Simulationsalgorithmus im Log-IG Fall:

1. Generiere 2 rechteckverteilte Variablen: $U_1, U_2 \sim R(0,1)$
2. Mit Box-Muller Transformation erhalte 2 unabhängige $C_{(1)}^2$ -verteilte Variablen:

$$Z_1 = [\sqrt{-2\ln(U_1)} \cos(2\pi U_2)]^2 \text{ und } Z_2 = [\sqrt{-2\ln(U_1)} \sin(2\pi U_2)]^2$$

3. Für jedes Z_i berechne $X_i = \hat{m} + \frac{Z_i - (Z_i^2 + 4\hat{b}\hat{m}Z_i)^{0,5}}{2\hat{b}}$.

4. Für jedes X_i simuliere $V_i \sim R(0,1)$.

5. Ist $V_i \leq \frac{\hat{m}}{\hat{m} + X_i}$, dann setze:

$$Y_i = \exp(\mathbf{d} + X_i), \text{ sonst } Y_i = \exp(\mathbf{d} + \frac{\hat{m}^2}{X_i}).$$

Ein so konstruiertes Y_i ist dann Log-IG verteilt mit Parametern \hat{m}, \hat{b} und Lageparameter \mathbf{d} .

6. Für jedes Anfalljahr $k = 1, \dots, n$ wiederhole Schritte 1-5 5.000 mal und erhalte 10.000

Log-IG verteilte Zufallsvariablen $U_{k1}^*, U_{k2}^*, \dots, U_{k,10.000}^*$, wobei

$$\mathbf{d} = \ln C_{k,1} \text{ und } \hat{m} = \sum_{j=2}^n \hat{m}_j$$

mit \hat{m}_j und \hat{b} aus der Iteration. Damit ist der Gesamtschaten:

$$S_k^* = \sum_{j=1}^n U_{j,k}^*, \quad k = 1, \dots, 10.000.$$

7. Mit diesen 10.000 Werten können Erwartungswert, Standardabweichung und Quantile des Endschatens für jedes Anfalljahr geschätzt werden.

2.3.5 Numerische Berechnung

Die oben beschriebenen Algorithmen werden wie immer zuerst auf das Dreieck angefallener und anschließend auf das Dreieck bezahlter Schäden angewendet.

Ergebnisse:

Aus dem Dreieck der kumulierten Schäden erhält man das Dreieck der Faktoren.

Für angefallene Schäden:

	2	3	4	5	6	7
1	1,2531	1,1445	1,1171	1,0494	1,0207	1,0222
2	1,3891	1,1563	1,0533	1,0660	1,0454	
3	1,2036	1,0635	1,0965	1,0730		
4	1,2479	1,1442	1,1257			
5	1,2488	1,0968				
6	1,2246					

Und entsprechend für bezahlte Schäden:

	2	3	4	5	6	7
1	1,5195	1,2615	1,2419	1,1760	1,1456	1,0625
2	1,5895	1,2148	1,1794	1,1861	1,1169	
3	1,5363	1,0984	1,1260	1,1341		
4	1,4359	1,1644	1,1734			
5	1,6134	1,0913				
6	1,4532					

Mit den oben beschriebenen Algorithmen erhält man folgende Schätzer für die Parameter.

Für angefallene Schäden:

Abw. Periode	LogNormal		LogGamma		Log IG	
	mu	sigma	alpha	lamda	mu	beta
1 - 2	0,23098	0,00211	0,32957	1,55562	0,23126	127,81209
2 - 3	0,11378	0,00101	0,20438	1,55562	0,10980	127,81209
3 - 4	0,09329	0,00066	0,16889	1,55562	0,09124	127,81209
4 - 5	0,06087	0,00009	0,11105	1,55562	0,06632	127,81209
5 - 6	0,03241	0,00014	0,05935	1,55562	0,03437	127,81209
6 - 7	0,02199	0,00014	0,04029	1,55562	0,02811	127,81209

Für bezahlte Schäden:

Abw. Periode	LogNormal		LogGamma		Log IG	
	mu	sigma	alpha	lamda	mu	beta
1 - 2	0,42085	0,00182	1,95451	3,53846	0,42483	117,24370
2 - 3	0,15205	0,00317	0,32959	3,53846	0,13920	117,24370
3 - 4	0,16505	0,00121	0,32959	3,53846	0,16585	117,24370
4 - 5	0,15288	0,00038	0,32959	3,53846	0,15829	117,24370
5 - 6	0,12324	0,00016	0,32959	3,53846	0,12995	117,24370
6 - 7	0,06060	0,00016	0,32959	3,53846	0,06818	117,24370

Da für das letzte Abwicklungsjahr nur eine Beobachtung vorliegt, ist es nicht möglich die Varianz zu berechnen, dafür wird die Varianz aus dem vorletztem Jahr übernommen.

Die Parameter für Log-Gamma- und Log-IG-Modelle werden durch Iterations-Algorithmen erzeugt, die mit Matlab programmiert wurden.

Beim Log-Gamma-Modell könnte die Matrix **J**, mit zu großen Werten auf der Diagonalen und zu kleinen außerhalb, beim Invertieren Probleme bereiten.

Mit den geschätzten Parametern kann man den erwarteten Endschaden für jedes Anfalljahr berechnen. Zum Vergleich werden noch mal die Chain-Ladder-Ergebnisse aufgelistet.

Die Ergebnisse für angefallene Schäden:

AnfallJahr	Chain-Ladder	LogNormal	LogGamma	Log IG
1	204.064.977	202.626.499	297.943.222	204.321.992
2	170.999.555	154.568.703	227.278.750	155.862.069
3	192.236.878	210.817.772	309.987.719	212.581.806
4	263.474.571	254.857.509	374.744.013	256.990.049
5	251.205.651	258.857.095	380.625.028	261.023.102
6	362.252.466	372.826.164	548.205.832	375.945.816
7	515.023.714	518.703.769	762.705.140	523.044.064
gesamt:	1.959.257.812	1.973.257.511	2.901.489.704	1.989.768.899

Die Ergebnisse mit dem Log-Gamma-Modell sind weniger zuverlässig als mit anderen Verteilungen, da die Matrix **J** mit angefallenen Schäden fast singular ist¹, was die Ergebnisse sichtlich beeinträchtigt. Die Ergebnisse mit stochastischen Faktoren sind meist höher als beim

¹ D. h. die Determinante von **J** ist nahe bei null.

Chain-Ladder Verfahren. Nur die Anfalljahre 1, 2, 4 bei Log-Normal und 2 bei Log-IG weisen niedrigere Ergebnisse aus.

Die entsprechenden Ergebnisse mit bezahlten Schäden:

AnfallJahr	Chain-Ladder	LogNormal	LogGamma	Log IG
1	179.534.285	154.752.878	174.314.261	156.857.971
2	139.155.430	127.509.516	143.627.230	129.244.019
3	174.926.084	198.101.772	223.142.629	200.796.537
4	247.130.408	264.163.906	297.555.282	267.757.310
5	237.766.063	240.423.248	270.813.709	243.693.709
6	367.158.792	389.727.583	438.990.709	395.029.020
7	470.728.937	480.876.165	541.660.837	487.417.490
gesamt:	1.816.400.000	1.855.555.068	2.090.104.657	1.880.796.055

Hier ist beim Log-Gamma-Modell die Matrix **J** „gut“ invertierbar, was auch bessere Ergebnisse liefert.

Ansonsten kann man beobachten, dass die Ergebnisse mit stochastischen Faktoren meist höher sind als die vom deterministischen Chain-Ladder-Verfahren. Nur bei den ersten Anfalljahren fallen die Ergebnisse niedriger aus.

Die geschätzten Parameter und die Kenntnis der Verteilung des Endschadens ermöglichen die Simulation des Endschadens, der Standardabweichung und der Quantile. Mit Hilfe von Matlab wurden 10.000 Simulationen für jedes Verteilungsmodell durchgeführt.

Die Ergebnisse mit angefallenen Schäden:

Anf. Jahr	LogNormal		LogGamma		Log IG	
	Endschaden	St. Abweichung	Endschaden	St. Abweichung	Endschaden	St. Abweichung
1	202.596.229	13.199.322	298.187.368	800.320.879	204.330.175	13.917.580
2	154.665.390	9.991.101	216.843.225	471.972.085	155.906.015	10.549.010
3	211.042.102	13.770.469	325.191.675	1.940.094.529	212.581.490	14.442.519
4	254.999.319	16.631.825	399.038.336	2.657.004.529	257.074.205	17.321.405
5	258.744.538	16.742.601	377.784.966	825.957.849	261.090.150	17.647.976
6	373.096.631	24.294.822	518.199.267	1.133.887.883	375.607.081	25.043.412
7	518.951.204	33.767.872	757.305.963	1.667.051.473	523.506.764	35.751.487
ges:	1.974.095.413	128.398.012	2.892.550.800	9.496.289.227	1.990.095.879	134.673.387

Mit bezahlten Schäden:

Anf. Jahr	LogNormal		LogGamma		Log IG	
	Endschaden	St. Abweichung	Endschaden	St. Abweichung	Endschaden	St. Abweichung
1	154.691.734	12.931.789	174.563.103	144.531.778	157.161.181	15.554.172
2	127.642.985	10.806.609	143.377.682	117.290.268	129.378.285	12.735.255
3	198.174.457	16.599.668	225.543.479	189.491.572	200.947.866	19.780.489
4	264.504.817	21.888.362	298.501.932	274.298.868	267.834.592	26.513.057
5	241.104.926	20.104.082	267.497.313	216.198.785	244.036.731	24.013.671
6	389.817.981	32.694.542	442.494.568	386.718.545	394.600.803	38.417.059
7	481.598.750	40.383.778	540.419.448	494.335.395	487.490.849	48.180.157
ges:	1.857.535.649	155.408.830	2.092.397.525	1.822.865.211	1.881.450.308	185.193.861

Die 90%-, 95%-, 98%- und 99%-Quantile für Endschäden aus den Simulationen ermöglichen eine risikogerechte Einschätzung der Reserve.

Die Ergebnisse mit angefallenen Schäden (hier nur 90 und 99%):

Anf. Jahr	Log-Normal		Log-Gamma		Log-IG	
	90%	99%	90%	99%	90%	99%
1	219.365.887	234.435.078	480.276.017	2.024.522.713	222.456.639	241.066.199
2	167.461.030	178.864.912	346.308.002	1.350.410.499	169.623.291	184.182.263
3	228.589.428	244.297.689	466.550.745	1.986.395.970	231.356.188	251.773.155
4	275.974.131	296.786.120	590.690.391	2.404.272.280	279.953.106	304.510.699
5	280.546.799	299.080.529	614.395.710	2.720.681.712	284.051.409	309.242.480
6	404.199.270	432.159.772	837.124.150	3.288.481.255	408.059.985	444.569.508
7	562.248.028	600.202.932	1.201.802.971	5.580.784.072	570.558.371	620.129.522
ges:	2.138.384.573	2.285.827.032	4.537.147.986	19.355.548.502	2.166.058.989	2.355.473.826

Die Ergebnisse mit bezahlten Schäden:

Anf. Jahr	Log-Normal		Log-Gamma		Log-IG	
	90%	99%	90%	99%	90%	99%
1	171.648.135	186.658.732	300.506.667	754.436.039	177.872.165	199.722.370
2	141.463.047	154.769.893	249.872.530	592.240.060	146.068.282	165.557.883
3	219.988.079	239.065.641	395.772.259	999.949.535	226.852.139	256.164.874
4	292.914.091	319.602.398	514.169.798	1.320.651.582	302.606.029	342.598.231
5	267.195.300	290.489.883	458.494.036	1.181.344.223	275.600.997	309.249.097
6	431.330.172	472.421.139	768.359.396	1.932.867.208	444.241.217	500.173.573
7	534.011.998	581.540.169	923.467.809	2.342.514.226	550.094.244	618.079.371
ges:	2.058.550.822	2.244.547.856	3.610.642.495	9.124.002.874	2.123.335.073	2.391.545.399

2.3.6 Kolmogorov-Test

Um die Entscheidung zu treffen, welches Verteilungsmodell zu den gegebenen Daten besser passt, sollte man zuerst den Kolmogorov-Test durchführen, der die Verteilungsannahme überprüft.

Es wird die Nullhypothese $H_0 : F(x) = F_o(x)$ für alle x (dass die Verteilungsfunktion einer Zufallsvariablen X gleich einer bestimmten vorgegebenen stetigen Verteilungsfunktion F_o ist) gegen die Alternativhypothese $H_1 : F(x) \neq F_o(x)$ für mindestens ein x geprüft.

Als Testvariable wird

$$D_n = \sup |F_n(x) - F_o(x)|$$

verwendet. Dabei bezeichnet F_n die Verteilungsfunktion der Stichprobe (X_1, \dots, X_n) vom Umfang n . Unter der Nullhypothese hängt die Verteilungsfunktion von D_n nur von n ab.

Für $n \rightarrow \infty$ konvergiert die Verteilungsfunktion von $D_n \sqrt{n}$ gegen die Kolmogorov-Verteilung. Ist $D_n > d_{n;1-a}$, so wird H_0 abgelehnt. Dabei sind $d_{n;1-a}$ das Quantil der Ordnung $1-a$ der Verteilung von D_n und a das Signifikanzniveau.¹ Im Gegensatz zum Chi-Quadrat-Anpassungs-Test, der eine Klassenteilung mit Mindestzahlen der Besetzung voraussetzt, ist der Kolmogorov-Test auch für kleine Stichproben anwendbar, was bei diesem Problem meistens der Fall ist.

¹ Die Quantile $d_{n;1-a}$ wurden dem *Lexikon Statistik*, Seite 186, entnommen.

Es ist sinnvoll, diesen Test nur für die Spalten 1 und 2 durchzuführen, die für unser Dreieck 6 bzw. 5 Elemente enthalten. Die empirische Verteilungsfunktion entsteht als Treppenfunktion durch Summation der relativen Häufigkeit der Einzelwerte von $\frac{1}{6}$ bzw. $\frac{1}{5}$. Die Werte von

$F_0(x)$ erhält man durch die in Excel definierten statistischen Funktionen. So gilt für die Log-Normalverteilung: $F_0(x) = \text{Lognormvert}(x; \mathbf{m}, \mathbf{s}^2)$. Die Werte von $F_0(x)$ für die Log-Gamma-Verteilung erhält man durch die Transformation der Gammaverteilung:

$$F(x) = P(X \leq x) = P(e^Y \leq x) = P(Y \leq \ln(x)),$$

wobei hier $X \sim \text{LogGamma}$ und $Y \sim \text{Gamma}(\mathbf{a}, \mathbf{I})$ verteilt sind. Also erhält man die Log-Gamma-Verteilungsfunktion mit Excel als $F_0(x) = \text{Gamma}(\ln(x); \mathbf{a}; \mathbf{I})$.

Die Werte der Log-IG-Verteilung wurden mit Hilfe von ExpertFit durch die Option „Distribution Viewer“ berechnet.

Man erhält somit folgende Teststatistiken für die Spalte 1 und 2 bei den angefallenen Schäden:

Abw.-Periode	n	Log-Normal	Log-Gamma	Log-IG	95%-Quantil
1-2	6	0,37952	0,44784	0,74241	0,51926
2-3	5	0,14969	0,36059	0,73596	0,56328

und entsprechend bei bezahlten Schäden:

Abw.-Periode	n	Log-Normal	Log-Gamma	Log-IG	95%-Quantil
1-2	6	0,19858	0,94398	0,79258	0,51926
2-3	5	0,24965	0,57897	0,68258	0,56328

Im Falle der angefallenen Schäden sind die Teststatistiken für Log-Normal und Log-Gamma immer kleiner als die Tafelwerte 0,51926 und entsprechend für die zweite Spalte 0,56328.

Damit können bei einer Irrtumswahrscheinlichkeit von 5% die Behauptungen, dass die Faktoren einer Log-Normal- und Log-Gammaverteilung folgen, nicht abgelehnt werden.

Im Falle der bezahlten Schäden sind die Teststatistiken im Log-Normal-Modell für beide Spalten kleiner als die Tafelwerte und im Log-Gamma- und Log-IG-Modell sind die Teststatistiken für beiden Spalte größer als die Tafelwerte. Damit wird bei einer Irrtumswahrscheinlichkeit von 5% die Behauptung, dass die Faktoren einer Log-Normalverteilung folgen, angenommen, während die Annahmen der Log-Gamma- und Log-IG-Verteilung verworfen werden.

Was sollte man tun, wenn der Kolmogorov-Test beide Verteilungen für richtig hält?

Um eine optimale Lösung zu finden, könnte man für jede Spalte die Summe $S = \sum_x (F_n(x) - F_o(x))^2$ der quadratischen Abweichungen zwischen theoretischen und empirischen Werten der Verteilungsfunktion berechnen, und sich für die Verteilung mit minimalem S entscheiden.

Die Tabelle unten listet folgende Werte S für die 1. und 2. Spalte, im Falle der angefallenen Schäden:

Abw.-Period	Log-Normal	Log-Gamma	Log-IG
1-2	0,21533	0,47870	2,42259
2-3	0,07546	0,28045	1,74201

Die Werte für die Log-Normalverteilung sind deutlich niedriger, was die Entscheidung für das Log-Normal-Modell begünstigt.

Im Falle der bezahlten Schäden wird auch die Richtigkeit der Log-Normalverteilungsannahme nochmal bestätigt. Es ergeben sich in diesem Falle folgende Werte für S :

Abw.-Period	Log-Normal	Log-Gamma	Log-IG
1-2	0,05588	2,41095	2,42259
2-3	0,07833	0,67404	2,05207

2.4 Die Kleinste-Quadrate-Regression

Es wird ein lineares Regressionsmodell¹ betrachtet, bei dem zusätzlich zu den Abwicklungsparametern die Anfalljahrparameter eingeführt werden.

Zuerst wird die Regressionsgleichung aufgestellt, dann die KQ-Methode erklärt und mögliche vorhandene Hilfsmittel zum Testen der Modellannahmen erläutert.

Anschließend wird die Methode auf beide Dreiecke der Zahlungszuwächse (bezahlt und angefallen) angewandt. Mit geschätzten Parametern werden die Prognosen (Werte im unteren Dreieck) erstellt und Reserven für jedes Anfalljahr gebildet.

Zum Schluss wird mit bekannten statistischen Tests – t-Test, F-Test, Bestimmtheitsmaß, Residuenanalyse – die Richtigkeit des Modells überprüft.

2.4.1 Modellierung für die KQ – Methode und Modellannahmen

Für das Modell wird angenommen, dass jedes Anfalljahr sein eigenes Anfalljahr-Niveau besitzt und die Aufteilung des Endschadens auf die Abwicklungsjahre im Schnitt für alle Anfalljahre gleich ist.

Mit den Bezeichnungen

U_i – Endschaden für das Anfalljahr i ,

B_j – Anteil des Abwicklungsjahres j am Endschaden,

$P_{i,j}$ – Schadenzuwächse (angefallene oder bezahlte) aus dem Anfalljahr i im Abwicklungsjahr j

kann man das Modell durch folgende Gleichung darstellen:

$$P_{i,j} = U_i \cdot B_j \quad i, j = 1, \dots, n$$

unter der Bedingung: $\sum_{j=1}^n B_j = 1$, wobei n für die Anzahl der Anfall- und Abwicklungsjahre

steht.

Durch Logarithmieren erhält man ein lineares Modell, zu dem ein Störterm $u_{i,j}$ zusätzlich eingeführt wird:

$$\ln(P_{i,j}) = \ln(U_i) + \ln(B_j) + u_{i,j} \quad i, j = 1..n$$

$$\ln(P_{i,j}) = Y_{i,j} = a_i + b_j + u_{i,j} \quad i, j = 1..n$$

¹ Dieses Modell wurde aus dem Artikel „Regression Modells based on Log-Incremental Payments“, S. Christofides in Claims Reserving Manual, Vol. 2, More Advanced Methods, 1990, S. 5.1 – 5.54, entnommen.

(b_1 wird auf Null gesetzt, um Singularitätsprobleme zu umgehen).

oder in Matrixschreibweise:

$$\mathbf{y} = \mathbf{X} \hat{\mathbf{a}} \quad (42)$$

mit $\hat{\mathbf{a}}' = (a_1, \dots, a_n, b_2, \dots, b_n)$.

Der Vektor \mathbf{y} enthält somit die logarithmierten Beobachtungen:

$$\mathbf{y}' = (Y_{1,1}, \dots, Y_{1,n}, Y_{2,1}, \dots, Y_{2,n-1}, \dots, Y_{n,1}).$$

Die $\frac{n(n+1)}{2} \times (2n-1)$ -Matrix \mathbf{X} wird Designmatrix genannt, deren genaue Struktur später erläutert wird und für die gilt:

\mathbf{X} ist nicht stochastisch mit $\text{Rang}(\mathbf{X}) = 2n-1 =: p$ (= Anzahl der zu schätzenden Parameter).

Die Restgröße \mathbf{u} stellt einen nicht beobachtbaren stochastischen Störterm dar, für den folgende Annahmen getroffen werden:

1. $E[\mathbf{u}] = \mathbf{0}$
2. $E[\mathbf{u}\mathbf{u}'] = \sigma^2 \mathbf{I}$ mit $\text{Var}(u_{i,j}) = \sigma^2$ für alle i, j
3. \mathbf{u} unterliegt einer mehrdimensionalen Normalverteilung

$E[\mathbf{u}\mathbf{u}']$ ist die Varianz-Kovarianz-Matrix der Störvariablen, die unter der Annahme 2 eine Diagonalmatrix ist. Annahme 2 besagt damit, dass die Störterme zweier unterschiedlicher Beobachtungen unkorreliert sind und die Varianz der Störterme konstant gleich σ^2 ist. Man bezeichnet Eigenschaft eins als fehlende Autokorrelation und Eigenschaft zwei als Homoskedastizität.

2.4.2 Schätzen im linearen Regressionsmodell

Die Methode der kleinsten Quadrate verlangt, dass die Parameter im Vektor $\hat{\mathbf{a}}$ so festzulegen sind, dass die Residuenquadratsumme $RSS = \mathbf{u}'\mathbf{u}$ minimiert wird. Durch Einsetzen von (42) erhält man

$$RSS = (\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\mathbf{a}} - \hat{\mathbf{a}}'\mathbf{X}'\mathbf{y} + \hat{\mathbf{a}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{a}}$$

und nach dem Differenzieren der Quadratsumme nach $\hat{\mathbf{a}}$

$$\frac{d(RSS)}{d\hat{\mathbf{a}}} = -2\mathbf{X}'\mathbf{y} + 2(\mathbf{X}'\mathbf{X})\hat{\mathbf{a}}.$$

Notwendige Bedingung für ein Minimum ist, dass diese Gleichung gleich 0 ist. Damit ergibt sich für die Lösung:

$$\hat{\mathbf{a}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (43)$$

Die Inverse der symmetrischen $p \times p$ -Matrix $\mathbf{X}'\mathbf{X}$ existiert nach der Annahme $\text{Rang}(\mathbf{X})=p$.

2.4.3 Schätzeigenschaften der KQ-Methode

Die Eignung der KQ-Methode zur Schätzung von Parametern für das vorliegende Modell hängt von den Eigenschaften ihrer Schätzfunktionen ab. Ihre Anwendung ist daher durch die Qualität der Schätzung zu begründen. Bei der Beurteilung der Güte der Schätzung unterscheidet man Schätzeigenschaften, die vom Stichprobenumfang unabhängig sind und solche, die ausschließlich bei großem Stichprobenumfang gelten. Insbesondere sind die Eigenschaften im ersten Fall vom Interesse, die auch bei kleinem Stichprobenumfang gelten.

Das sind die Erwartungstreue und Effizienz des KQ-Schätzers $\hat{\mathbf{a}}$.

Erwartungstreue besagt, dass der Erwartungswert des Schätzers gleich dem wahren Parameter ist. Diese Eigenschaft kann leicht gezeigt werden, indem die Schätzung in die Regressionsgleichung (42) eingesetzt wird:

$$E(\hat{\mathbf{a}}) = E\left[\hat{\mathbf{a}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right] = \hat{\mathbf{a}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}] = \hat{\mathbf{a}}$$

Damit ist jedoch noch nichts über die Genauigkeit der Schätzung ausgesagt. Je kleiner die Varianzen der geschätzten Regressionskoeffizienten sind, um so präziser kann $\hat{\mathbf{a}}$ geschätzt werden. Ein Schätzer mit kleinstmöglicher Varianz heißt effizient.

Diese beide Eigenschaften sind Aussage des Gauß-Markov-Theorems.

Außerdem gilt unter der Annahme 3 für Normalität der Residuen, dass der KQ-Schätzer (43) mit dem Maximum-Likelihood-Schätzer übereinstimmt¹.

Die Varianz - Kovarianzmatrix von $\hat{\mathbf{a}}$ lautet:

$$\text{Cov}(\hat{\mathbf{a}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Beweis:

$$\begin{aligned} \text{Cov}(\hat{\mathbf{a}}) &= E[(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})'] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}'] = \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}\mathbf{u}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \end{aligned}$$

¹ Zum Beweis siehe in H. Eckey, R. Kosfeld, C. Dreger (1995), a. a. O. : S. 57 f

$$s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

Q.E.D.

Diese ist eine symmetrische $p \times p$ -Matrix mit $Var(\hat{\mathbf{b}}_j)$ in der Hauptdiagonale und $Cov(\hat{\mathbf{b}}_i, \hat{\mathbf{b}}_j)$ zwischen den Schätzern $\hat{\mathbf{b}}_i$ und $\hat{\mathbf{b}}_j$ an den anderen Positionen. Diese Varianzen und Kovarianzen werden bei der Berechnung der Reservegenauigkeit benötigt.

2.4.4 Angepasste Werte, Residuen und Standardabweichung

Die geschätzte Regressionsgerade

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\mathbf{b}} \quad (44)$$

kann dann zur Schätzung der zukünftigen logarithmischen Zahlungszuwächse verwendet werden, indem man statt der Designmatrix \mathbf{X} die $\frac{n(n-1)}{2} \times p$ -Future-Designmatrix \mathbf{X}_f , die genau wie Matrix \mathbf{X} aufgebaut ist, in die Gleichung (44) einsetzt.

Die Residuen $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ geben die Differenz zwischen den beobachteten und den durch die Regression angepassten Werten an. Je kleiner die Residuen, desto besser die Anpassung. Mit

Hilfe der Residuen wird der Modellparameter σ^2 durch $s^2 = \frac{\mathbf{e}'\mathbf{e}}{N-p}$ geschätzt, wobei

$N = \frac{n(n+1)}{2}$ die Anzahl der vorliegenden Beobachtungen und p die Anzahl der Parameter darstellt.

$s = \sqrt{s^2}$ ist ein Maß für den Fehler, den man bei der Schätzung oder Voraussage von \mathbf{Y} aus vorgegebenen Werten \mathbf{X} macht. Diese Standardabweichung wird auch als Standardfehler der Schätzung oder als Standardfehler der Voraussage bezeichnet.

2.4.5 Das Bestimmtheitsmaß

Für die Güte der Anpassung des Regressionsmodells wird das Bestimmtheitsmaß R^2 verwendet, das definiert ist durch¹:

$$R^2 = \frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$$

¹ $R^2 = \frac{\text{durch die Regression erklärte Varianz}}{\text{gesamte Varianz}}$

Es gilt $0 \leq R^2 \leq 1$. Je näher R^2 an 1 liegt, desto besser wird die Abhängigkeit zwischen \mathbf{X} und \mathbf{y} durch das Modell erklärt.

2.4.6 Signifikanztest für die Regressionskoeffizienten

Obwohl das Bestimmtheitsmaß oft eine dominierende Rolle bei der Modellbewertung einnimmt, sollten zusätzlich Signifikanztests der Regressionskoeffizienten durchgeführt werden. Sind die Parameter geschätzt, so kann die Nullhypothese: $H_0 : \mathbf{b}_i = 0$ für alle Parameter \mathbf{b}_i $i = 1, \dots, p$ geprüft werden.

Die Prüfgröße des Signifikanztests

$$T = \frac{\hat{\mathbf{b}}_i}{\sqrt{\text{Var}(\hat{\mathbf{b}}_i)}} \quad (45)$$

ist t -verteilt mit $n-p$ Freiheitsgraden. Bei einem zweiseitigen Test wird die Nullhypothese bei einem Signifikanzniveau \mathbf{a} verworfen, wenn der absolute Wert der Prüfgröße T das $(1-\mathbf{a}/2)$ -Quantil $t_{1-\mathbf{a}/2; N-p}$ der t -Verteilung mit $N-p$ Freiheitsgraden übersteigt:

$$|T| > t_{1-\mathbf{a}/2; N-p} \Rightarrow H_0 \text{ ablehnen.}$$

In diesem Fall ist der Parameter \mathbf{b}_i signifikant von Null verschieden.

2.4.7 Signifikanz des Gesamtzusammenhangs

Die Nullhypothese für einen Test des Gesamtzusammenhangs ist durch

$$H_0 : \mathbf{b}_1 = \mathbf{b}_2 = \dots = \mathbf{b}_p = 0$$

gegeben. Unter der H_0 ist die Prüfgröße

$$F = \frac{R^2 / (p-1)}{(1-R^2) / (N-p)}$$

F -verteilt mit $p-1$ und $N-p$ Freiheitsgraden. Die Nullhypothese wird auf dem Signifikanzniveau \mathbf{a} verworfen, wenn die Prüfgröße F das $(1-\mathbf{a})$ -Quantil $F_{1-\mathbf{a}; p-1; N-p}$ der

F -Verteilung mit $p-1$ und $N-p$ Freiheitsgraden übersteigt:

$$F > F_{1-\mathbf{a}; p-1; N-p} \Rightarrow H_0 \text{ ablehnen.}$$

In diesem Fall ist mindestens ein Koeffizient bei einer Irrtumswahrscheinlichkeit \mathbf{a} für das Modell signifikant.

2.4.8 Numerische Berechnung

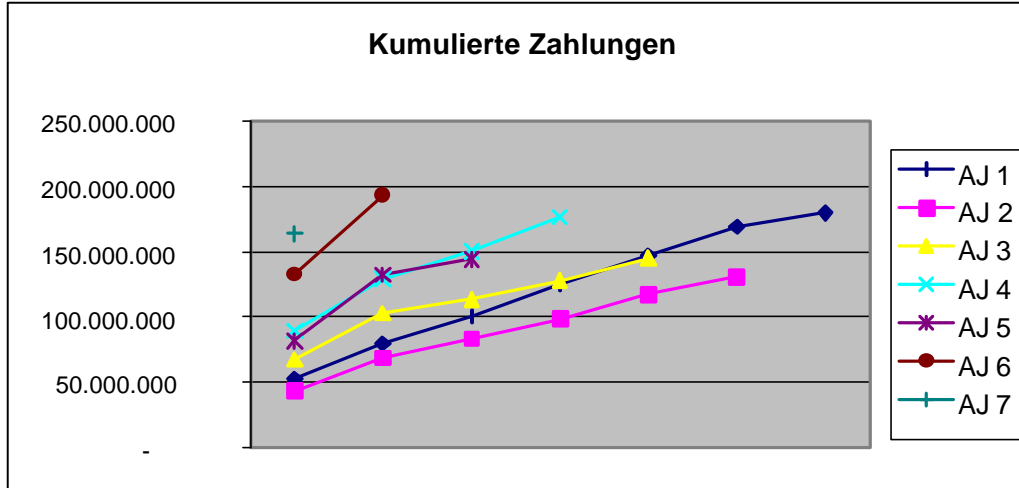
Wir wenden dieses Modell nacheinander auf beide Dreiecke (mit bezahlten und angefallenen Schadenzuwächsen) an und vergleichen anschließend die Ergebnisse.

2.4.8.1 Berechnung mit dem Abwicklungsdreieck der bezahlten Schadenzuwächse

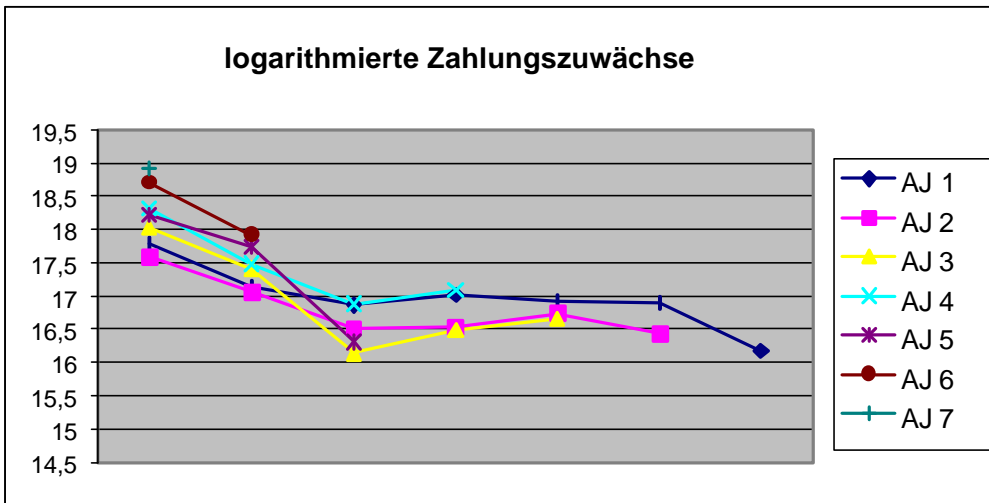
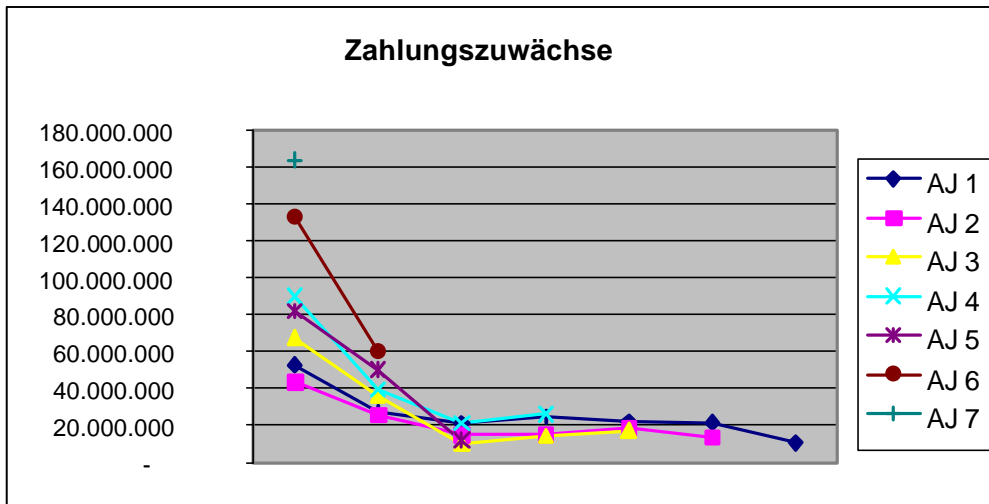
Gegeben ist das Dreieck der bezahlten Schadenzuwächse:

52.686.881	27.373.463	20.936.631	24.426.289	22.080.274	21.473.601	10.557.146
43.411.656	25.592.252	14.821.015	15.040.814	18.396.315	13.710.628	
67.445.366	36.170.840	10.192.846	14.334.935	17.183.602		
89.936.759	39.199.208	21.232.284	26.078.508			
81.854.058	50.212.911	21.232.284				
132.685.938	60.134.134					
163.718.217						

Die Zahlungsabwicklung kann man sich mit Hilfe von Plots veranschaulichen¹.



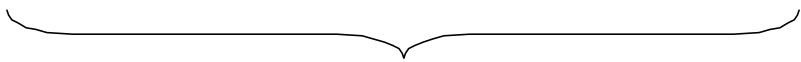
¹ Die Abkürzung AJ in den Abbildungen steht für das Anfalljahr.



Die obigen Graphiken legen ein exponentiell fallendes Modell nahe.

Die Regression wurde in Matlab und Excel programmiert (s. Anhang), dafür wurden die Daten folgendermaßen aufbereitet:

i	j	P_{ij}	Y_{ij}	a_1	a_2	a_3	a_4	a_5	a_6	a_7	b_2	b_3	b_4	b_5	b_6	b_7
1	1	52.686.881	17,7799	1	0	0	0	0	0	0	0	0	0	0	0	0
1	2	27.373.463	17,1251	1	0	0	0	0	0	0	1	0	0	0	0	0
1	3	20.936.631	16,8570	1	0	0	0	0	0	0	0	1	0	0	0	0
1	4	24.426.289	17,0112	1	0	0	0	0	0	0	0	0	1	0	0	0
1	5	22.080.274	16,9102	1	0	0	0	0	0	0	0	0	0	1	0	0
1	6	21.473.601	16,8823	1	0	0	0	0	0	0	0	0	0	0	1	0
1	7	10.557.146	16,1723	1	0	0	0	0	0	0	0	0	0	0	0	1
2	1	43.411.656	17,5862	0	1	0	0	0	0	0	0	0	0	0	0	0
2	2	25.592.252	17,0578	0	1	0	0	0	0	0	1	0	0	0	0	0
2	3	14.821.015	16,5116	0	1	0	0	0	0	0	0	1	0	0	0	0
2	4	15.040.814	16,5263	0	1	0	0	0	0	0	0	0	1	0	0	0
2	5	18.396.315	16,7277	0	1	0	0	0	0	0	0	0	0	1	0	0
2	6	13.710.628	16,4337	0	1	0	0	0	0	0	0	0	0	0	1	0
3	1	67.445.366	18,0268	0	0	1	0	0	0	0	0	0	0	0	0	0
3	2	36.170.840	17,4038	0	0	1	0	0	0	0	1	0	0	0	0	0
3	3	10.192.846	16,1372	0	0	1	0	0	0	0	0	1	0	0	0	0
3	4	14.334.935	16,4782	0	0	1	0	0	0	0	0	0	1	0	0	0
3	5	17.183.602	16,6595	0	0	1	0	0	0	0	0	0	0	1	0	0
4	1	89.936.759	18,3146	0	0	0	1	0	0	0	0	0	0	0	0	0
4	2	39.199.208	17,4842	0	0	0	1	0	0	0	1	0	0	0	0	0
4	3	21.232.284	16,8710	0	0	0	1	0	0	0	0	1	0	0	0	0
4	4	26.078.508	17,0766	0	0	0	1	0	0	0	0	0	1	0	0	0
5	1	81.854.058	18,2204	0	0	0	0	1	0	0	0	0	0	0	0	0
5	2	50.212.911	17,7318	0	0	0	0	1	0	0	1	0	0	0	0	0
5	3	12.053.610	16,3049	0	0	0	0	1	0	0	0	1	0	0	0	0
6	1	132.685.938	18,7035	0	0	0	0	0	1	0	0	0	0	0	0	0
6	2	60.134.134	17,9121	0	0	0	0	0	1	0	1	0	0	0	0	0
7	1	163.718.217	18,9137	0	0	0	0	0	0	1	0	0	0	0	0	0



Designmatrix X

Man erhält folgende Regressionsergebnisse:

- Standardfehler (s): 0,223732
- R^2 : 0,94058
- F : 24,6977
- Anzahl der Beobachtungen: 28
- Anzahl der Freiheitsgrade: 28-13=15

	<u>Parameter:</u>	<u>Standardfehler:</u>	<u>T-Werte:</u>
a₁:	18,0358	0,1300	138,7
a₂:	17,7488	0,1330	136,5
a₃:	17,8242	0,1330	134,0
a₄:	18,2643	0,1393	131,1
a₅:	18,1244	0,1500	120,8
a₆:	18,6341	0,1709	109,0
a₇:	18,9137	0,2238	84,5
b₂:	-0,6529	0,1292	-5,1
b₃:	-1,4630	0,1386	-10,6
b₄:	-1,1951	0,1500	-8,0
b₅:	-1,1037	0,1664	-6,6
b₆:	-1,2344	0,1944	-6,3
b₇:	-1,8635	0,2587	-7,2

Die Ergebnisse liefern also einen relativ hohen Wert für das Bestimmtheitsmaß, und die F - und T -Werte bestätigen die Signifikanz der Parameter zum Signifikanzniveau $\alpha = 5\%$, da beide Statistiken betragsmäßig deutlich größer als die entsprechenden Tabellenwerte $t(0,975;15) = 2,131$ bzw. $F(0,950;12;15) = 2,48$ sind.

Mit den geschätzten Parametern (b_2, \dots, b_7) , die die Aufteilung des Endschadens auf die Abwicklungsjahre bestimmen, kann das Zahlungsmuster abgeleitet werden:

	<u>Parameter:</u>	<u>exp(b_i):</u>	<u>Zahlung in % des Endschadens (nach KO):</u>	<u>Zahlung in % des Endschadens (nach CL):</u>
b₁:	0	1,000000	35,30%	34,78%
b₂:	-0,6529	0,520517	18,38%	17,74%
b₃:	-1,4630	0,231541	8,17%	8,10%
b₄:	-1,1951	0,302668	10,69%	10,78%
b₅:	-1,1037	0,331635	11,71%	11,68%
b₆:	-1,2344	0,291013	10,27%	11,04%
b₇:	-1,8635	<u>0,155124</u>	<u>5,48%</u>	<u>5,88%</u>
Summe:		2,832498515	100,00%	100,00%

Beide Verfahren haben annähernd gleiche Schadenabwicklungsschemata, was auch auf ähnliche Modellannahmen zurückzuführen ist.

Zahlungsprognose aufgrund der geschätzten Regression

Wir möchten jetzt die zukünftigen Zahlungen und deren Standardabweichungen mit geschätzten Parametern ermitteln.

Da die ganze Regression in Log-Skalierung erfolgte, sollten die angepassten Werte in die Original-Skalierung zurückgeführt werden. Dafür wird die Beziehung zwischen Log-Normal- und Normalverteilung verwendet, die schon in §2.3.2 beschrieben wurde.

Nach unseren Modellannahmen besitzt $Y_{i,j} = \ln(P_{i,j})$ eine Normalverteilung, also ist $P_{i,j}$ nach obiger Definition lognormalverteilt.

Für unser Dreieck sollen $\frac{n(n-1)}{2}$ Prognosewerte $P_{i,j}$ mit $i + j > n + 1$ ermittelt werden.

Man schätzt zuerst $\hat{Y}_{i,j}$, ermittelt die Varianz-Kovarianz-Matrix von $\hat{Y}_{i,j}$, und berechnet die Prognosewerte $\hat{P}_{i,j}$ und deren Standardabweichungen.

Die zukünftigen $\hat{Y}_{i,j}$ sind mit geschätzten Parametern sehr leicht zu ermitteln, z.B. gilt für Anfalljahr i und Abwicklungsjahr j :

$$\hat{Y}_{i,j} = a_i + b_j$$

oder für das gesamte Modell in Matrixschreibweise:

$$\hat{\mathbf{y}} = \mathbf{X}_f \cdot \hat{\mathbf{a}},$$

wobei \mathbf{X}_f die Future-Designmatrix bezeichnet, die analog zur Designmatrix \mathbf{X} aufgebaut ist, indem wir das Design (d.h. die Positionierung der Einsen) nicht an den beobachteten Zellen (i, k) , $i + k \leq n + 1$, sondern an den zu prognostizierenden Zellen (i, k) , $i + k > n + 1$, ausrichten.

Future Designmatrix \mathbf{X}_f :

<i>i</i>	<i>j</i>	a₁	a₂	a₃	a₄	a₅	a₆	a₇	b₂	b₃	b₄	b₅	b₆	b₇
2	7	0	1	0	0	0	0	0	0	0	0	0	0	1
3	6	0	0	1	0	0	0	0	0	0	0	0	1	0
3	7	0	0	1	0	0	0	0	0	0	0	0	0	1
4	5	0	0	0	1	0	0	0	0	0	0	1	0	0
4	6	0	0	0	1	0	0	0	0	0	0	0	1	0
4	7	0	0	0	1	0	0	0	0	0	0	0	0	1
5	4	0	0	0	0	1	0	0	0	0	1	0	0	0
5	5	0	0	0	0	1	0	0	0	0	0	1	0	0
5	6	0	0	0	0	1	0	0	0	0	0	0	1	0
5	7	0	0	0	0	1	0	0	0	0	0	0	0	1
6	3	0	0	0	0	0	1	0	0	1	0	0	0	0
6	4	0	0	0	0	0	1	0	0	0	1	0	0	0
6	5	0	0	0	0	0	1	0	0	0	0	1	0	0
6	6	0	0	0	0	0	1	0	0	0	0	0	1	0
6	7	0	0	0	0	0	1	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	1	1	0	0	0	0	0
7	3	0	0	0	0	0	0	1	0	1	0	0	0	0
7	4	0	0	0	0	0	0	1	0	0	1	0	0	0
7	5	0	0	0	0	0	0	1	0	0	0	1	0	0
7	6	0	0	0	0	0	0	1	0	0	0	0	1	0
7	7	0	0	0	0	0	0	1	0	0	0	0	0	1

Entsprechend gilt für die Varianz-Kovarianz-Matrix von $\hat{\mathbf{y}}$:

$\text{Cov}(\hat{\mathbf{y}}) = \mathbf{s}^2 \mathbf{X}_f (\mathbf{X}\mathbf{C}\mathbf{X})^{-1} \mathbf{X}_f \mathbf{C}$ mit Varianzen von Prognosewerten auf der Hauptdiagonale und Kovarianzen zwischen den Prognosenwerten auf den anderen Positionen.

Für die Prognose der Zahlung gilt dann:

$$\hat{P}_{i,j} = \exp(\hat{Y}_{i,j} + 0,5 \text{Var}(\hat{Y}_{i,j}))$$

und für die Standardabweichung:

$$\text{s.e.}(\hat{P}_{i,j}) = \sqrt{\text{Var}(\hat{P}_{i,j})} = \hat{P}_{i,j} \sqrt{\exp(\text{Var}(\hat{Y}_{i,j})) - 1}.$$

Es haben sich folgende Werte für die Prognose von $\hat{Y}_{i,j}$ in Log-Skalierung bzw. $\hat{P}_{i,j}$ in Original-Skalierung für $i + j > n + 1$ (unteres Dreieck) ergeben mit entsprechenden Werten für die Standardabweichungen:

i	j	Y_{ij}	$\text{Var}(Y_{ij})$	P_{ij}	$\text{s.e.}(P_{ij})$
2	7	15,8852	0,0667	8.191.287	2.151.281
3	6	16,5898	0,04	16.351.238	3.303.224
3	7	15,9607	0,0692	8.844.207	2.367.383
4	5	17,1606	0,0334	28.841.164	5.315.235
4	6	17,0299	0,0434	25.435.238	5.356.854
4	7	16,4008	0,0726	13.757.644	3.775.222
5	4	16,9293	0,0334	22.885.872	4.217.714
5	5	17,0207	0,0386	25.141.499	4.987.573
5	6	16,8901	0,0486	22.172.476	4.948.008
5	7	16,2609	0,0778	11.992.851	3.411.254
6	3	17,1711	0,04	29.240.506	5.907.073
6	4	17,4389	0,0434	38.287.874	8.063.717
6	5	17,5303	0,0486	42.061.518	9.386.445
6	6	17,3997	0,0586	37.094.368	9.112.767
6	7	16,7705	0,0878	20.063.940	6.078.074
7	2	18,2607	0,0667	88.107.979	23.139.836
7	3	17,4507	0,0692	39.242.071	10.504.166
7	4	17,7185	0,0726	51.384.045	14.100.247
7	5	17,8099	0,0778	56.448.445	16.056.230
7	6	17,6793	0,0878	49.782.306	15.080.814
7	7	17,0501	0,117	26.926.708	9.486.437

Schätzung der Reserve

Wir können jetzt die Reserven für jedes Anfalljahr $i = 2, \dots, n$ und deren Standardabweichungen schätzen.

Zur Berechnung des Standardfehlers des *Reserveschätzers*

$$\hat{R}_i = \hat{P}_{i,n+2-i} + \dots + \hat{P}_{i,n} \quad \text{mit } n = 7 - \text{die Dreiecksgröße}$$

brauchen wir wegen

$$\text{Var}(\hat{R}_i) = \sum_{k=n+2-i}^n \text{Var}(\hat{P}_{i,k}) + 2 \sum_{k < j} \text{Cov}(\hat{P}_{i,k}, \hat{P}_{i,j})$$

auch Schätzer für die Kovarianzen, da die benachbarten Spalten des Dreieckes miteinander korreliert sind.

Für die Kovarianz der Log-normalverteilten Zahlungsprognosen gilt

$$\text{Cov}(\hat{P}_{i,k}, \hat{P}_{i,j}) = E(\hat{P}_{i,k}) E(\hat{P}_{i,j}) (\exp(\text{Cov}(\ln(\hat{P}_{i,k}), \ln(\hat{P}_{i,j}))) - 1)$$

$$\text{Cov}(\hat{P}_{i,k}, \hat{P}_{l,j}) = 0 \text{ für } i \neq l.$$

Gemäß unserem Regressionsansatz ist

$$\ln(P_{i,j}) = Y_{i,j} = a_i + b_j + u_{i,j}$$

$$\ln(\hat{P}_{i,j}) = \hat{Y}_{i,j} \text{ und für } j \neq k \text{ gilt}$$

$$\text{Cov}(\ln(\hat{P}_{i,k}), \ln(\hat{P}_{i,j})) = \text{Cov}(\hat{Y}_{i,k}, \hat{Y}_{i,j}).$$

Letzteres wird wie schon oben erwähnt aus der Matrix $\text{Cov}(\hat{\mathbf{y}}) = \mathbf{s}^2 \mathbf{X}_f (\mathbf{X}\mathbf{X})^{-1} \mathbf{X}_f \mathbf{t}$ der künftigen $Y_{i,j}$, $i + k > n + 1$ entnommen. Diese Kovarianzen bilden alle Matrixelemente außerhalb der Hauptdiagonale von $\text{Cov}(\hat{\mathbf{y}})$.

Das alles mag wegen der vielen Kovarianzen vielleicht etwas mühsam erscheinen, ist jedoch mit Hilfe einiger Matrixoperationen recht einfach rechenbar und in Matlab, das gerade für den Matrizenkalkül sehr gut ist, leicht zu programmieren.

Folgende Tabelle zeigt die ermittelten Reserven, deren Varianzen und Standardfehler für jedes Anfalljahr und entsprechend für alle Anfalljahre zusammen.

Anf. Jahr	Reserve	Varianz	St. Fehler
2	8.191.287	4,62801E+12	2.151.281
3	25.195.444	20.886.905.676.004	4.570.219
4	68.034.046	122.267.808.383.097	11.057.477
5	82.192.698	186.854.168.780.379	13.669.461
6	166.748.207	998.071.077.326.007	31.592.263
7	311.891.554	6.241.729.044.426.100	79.004.614
Gesamt =	662.253.236	7,57444E+15	87.031.242

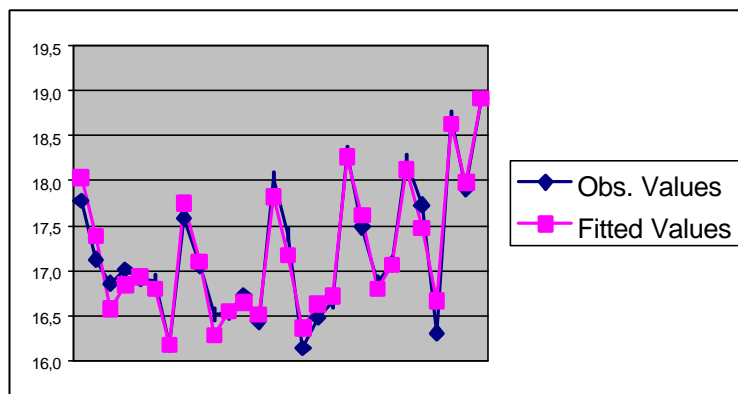
Der Standardfehler der Gesamtreserve beträgt 87.031.242, was 13% der Gesamtreserve von 662.253.236 ausmacht. Im Vergleich: Die Gesamtreserve nach Chain-Ladder betrug 683.459.819. Dies ist aber eine Punktschätzung, was sicherlich weniger aussagt als eine Schätzung mit Standardfehler.

Hilfsmittel zum Testen der Modellannahmen

In großem Maß bilden geeignete graphische Darstellungen eines Modells wichtige Hilfsmittel zur Prüfung und – daraus folgend – zur Verbesserung eines Modells.

Vor der Anwendung eines ökonometrischen Modells zum Zwecke einer Prognose sollte die Prognosequalität für den bekannten Zeitraum in Vergangenheit überprüft werden. Werden die Abweichungen zwischen den beobachteten und den durch die Regression angepassten Werten gering, so kann man auch auf eine richtige Prognose hoffen.

Folgender Plot zeigt der Verlauf den beobachteten gegen die angepassten Werte.



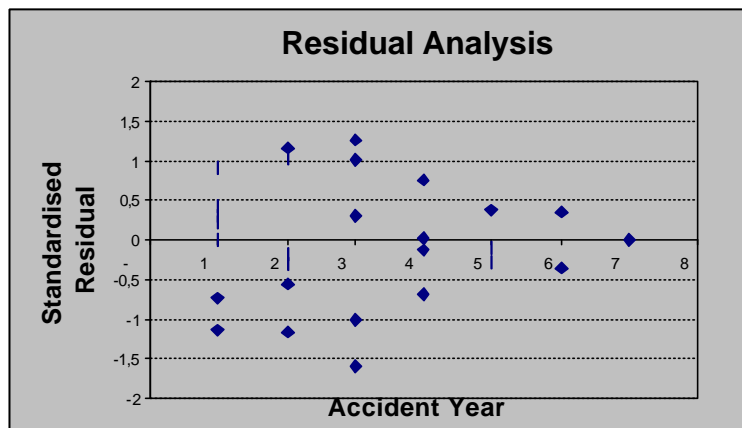
Wie man sieht, verlaufen die angepassten Werte ziemlich dicht an den beobachteten, was die Richtigkeit des Modells nahelegt.

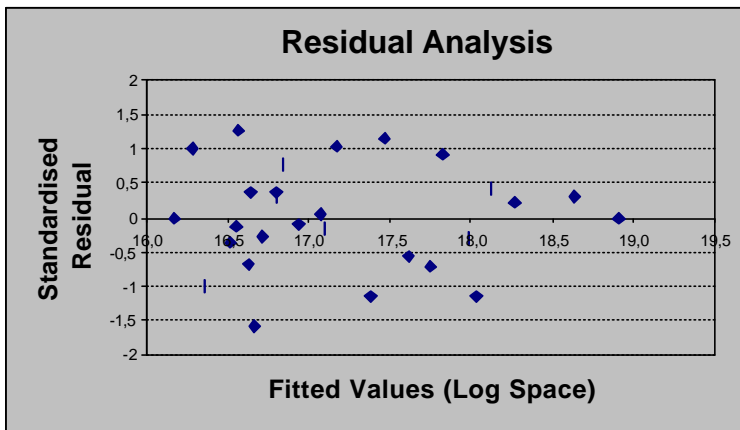
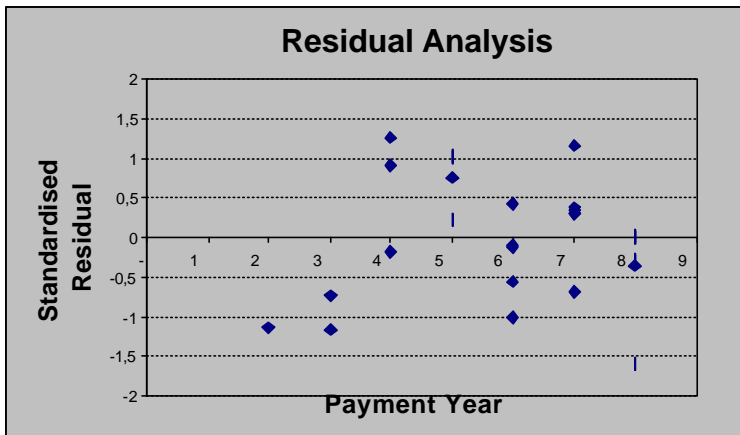
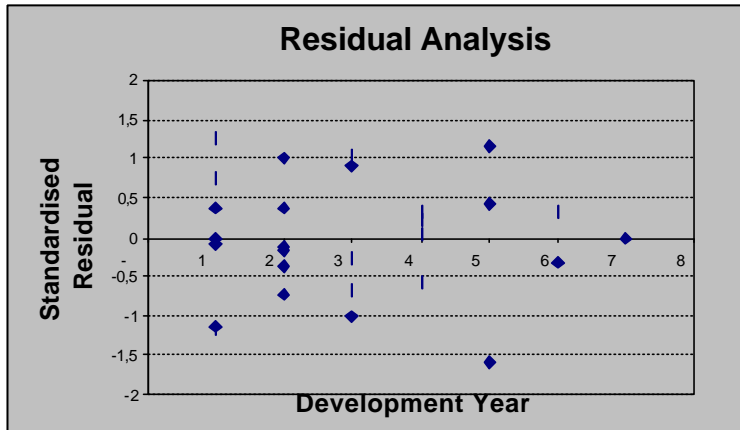
Als nächstes erstellen wir Plots der standardisierten Residuen $\frac{(Y_{i,j} - \hat{Y}_{i,j})}{s}$ sowohl in Abhängigkeit vom Anfalljahr i als auch vom Abwicklungsjahr j als auch vom Kalenderjahr $i + j$ als auch von den geschätzten Werten $\hat{y} = \mathbf{X} \hat{\mathbf{a}}$. Durch diese graphische Darstellungen können Fehlspezifikationen des Modells, wie Heteroskedastizität und Autokorrelation, erkannt werden.

Die dafür benötigten Werte werden in einer Residuentabelle zusammengefasst.

i	j	$i + j$	Y_{i,j_Obs}	Y_{i,j_Fit}	$Y_{Obs} - Y_{Fit}$	st. Residuen
1	1	2	17,7799	18,0358	-0,2560	-1,1441
1	2	3	17,1251	17,3829	-0,2578	-1,1524
1	3	4	16,8570	16,5728	0,2842	1,2701
1	4	5	17,0112	16,8407	0,1704	0,7618
1	5	6	16,9102	16,9321	-0,0219	-0,0980
1	6	7	16,8823	16,8015	0,0809	0,3615
1	7	8	16,1723	16,1723	0,0000	0,0000
2	1	3	17,5862	17,7488	-0,1625	-0,7264
2	2	4	17,0578	17,0958	-0,0380	-0,1700
2	3	5	16,5116	16,2858	0,2258	1,0092
2	4	6	16,5263	16,5536	-0,0274	-0,1223
2	5	7	16,7277	16,6450	0,0826	0,3693
2	6	8	16,4337	16,5144	-0,0807	-0,3607
3	1	4	18,0268	17,8242	0,2026	0,9057
3	2	5	17,4038	17,1713	0,2325	1,0392
3	3	6	16,1372	16,3612	-0,2240	-1,0012
3	4	7	16,4782	16,6291	-0,1509	-0,6743
3	5	8	16,6595	16,7205	-0,0610	-0,2727
4	1	5	18,3146	18,2643	0,0503	0,2248
4	2	6	17,4842	17,6114	-0,1272	-0,5687
4	3	7	16,8710	16,8013	0,0697	0,3115
4	4	8	17,0766	17,0692	0,0074	0,0331
5	1	6	18,2204	18,1244	0,0960	0,4291
5	2	7	17,7318	17,4715	0,2603	1,1633
5	3	8	16,3049	16,6615	-0,3566	-1,5938
6	1	7	18,7035	18,6341	0,0694	0,3103
6	2	8	17,9121	17,9811	-0,0690	-0,3086
7	1	8	18,9137	18,9137	0,0000	0,0000

Unter der Richtigkeit der Modellannahmen sollen die Residuenplots kein systematisches, sondern zufälliges Muster aufweisen. Folgende Abbildungen zeigen keinen systematischen Trend.





Simulation zur Prüfung des Modells

Mit Hilfe von Matlab wurden 21 neue Prognosewerte simuliert und zwar nach folgendem Algorithmus:

1. Generiere 2 rechteckverteilte Variablen: $U_1, U_2 \sim R(0,1)$
2. Mit Box-Muller Transformation erzeuge 2 unabhängige normalverteilte Variablen:
 $X_1 = \sqrt{-2\ln(U_1)} \cos(2\pi U_2)$ und $X_2 = \sqrt{-2\ln(U_1)} \sin(2\pi U_2)$
3. Mit Transformation erzeuge 2 normalverteilte Variablen mit Parametern: $\hat{\mathbf{m}}_k$ und $\hat{\mathbf{s}}_{ik}^2$:
 $Y_1 = \hat{\mathbf{m}}_k + \hat{\mathbf{s}}_{ik} X_1$ und $Y_2 = \hat{\mathbf{m}}_k + \hat{\mathbf{s}}_{ik} X_2$.
Für $\hat{\mathbf{m}}_k$ und $\hat{\mathbf{s}}_{ik}^2$ werden die aus der Regression geschätzten Werte $\hat{Y}_{i,k}$ und $\text{Var}(\hat{Y}_{i,k})$ eingesetzt.
4. Durch 10.000-maliges Wiederholen der Schritte 1-3 erhält man 20.000 normalverteilte Zufallsvariablen $Y_i, i = 1, \dots, 20.000$.
5. Berechne neuen Prognosewert: $\hat{Y}_{i,k} = \frac{\sum_{j=1}^{20.000} Y_j}{20.000}$ mit $\text{Var}(\hat{Y}_{i,k}) = \frac{1}{20.000-1} \sum Y_j^2 - \bar{Y}^2$ für $i + k > n + 1$.
6. Wiederhole Schritte 1-5 und erhalte 21 neue Prognosewerte: $\hat{Y}_{i,k}$ mit $i + k > n + 1$.

Z.B. wird $\hat{Y}_{2,7}$ mit den aus der Regression geschätzten Parametern $\hat{\mathbf{m}}_{2,7} = 15,8852$ und $\hat{\mathbf{s}}_{2,7}^2 = 0,0667$ simuliert.

Die aus der Simulation gewonnene Ergebnisse unterscheiden sich nur gering von den Regressionsschätzern:

$Y_{i,j}$	Y_{i,j_Sim}	$Var(Y_{i,j})$	$Var(Y_{i,j})_{Sim}$	$P_{i,j}$	P_{i,j_Sim}
15,8852	15,8890	0,0667	0,0784	8.191.287	8.270.454
16,5898	16,5879	0,04	0,0539	16.351.238	16.433.788
15,9608	15,9625	0,0692	0,0831	8.845.420	8.922.172
17,1606	17,1607	0,0334	0,0482	28.840.447	29.057.872
17,0297	17,0318	0,0434	0,0588	25.429.690	25.679.408
16,4007	16,4006	0,0726	0,0862	13.756.529	13.848.673
16,9296	16,9283	0,0334	0,0481	22.891.075	23.030.961
17,0209	17,0211	0,0386	0,0528	25.146.589	25.329.999
16,8901	16,8905	0,0486	0,0624	22.172.679	22.335.770
16,2611	16,2604	0,0778	0,0915	11.994.606	12.068.991
17,1711	17,1724	0,04	0,0551	29.240.506	29.501.451
17,4389	17,4358	0,0434	0,0586	38.285.976	38.458.873
17,5303	17,5284	0,0486	0,0634	42.058.388	42.291.640
17,3994	17,3991	0,0586	0,0730	37.084.437	37.340.908
16,7704	16,7672	0,0878	0,1015	20.061.320	20.134.594
18,2606	18,2586	0,0667	0,0830	88.096.474	88.640.551
17,4508	17,4490	0,0692	0,0834	39.247.196	39.456.131
17,7186	17,7166	0,0726	0,0891	51.388.208	51.709.393
17,8100	17,8091	0,0778	0,0915	56.451.615	56.788.818
17,6791	17,6794	0,0878	0,1021	49.775.478	50.146.035
17,0501	17,0534	0,117	0,1323	26.926.708	27.222.380

Und daraus Ergebnisse aus der Simulation gesamt:

Anf.Jahr:	Reserve:	Varianz:	Standardfehler:
2	8.270.454	5,57841E+12	2.361.866
3	25.355.960	26.225.724.883.395	5.121.106
4	68.585.954	149.954.151.268.540	12.245.577
5	82.765.722	215.104.744.390.753	14.666.450
6	167.727.466	1.094.453.270.189.360	33.082.522
7	313.963.308	6.570.457.641.733.470	81.058.359
Gesamt_Sim:	666.668.864	8,06177E+15	89.787.382
Gesamt_Regr:	662.235.865	7,574E+15	87.028.753

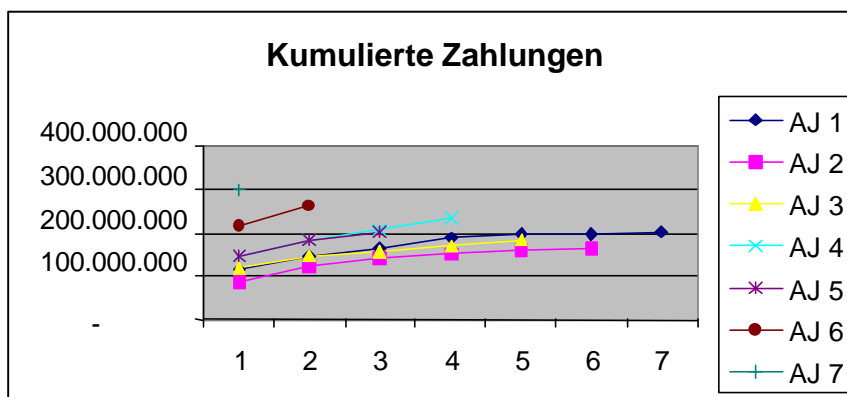
2.4.8.2 Berechnung mit dem Abwicklungsdreieck der angefallenen Schadenzuwächse

Analog zu Kapitel 2.4.8.1 führen wir jetzt dieselben Berechnungen mit dem Dreieck der angefallenen Schadenzuwächse durch. Hier wird auf die theoretischen Erläuterungen des vorigen Kapitels verzichtet und viel mehr nur auf die Ergebnisse eingegangen.

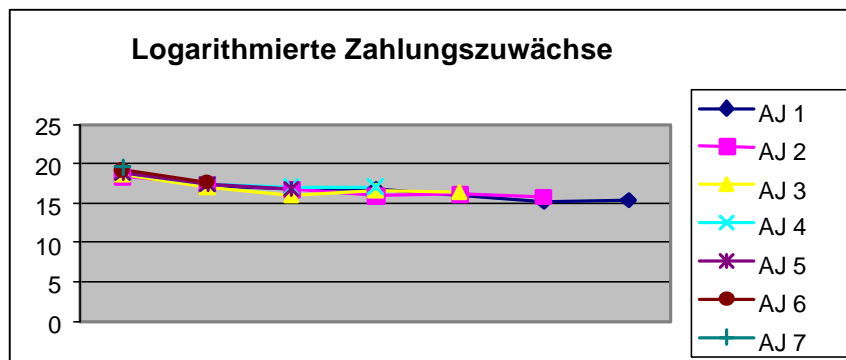
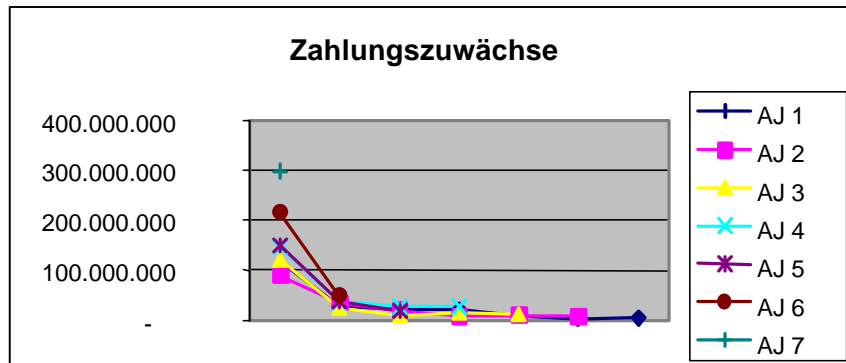
Gegeben das Dreieck der angefallenen Schadenzuwächse:

116.324.464	29.446.291	21.068.977	19.532.007	9.213.354	4.041.664	4.438.220
88.735.292	34.528.166	19.260.713	7.590.635	9.906.000	7.259.671	
121.026.936	24.638.451	9.249.534	14.950.622	12.397.939		
146.309.408	36.264.614	26.326.939	26.266.618			
148.605.503	36.976.087	17.970.956				
214.033.228	48.070.326					
297.779.107						

Mit Hilfe von Plots kann man sich die Zahlungsabwicklung veranschaulichen.



Mit dem Plot der kumulierten Zahlungen allein kann man das Modell nur schwer erkennen. Der Plot der Zahlungszuwächse zeigt deutlich, dass die Zuwächse exponentiell abfallen.



Analog zu den bezahlten Schäden wurde auf die angefallenen Schadenzuwächse das gleiche Regressionsmodell angewandt.

Man erhält folgende Ergebnisse:

Standardabweichung: 0,296229

R^2 : 0,95813

F : 33,9479

Anzahl der Beobachtungen: 28

Anzahl der Freiheitsgrade: 28-13=15

	<u>Parameter:</u>	<u>Standardfehler:</u>	<u>T-Werte:</u>
a ₁ :	18,5620	0,1721	107,8
a ₂ :	18,4805	0,1721	107,4
a ₃ :	18,4423	0,1762	104,7
a ₄ :	18,9319	0,1844	102,7
a ₅ :	18,7507	0,1986	94,4
a ₆ :	19,1170	0,2263	84,5
a ₇ :	19,5119	0,2964	65,8
b ₂ :	-1,3651	0,1711	-8,0
b ₃ :	-1,9397	0,1835	-10,6
b ₄ :	-2,0455	0,1986	-10,3
b ₅ :	-2,3354	0,2204	-10,6
b ₆ :	-3,0164	0,2574	-11,7
b ₇ :	-3,2562	0,3425	-9,5

Die Parameter sind laut *T*- und *F*-Statistiken wieder signifikant von Null verschieden.

Der Zahlungsmuster ist der Chain-Ladder Abwicklung sehr ähnlich.

	<u>Parameter:</u>	<u>exp(b_i):</u>	<u>Zahlung in % des Endschaten</u> <u>(nach KQ):</u>	<u>Zahlung in % des Endschaten</u> <u>(nach CL):</u>
b ₁ :	0	1	58%	58%
b ₂ :	-1,3651	0,1711	15%	15%
b ₃ :	-1,3651	0,1835	8%	9%
b ₄ :	-1,3651	0,1986	8%	8%
b ₅ :	-1,3651	0,2204	6%	6%
b ₆ :	-1,3651	0,2574	3%	3%
b ₇ :	-1,3651	<u>0,3425</u>	<u>2%</u>	<u>2%</u>
Summe:		2,3734	100%	100%

Zahlungsprognose aufgrund der geschätzten Regression

Analog berechnen wir die zukünftigen Zahlungen und deren Standardabweichungen mit geschätzten Parametern und führen die angepassten Werte in die Original-Skalierung über.

Die Ergebnisse sind in der folgenden Tabelle aufgelistet:

i	j	Y_{ij}	$\text{Var}(Y_{ij})$	P_{ij}	$\text{s.e.}(P_{ij})$
2	7	15,2243	0,1169	4.337.318	1.527.573
3	6	15,4259	0,0701	5.183.552	1.397.057
3	7	15,1862	0,1213	4.184.128	1.502.662
4	5	16,5965	0,0586	16.614.333	4.079.840
4	6	15,9155	0,0761	8.482.689	2.385.010
4	7	15,6757	0,1273	6.847.168	2.522.569
5	4	16,7052	0,0586	18.521.410	4.548.145
5	5	16,4153	0,0677	13.924.162	3.684.257
5	6	15,7343	0,0852	7.109.183	2.120.082
5	7	15,4945	0,1364	5.738.484	2.193.624
6	3	17,1774	0,0701	29.872.370	8.051.122
6	4	17,0715	0,0761	26.951.114	7.577.629
6	5	16,7816	0,0852	20.261.508	6.042.334
6	6	16,1006	0,1027	10.344.807	3.402.665
6	7	15,8608	0,1539	8.350.257	3.406.213
7	2	18,1468	0,1169	80.620.103	28.393.843
7	3	17,5722	0,1213	45.483.864	16.334.791
7	4	17,4663	0,1273	41.035.941	15.118.073
7	5	17,1765	0,1364	30.850.303	11.793.005
7	6	16,4955	0,1539	15.751.070	6.425.130
7	7	16,2557	0,2051	12.714.156	6.066.373

Schätzung der Reserve

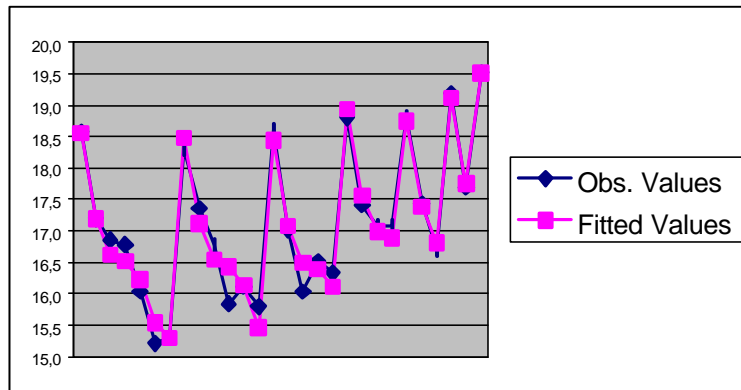
Jetzt können wir die Reserven und deren Standardabweichungen für jedes Anfalljahr $i = 2, \dots, 7$ schätzen.

Anf. Jahr:	Reserve:	Varianz:	St. Fehler:
2	4.337.318	2.333.480.170.662	1.527.573
3	9.367.680	5.365.536.642.316	2.316.363
4	31.944.189	47.638.307.650.986	6.902.051
5	45.293.240	98.024.698.338.080	9.900.742
6	95.780.056	565.501.421.612.776	23.780.274
7	226.455.438	5.838.066.570.098.960	76.407.242
Gesamt =	413.177.920	6.556.930.014.513.780	80.974.873

Der Standardfehler der Gesamtreserve beträgt 20% der Gesamtreserve von 413.177.922. Die Gesamtreserve nach Chain-Ladder betrug 407.046.090.

Graphische Darstellungen

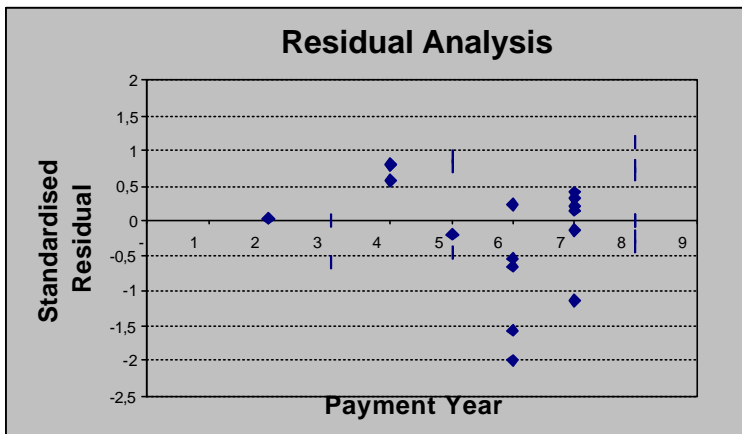
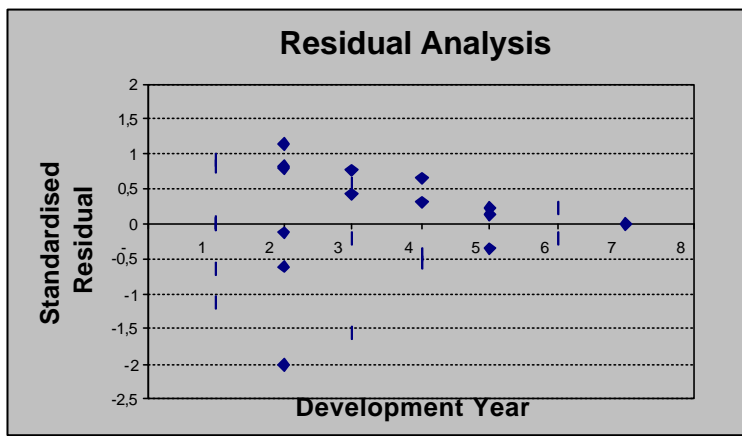
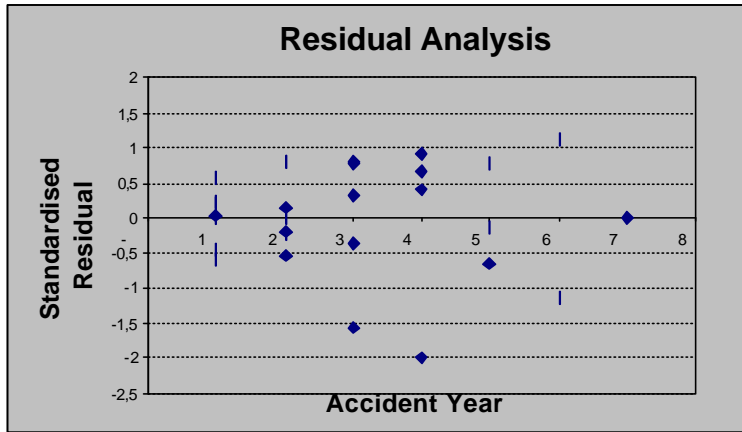
Der Plot der beobachteten Werte gegen die angepassten zeigt eine relativ gute Anpassung.

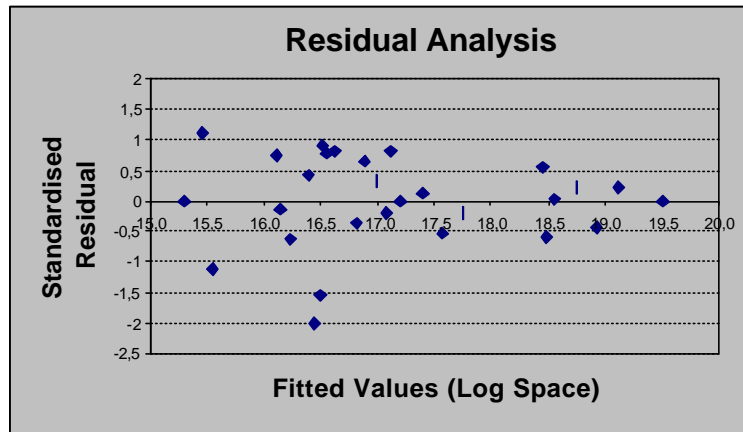


Die folgende Residuentabelle enthält alle zur Residuenanalyse verwendeten Werte:

<i>i</i>	<i>j</i>	<i>i + j</i>	Y_{i,j_Obs}	Y_{i,j_Fit}	$Y_{Obs} - Y_{Fit}$	St. Residuen
1	1	2	18,5719	18,5620	0,0099	0,0336
1	2	3	17,1981	17,1969	0,0012	0,0040
1	3	4	16,8633	16,6223	0,2410	0,8136
1	4	5	16,7876	16,5164	0,2712	0,9153
1	5	6	16,0362	16,2266	-0,1904	-0,6427
1	6	7	15,2122	15,5456	-0,3334	-1,1254
1	7	8	15,3058	15,3058	0,0000	0,0000
2	1	3	18,3012	18,4805	-0,1793	-0,6054
2	2	4	17,3573	17,1154	0,2419	0,8165
2	3	5	16,7736	16,5408	0,2327	0,7857
2	4	6	15,8424	16,4350	-0,5925	-2,0002
2	5	7	16,1087	16,1451	-0,0364	-0,1230
2	6	8	15,7978	15,4641	0,3338	1,1267
3	1	4	18,6115	18,4423	0,1692	0,5711
3	2	5	17,0198	17,0773	-0,0575	-0,1940
3	3	6	16,0401	16,5027	-0,4626	-1,5617
3	4	7	16,5203	16,3968	0,1235	0,4168
3	5	8	16,3330	16,1069	0,2261	0,7633
4	1	5	18,8012	18,9319	-0,1307	-0,4411
4	2	6	17,4064	17,5668	-0,1605	-0,5417
4	3	7	17,0861	16,9922	0,0939	0,3168
4	4	8	17,0838	16,8864	0,1974	0,6665
5	1	6	18,8168	18,7507	0,0661	0,2232
5	2	7	17,4258	17,3856	0,0401	0,1355
5	3	8	16,7043	16,8111	-0,1068	-0,3605
6	1	7	19,1816	19,1170	0,0646	0,2181
6	2	8	17,6882	17,7520	-0,0638	-0,2153
7	1	8	19,5119	19,5119	0,0000	0,0000

Alle Residuenplots zeigen kein systematisches Muster, was auf die Richtigkeit des Modells hinweist.





Simulationsergebnisse:

Mit Matlab werden nach dem gleichen Algorithmus wie oben 21 neue Prognosewerte simuliert, die in folgender Tabelle im Vergleich zu den Regressionswerten aufgelistet sind:

Y_{ij}	Y_{i,j_Sim}	$Var(Y_{ij})$	$Var(Y_{i,j})_{Sim}$	P_{ij}	P_{i,j_Sim}
15,2243	15,2254	0,1169	0,1275	4.337.318	4.365.093
15,4259	15,4266	0,0701	0,0809	5.183.552	5.215.002
15,1862	15,1869	0,1213	0,134	4.184.128	4.213.904
16,5965	16,5997	0,0586	0,0727	16.614.333	16.785.903
15,9155	15,9132	0,0761	0,0885	8.482.689	8.515.938
15,6757	15,673	0,1273	0,138	6.847.168	6.865.367
16,7052	16,7088	0,0586	0,0729	18.521.410	18.722.751
16,4153	16,4145	0,0677	0,0813	13.924.162	14.008.152
15,7343	15,7348	0,0852	0,0971	7.109.183	7.155.201
15,4945	15,489	0,1364	0,1497	5.738.484	5.745.055
17,1774	17,1761	0,0701	0,0845	29.872.370	30.049.302
17,0715	17,0675	0,0761	0,0908	26.951.114	27.041.952
16,7816	16,7864	0,0852	0,0992	20.261.508	20.501.323
16,1006	16,0969	0,1027	0,1168	10.344.807	10.379.051
15,8608	15,8667	0,1539	0,1676	8.350.257	8.456.957
18,1468	18,1483	0,1169	0,1323	80.620.103	81.364.531
17,5722	17,5702	0,1213	0,135	45.483.864	45.704.165
17,4663	17,4671	0,1273	0,1418	41.035.941	41.367.247
17,1765	17,1789	0,1364	0,1511	30.850.303	31.153.900
16,4955	16,4915	0,1539	0,1703	15.751.070	15.817.856
16,2557	16,2507	0,2051	0,2233	12.714.156	12.766.690

Die Reserveschätzer nach Simulation:

Anf.Jahr:	Reserve:	Varianz:	St. Fehler:
2	4.365.093	2,59106E+12	1.609.677
3	9.428.905	5.993.622.852.993	2.448.188
4	32.167.208	53.874.264.264.288	7.339.909
5	45.631.159	108.137.897.222.867	10.398.937
6	96.428.586	602.922.373.366.398	24.554.478
7	228.174.390	6.063.911.442.621.890	77.871.121
Gesamt_Sim:	416.195.341	6,83743E+15	82.688.758
Gesamt_Regr:	413.177.920	6,55693E+15	80.974.873

Die Ergebnisse der Simulation und der Regression unterscheiden sich wieder nur sehr gering.

Abschließend muss noch betont werden, dass das Verfahren recht empfindlich auf einzelne Ausreißer reagiert, denn diese beeinflussen sehr die Höhe von \hat{S}^2 . Es ist dann besser, solche Ausreißer durch einen interpolierten Wert zu ersetzen.

Auf jeden Fall scheint dieses Verfahren geeigneter als Chain-Ladder. Während bei Chain-Ladder nur die Abwicklungsparameter bestimmt werden, werden in diesem Modell auch Anfalljahrparameter bestimmt. Die größere Anzahl der Parameter garantiert im Allgemeinen eine bessere Anpassung, andererseits könnte dabei die Stabilität verloren gehen, wie z.B. Sensibilität gegenüber Datenänderungen.

Bei Reservenbildung und Prognosen werden Beziehungen zwischen einzelnen Spalten und Zeilen des Dreieckes viel mehr ausgenutzt. So wird beim Chain-Ladder-Verfahren der Prognosewert $\hat{P}_{i,j}$ im unteren Dreieck einzig und allein durch den bekannten Vorgänger $P_{i,j-1}$ bestimmt, während bei diesem Modell viel mehr Daten berücksichtigt werden.

3 Zusammenfassung

3.1 Vergleich der Verfahren

Aus der Fülle der vorliegenden Verfahren sollte zum Schätzen der Reserve das gewählt werden, welches auf die Daten am besten „passt“.

Welches Verfahren diese Eigenschaft hat, lässt sich z. B. mit folgenden drei Methoden analysieren, bzw. testen:

Methode 1: durch die Verwendung des bekannten Dreiecks.

Man erzeugt das bekannte (obere) Dreieck durch das vorliegende Modell, wobei die erste Spalte die Anfangswerte liefert.

Anschließend vergleicht man die bekannten Werte mit den durch das Modell erzeugten unter Verwendung von Prognosefehlermaßen.¹

Die durch das Modell erzeugten Werte werden in Zukunft als Prognosewerte bezeichnet.

Methode 2: durch die Verwendung eines Teils des bekannten Dreiecks.

Aus dem Abwicklungsdreieck schneidet man ein kleineres Dreieck der Seitenlänge n^* aus und zwar so, dass das zu diesem Dreieck gehörende Rechteck immer noch im Dreieck enthalten ist. Jetzt kann man das kleine Dreieck mit Hilfe des zu untersuchenden Verfahrens zum Rechteck vervollständigen. Somit kann man die Prognosewerte mit tatsächlichen Werten vergleichen. Die Seitenlänge n^* des kleinen Dreiecks soll so gewählt werden, dass möglichst viele Werte zum Vergleich zur Verfügung stehen. Also soll die Anzahl der Elemente im Schnittbereich des unteren kleinen Dreiecks mit dem bekannten Dreieck maximiert werden.

In unserem Fall für $n = 7$ wird man das Dreieck der Seitenlänge $n^* = 4$ rausschneiden. Damit erhält man 6 Werte zum Vergleich.

Methode 3: mit Hilfe von Simulation.

Man simuliert nach dem Modell mit Hilfe der geschätzten Parameter das ganze Rechteck, entnimmt nur das obere Dreieck und wendet darauf das zu untersuchende Verfahren an. Anschließend vergleicht man das untere Dreieck aus der Simulation mit dem neu berechneten mittels der Vergleichsmaße.

Nach $n-1$ Jahren, wenn das untere Dreieck vollständig ist, empfiehlt es sich die Ergebnisse zu überprüfen, um die möglicherweise entstandenen Fehlentscheidungen zu analysieren.

¹ Die Prognosefehlermaße werden im folgenden Kapitel „Prognosefehlermaße“ beschrieben.

3.2 Prognosefehlermaße

Durch Prognosefehlermaße erhält man Informationen über die Güte konkurrierender Verfahren.

Sie geben in der Regel einen Durchschnitt der Abweichungen zwischen den tatsächlichen $C_{i,j}$, $i = 1, \dots, n, j = 2, \dots, n-i+1$, und den Prognosewerten $\hat{C}_{i,j}$, $i = 1, \dots, n, j = 2, \dots, n-i+1$ wieder.

In dieser Arbeit wurden 4 Prognosefehlermaße verwendet.

Der mittlere absolute Prognosefehler (mean absolute error) wird aus den absoluten Abweichungen $|C_{i,j} - \hat{C}_{i,j}|$ gebildet,

$$MAE = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=2}^{n-i+1} |C_{i,j} - \hat{C}_{i,j}| \quad (46)$$

während sich die Wurzel des mittleren quadratischen Fehlers (root mean square error) aus den quadratischen Abweichungen $(C_{i,j} - \hat{C}_{i,j})^2$ zusammensetzt:

$$RMSE = \sqrt{\frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=2}^{n-i+1} (C_{i,j} - \hat{C}_{i,j})^2} \quad (47)$$

Beide Maße sind in der Größenordnung vergleichbar, da beim RMSE die Quadrierung der Einzelabweichungen durch das Ziehen der Wurzel aus der Summe der Abweichungsquadrate größenmäßig wieder rückgängig gemacht wird.

Allgemein ist der mittlere absolute Fehler stets kleiner oder gleich der Wurzel des mittleren quadratischen Fehlers, da sich größere Prognosefehler erheblich stärker im *RMSE* als im *MAE* niederschlagen.

Zwei andere vergleichbare Prognosefehlermaße sind relative Maße.

Mit den relativen Fehlern $\frac{|C_{i,j} - \hat{C}_{i,j}|}{C_{i,j}}$ ergibt sich der mittlere prozentuale Prognosefehler

(mean absolute percentage error) in der Form:

$$MAPE = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=2}^{n-i+1} \frac{|C_{i,j} - \hat{C}_{i,j}|}{C_{i,j}} \quad (48)$$

Er gibt an, um wieviel Prozent die Prognosewerte von den tatsächlichen abweichen.

Ein entsprechender prozentualer Durchschnitt auf der Grundlage der Wurzel des mittleren quadratischen Fehlers ist durch

$$RMSPE = \sqrt{\frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=2}^{n-i+1} \left(\frac{C_{i,j} - \hat{C}_{i,j}}{C_{i,j}} \right)^2} \quad (49)$$

(root mean square percentage error) gegeben.

3.3 Numerische Ergebnisse

Berechnung nach Methode 1

Mit der Excel-Datei „Verfahren“¹ wurden die oben definierten Maße berechnet,² zuerst mit dem Dreieck der angefallenen Schadenzuwächse und anschließend mit dem Dreieck der bezahlten Schadenzuwächse.

Die in der Tabelle unten aufgelisteten Maße zeigen in beiden Fällen einen deutlichen Vorteil der Kleinste-Quadrate-Regression gegenüber den anderen Verfahren. Alle Maße sind für die Kleinste-Quadrate-Regression wesentlich kleiner als für die anderen Verfahren.

An der zweiten Stelle steht das Verfahren nach dem Log-inversen gaußschen Modell. Die Verfahren Chain-Ladder und robuste Faktoren, die sich den dritten Platz teilen, haben sehr ähnliche Maße, was nicht überraschend ist, da beide Verfahren, solange in den Daten keine Ausreißer vorhanden sind, auf ähnlichen Berechnungen basieren.

Die Verfahren nach den Log-Gamma- und Log-Normal-Modellen sollen den Maßen entsprechend nicht verwendet werden, da sie die höchsten Fehlermaße aufweisen.

Prognosefehlermaße für das Dreieck der angefallenen Schadenzuwächse

Maß	Chain-Ladder	robuste Faktoren	KQ-Regression	Log Normal	Log Gamma	Log IGauß
MAE	8.317.867	8.322.466	4.664.037	19.847.178	18.697.622	8.269.126
RMSE	10.266.711	10.257.738	5.341.602	27.327.515	27.809.863	10.217.242
MAPE	0,05011	0,05049	0,02747	0,10800	0,10430	0,04973
RMSPE	0,06426	0,06468	0,03281	0,14291	0,14692	0,06348

¹ Siehe die Datei „verfahren.xls“ auf der beiliegenden Diskette

² Im Tabellenblatt „Daten“ kann man das entsprechende Dreieck (bezahlt bzw. angefallen) auswählen. Im Tabellenblatt „Dreieck“ können die Berechnungen durchgeführt werden, indem man den Button für das gewählte Verfahren anklickt. Nach dem alle Verfahren berechnet wurden, bildet sich im Tabellenblatt „Vergleich“ die Tabelle mit den Prognosefehlermaßen.

Prognosefehlermaße für das Dreieck der bezahlten Schadenzuwächse

Maß	Chain-Ladder	robuste Faktoren	KQ-Regression	Log Normal	Log Gamma	Log IGauß
MAE	10.559.655	10.835.183	3.478.098	24.579.995	21.450.852	10.009.134
RMSE	12.964.047	13.699.126	4.528.998	36.440.076	26.891.070	11.985.974
MAPE	0,07864	0,08208	0,02914	0,17353	0,16465	0,07379
RMSPE	0,09136	0,09799	0,03967	0,22846	0,19627	0,08442

Berechnung nach Methode 2

Mit derselben Excel-Datei „Verfahren“ vervollständigt man die 2 kleineren Dreiecke jeweils mit bezahlten und angefallenen Schadenzuwächsen zu den Rechtecken. Die resultierenden Prognosefehlermaße zeigen, dass sowohl die Kleinste-Quadrate-Methode als auch die Methode der robusten Faktoren sehr niedrige Maße aufweisen.

Prognosefehlermaße für das Dreieck der bezahlten Schadenzuwächse

Maß	Chain-Ladder	robuste Faktoren	KQ-Regression	Log Normal	Log Gamma	Log IGauß
MAE	20.090.283	20.113.947	20.112.120	38.063.311	66.351.235	20.394.421
RMSE	23.137.539	23.181.398	23.158.049	42.063.918	69.824.259	23.552.229
MAPE	0,14353	0,14376	0,14339	0,28033	0,48935	0,14565
RMSPE	0,15899	0,15946	0,15855	0,30894	0,49816	0,16238

Prognosefehlermaße für das Dreieck der angefallenen Schadenzuwächse

Maß	Chain-Ladder	robuste Faktoren	KQ-Regression	Log Normal	Log Gamma	Log IGauß
MAE	8.553.888	7.739.947	8.951.809	24.822.960	63.815.307	12.508.683
RMSE	9.925.679	9.616.309	9.556.585	28.933.052	72.259.560	16.453.152
MAPE	0,05051	0,04659	0,05187	0,14114	0,35070	0,07273
RMSPE	0,06028	0,05902	0,05730	0,16845	0,39474	0,09893

Berechnung nach Methode 3

Alle Verfahren, die in dieser Arbeit behandelt wurden, legen stets den Daten eine Verteilung zugrunde. So sind die Zahlungen nach Chain-Ladder und der Methode der robusten Faktoren normalverteilt mit $E(C_{i,j}) = f_j C_{i,j-1}$ und $Var(C_{i,j}) = s_j^2$. Anhand der ersten Spalte und der in §2.1 und §2.2 ermittelten Faktoren berechnet man die Erwartungswerte für die sechs anderen Spalten. Als Schätzer für die Varianz s_j^2 wird

$$s_j^2 = \frac{1}{n-j} \sum_{i=2}^{n-j+1} (C_{i,j} - f_j C_{i,j-1})^2 \quad (50)$$

verwendet.

Der Simulationsalgorithmus zur Erzeugung einer normalverteilten Variablen ist auf der Seite 57 beschrieben.

Für die KQ-Methode erzeugt man analog normalverteilte Variablen mit den im §2.4 berechneten Parametern. Aufgrund der Annahme der Log-normalverteilten Schadenzuwächse werden die erzeugten Daten in die Original-Skalierung geführt.

Für die Methode der stochastischen Faktoren erzeugt man nach den berechneten Parametern die entsprechend verteilten Zufallszahlen. Die Simulationsalgorithmen wurden bereits im §2.3 beschrieben.

Die Prognosefehlermaße zeigen im Fall der bezahlten Schadenzuwächse einen deutlichen Vorteil der KQ-Methode und im Falle der angefallenen Schadenzuwächse sind das Log-inverse gaußsche Modell gefolgt von der Methode der robusten Faktoren besser als die KQ-Methode.

Prognosefehlermaße für das Dreieck der bezahlten Schadenzuwächse

Maß	Chain-Ladder	robuste Faktoren	KQ-Regression	Log Normal	Log Gamma	Log IGauß
MAE	34.552.334	38.450.124	16.163.892	46.272.416	248.501.773	20.482.125
RMSE	51.340.997	51.546.552	25.040.515	53.875.853	314.599.025	25.896.090
MAPE	0,12829	0,16015	0,04960	0,17244	0,59121	0,07502
RMSPE	0,18137	0,22420	0,07080	0,19834	0,70839	0,09595

Prognosefehlermaße für das Dreieck der angefallenen Schadenzuwächse

Maß	Chain-Ladder	robuste Faktoren	KQ-Regression	Log Normal	Log Gamma	Log IGauß
<i>MAE</i>	24.012.047	18.473.971	21.409.838	21.939.153	955.311.594	13.410.229
<i>RMSE</i>	32.268.746	25.280.444	26.876.929	26.773.815	1.347.217.719	18.756.142
<i>MAPE</i>	0,08446	0,06884	0,06557	0,06627	1,22635	0,03795
<i>RMSPE</i>	0,11481	0,10334	0,08698	0,07955	1,68656	0,04754

4 Literaturverzeichnis

T. Mack, Schadenversicherungsmathematik, DGVM-Schriftenreihe, Verlag Versicherungswirtschaft, Karlsruhe 1997.

Institut of Actuaries (Hg.), Claims Reserving Manual, Band 1-2, 1989.

A.M. Law, W.D. Kelton, Simulation Modelling and Analysis, McGraw-Hill, 1991.

K. Behnen, G. Neuhaus, Grundkurs Stochastik, Teubner Verlag, 1995.

J. Johnston, J. DiNardo, Econometric Methods, McGraw-Hill, 1997.

V. Kelly, Practical loss reserving method with stochastic development factors, PCAS, 1997.

E. Kremer, Robust Lagfactors, Blätter der deutschen Gesellschaft für Versicherungsmathematik, 1997.

E. Scanlon, Residuals and Influence in Regression, PCAS, 1994.

R. Verral, Statistical Methods for the Chain Ladder Technique, PCAS, 1994.

Johnson, Kotz, Balakrishnan, Continuous Univariate Distributions, Vol. 1, John Wiley & Sons, 1994.

R. Hogg, S. Klugmann, Loss Distributions, John Wiley & Sons, 1984.

S. Brandt, Datenanalyse, B. I. Wissenschaftsverlag, 1981.

H. Schneeweiß, Ökonometrie, Physica – Verlag, 1974.

H. Eckey, R. Kosfeld, C. Dreger, Ökonometrie, Gabler – Verlag, 1995.

S. Press, Applied Multivariate Analysis, Series in quantitative Methods for Decision - Making, 1971.

R. Ansorge, H. Oberle, Mathematik für Ingenieure, Band 1, Akademie Verlag, 1994.

P. Holland, Robust Regression using iteratively Reweighted Least - Squares, Commun. Statist. – Theor. Meth.,
A6, 1997, 813 - 827.

D. Andrews, A Robust Method for Multiple Linear Regression, Technometrics, Vol. 16, 1974, 523 - 531.

L. Doray, IBNR Reserve Under a Loglinear Location-Scale Regression Model, PCAS 1994.

P. Huber, Robust Statistics, Wiley series in probability and mathematical statistics, 1981.

D. Pfeifer, Bedingte Verteilungen, Skript zur Vorlesung “Versicherungsmathematik 2”.

Anhang A: Hilfssätze

Hilfssätze aus dem Skript „Bedingte Verteilungen“ von D. Pfeifer.

Satz 1. (Eigenschaften des allgemeinen bedingten Erwartungswerts)

Es sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum, X eine nicht negative bzw. Integrierbare Zufallsvariable und $\mathcal{C} \subseteq \mathcal{A}$ eine Teil- \mathbf{s} -Algebra. $\mathcal{P}_{\mathcal{C}}$ ist die Einschränkung von \mathcal{P} auf \mathcal{C} . Dann gilt:

a) $E[E(X | \mathcal{C})] = E(X)$

b) Ist X bereits \mathcal{C} -meßbar, so ist $E(X | \mathcal{C}) = X$ $\mathcal{P}_{\mathcal{C}}$ -f.s.

c) Ist Y eine weitere, \mathcal{C} -meßbare Zufallsvariable, so dass XY nicht negativ bzw. integrierbar ist, so gilt

$$E(XY | \mathcal{C}) = YE(X | \mathcal{C}) \quad \mathcal{P}_{\mathcal{C}}\text{-f.s.}$$

d) Ist $\mathcal{D} \subseteq \mathcal{C}$ eine Teil- \mathbf{s} -Algebra von \mathcal{C} , so ist

$$E[E(X | \mathcal{C}) | \mathcal{D}] = E(X | \mathcal{D}) \quad \mathcal{P}_{\mathcal{D}}\text{-f.s.}$$

e) Ist die \mathbf{s} -Algebra $\mathcal{D} \subseteq \mathcal{A}$ unabhängig von $\mathbf{s}(\mathbf{s}(X), \mathcal{C})$, so ist

$$E[X | \mathbf{s}(\mathcal{C}, \mathcal{D})] = E(X | \mathcal{C}) \quad \mathcal{P}_{\mathbf{s}(\mathcal{C}, \mathcal{D})}\text{-f.s.}$$

wobei $\mathbf{s}(\mathcal{C}, \mathcal{D})$ die von \mathcal{C} und \mathcal{D} gemeinsam erzeugte \mathbf{s} -Algebra ist.

Ist speziell X unabhängig von \mathcal{D} , so gilt

$$E(X | \mathcal{D}) = E(X) \quad \mathcal{P}_{\mathcal{D}}\text{-f.s.}$$

Lemma 1. (Ersetzungslemma)

X und Y seien Zufallselemente auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$ mit Werten in Meßräumen $(\mathcal{X}, \mathcal{B})$ bzw. $(\mathcal{Y}, \mathcal{D})$. Die bedingte Verteilung $\mathcal{P}(\cdot | \mathbf{Y} = \mathbf{y}), \mathbf{y} \in \mathcal{Y}$, sei regulär,

$G: (\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{D}) \rightarrow (\mathcal{Z}, \mathcal{F})$ sei eine meßbare Abbildung. Dann gilt für alle $A \in \mathcal{F}$:

$$\begin{aligned} \mathcal{P}(G(\mathbf{X}, \mathbf{Y}) \in A | \mathbf{Y} = \mathbf{y}) &= \mathcal{P}(G(\mathbf{X}, \mathbf{y}) \in A | \mathbf{Y} = \mathbf{y}) \\ &:= \mathcal{P}(G(\mathbf{X}, \mathbf{z}) \in A | \mathbf{Y} = \mathbf{y}) \Big|_{\mathbf{z} = \mathbf{y}} \quad \mathcal{P}^{\mathbf{Y}}\text{-f.s.} \end{aligned} \quad (1)$$

Sind insbesondere \mathbf{X} und \mathbf{Y} stochastisch unabhängig, so gilt auch

$$\mathcal{P}(G(\mathbf{X}, \mathbf{Y}) \in A \mid \mathbf{Y} = \mathbf{y}) = \mathcal{P}(G(\mathbf{X}, \mathbf{y}) \in A) \quad \mathcal{P}^{\mathbf{Y}}\text{-f.s.} \quad (2)$$