

Ausgleich von Invalidierungshäufigkeiten mit Bailey & Simon

Autor(en): **Thurnherr, Willi**

Objektyp: **Article**

Zeitschrift: **Mitteilungen / Schweizerische Aktuarvereinigung = Bulletin / Association Suisse des Actuaires = Bulletin / Swiss Association of Actuaries**

Band (Jahr): **- (1997)**

Heft 2

PDF erstellt am: **01.05.2024**

Persistenter Link: <https://doi.org/10.5169/seals-967342>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

B. Wissenschaftliche Mitteilungen

WILLI THURNHERR, Basel

Ausgleich von Invalidierungshäufigkeiten mit Bailey & Simon

Für die Tarifierung der Invaliditätsversicherung müssen zwei verschiedene Komplexe bearbeitet werden. Zuerst muss ein Tarifmodell festgelegt werden, bevor in einem zweiten Schritt die Tarifvariablen durch eine geeignete Methode bestimmt werden können. Das Ziel dieser Arbeit besteht darin, ausgehend von einem vorgegebenen Tarifmodell ein Schätzverfahren für die Invalidierungshäufigkeiten vorzustellen, welches auch bei teilweise unplausiblen (bezüglich des vorgegebenen Tarifmodells) statistischen Rohdaten Schätzungen für die Invalidierungshäufigkeiten liefert, welche den Anforderungen des gewählten Modells genügen. Das Schätzverfahren wurde von Bailey & Simon [1] zur Tarifierung in der kanadischen Automobilversicherung entwickelt. Die Methode von Bailey & Simon ist ein Repräsentant aus einer Gruppe von pragmatischen statistischen Parameterschätzungsverfahren, welche Schätzungen aufgrund eines mathematischen Modells liefern. Das Verfahren wird am Beispiel der Daten aus der VPL¹-Gemeinschaftsstatistik 1986–1990 [3] expliziert. Den erzielten Resultaten werden die mittels Spline-Ausgleichung geglätteten Werte gegenübergestellt.

1 Invaliditätsmodell

Während bei der Todesfallversicherung üblicherweise nach den Merkmalen Geschlecht, Alter sowie Deckung unterschieden wird, kommt bei der Invaliditätsversicherung zusätzlich das Merkmal Wartefrist hinzu. Wird für alle Wartefristen und Deckungen ein einheitliches Tarifierungsmodell gewählt, um eine geschlossene und konsistente Tarifierung der Invaliditätsversicherung zu erreichen, so ergeben sich zwischen den einzelnen Wartefristen Abhängigkeiten, welche durch Sterblichkeit und Reaktivierung der Invaliden bestimmt werden.

¹ Verein Privater Lebensversicherer (Schweiz)

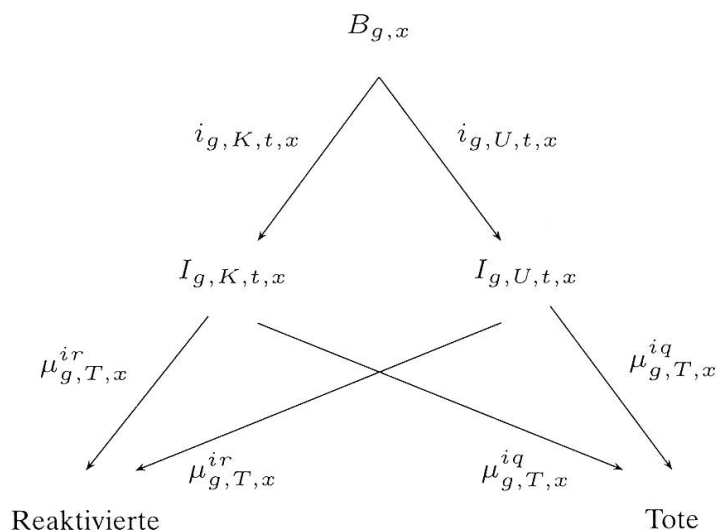
Im Folgenden werden als Bezeichnungen verwendet:

g	Geschlecht (Wertebereich: Mann / Frau)
d	Ursache der Invalidität (Wertebereich: U = Invalidität durch Unfall, K = Invalidität durch Krankheit, $U + K$ = Invalidität durch Unfall oder Krankheit)
w	Wartezeit (Wertebereich: 3, 6, 12, 24 Monate)
x	Alter (Wertebereich: 15, 16, ..., 64)
$B_{g,x}$	Bestand unter Risiko (x -jährige aktive Versicherte mit Geschlecht g , welche dem Invalidierungsrisiko unterliegen)
$i_{g,d,t,x}$	Wahrscheinlichkeit für einen x -jährigen Versicherten mit Geschlecht g durch die Ursache d invalid zu werden und nach Ablauf der Zeit t noch invalid zu sein
$I_{g,d,t,x}$	Bestand der aus $B_{g,x}$ durch die Ursache d hervorgegangenen Invaliden (mit Geschlecht g), welche im Alter x invalid wurden und nach Ablauf der Zeit t noch invalid sind
$\mu_{g,t,x}^{iq}$	Sterblichkeitsintensität für Personen mit Geschlecht g , welche im Alter x invalid wurden, nach Ablauf der Invaliditätsdauer t ($\mu_{g,t,x}^{iq} > 0$)
$\mu_{g,t,x}^{ir}$	Reaktivierungsintensität für Personen mit Geschlecht g , welche im Alter x invalid wurden, nach Ablauf der Invaliditätsdauer t ($\mu_{g,t,x}^{ir} > 0$)

Die Invalidierungswahrscheinlichkeit $i_{g,d,t,x}$ berechnet sich aus dem Bestand unter Risiko $B_{g,x}$ und dem daraus hervorgegangenen Invalidenbestand $I_{g,d,t,x}$ als

$$i_{g,d,t,x} = \frac{I_{g,d,t,x}}{B_{g,x}} \quad (1)$$

Für die Tarifierung der Invaliditätsversicherung wird folgendes Modell gewählt:



Dabei wird vernachlässigt, dass die Reaktivierten im Alter $x + t$ erneut dem Invaliditätsrisiko unterliegen. Für die folgenden Zusammenhänge ist das jedoch ohne Belang:

$$i_{g,d,t_1,x} < i_{g,d,t_2,x} \Leftrightarrow t_1 > t_2 \quad (2)$$

Grund:

$$I_{g,d,t_1,x} = I_{g,d,t_2,x} \cdot (t_1 - t_2) p_{g,x,t_2}^i < I_{g,d,t_2,x}$$

wobei ${}_T p_{g,t,x}^i = e^{-\int_t^{t+T} (\mu_{g,s,x}^{iq} + \mu_{g,s,x}^{ir}) ds}$

Bezüglich Ursache gilt:

$$i_{g,U+K,t,x} = i_{g,K,t,x} + i_{g,U,t,x} \quad (3)$$

Aus (3) folgt insbesondere auch $i_{g,U+K,t,x} \geq i_{g,K,t,x}$.

2 Datenmaterial

Als Datenbasis für die Ausgleichungen verwenden wir die Rohdaten aus der Gemeinschaftsstatistik VPL über den Verlauf der Invalidität in der Kollektivversicherung aus der Erhebungsperiode vom 1.1.1986 bis 31.12.1990².

Das Beobachtungsmaterial wurde nach Geschlecht und Alter unterteilt. Für die Bestimmung der Invalidität war das massgebende Kriterium, ob eine Invalidenrente resp. Prämienbefreiung zur Auszahlung gelangte. Deshalb wurde bei den Beständen unter Risiko nach versicherter Wartefrist differenziert und die sich daraus ergebenden Invaliditätsfälle erfasst. Was die Ursache der Invalidität betrifft, wurden die Bestände unter Risiko aufgrund der versicherten Deckung (entweder Volldeckung oder Krankheitsdeckung) separiert und die daraus hervorgegangenen Invaliden getrennt erfasst.

Bezeichnet man mit $B_{g,d,w,x}$ die Anzahl aktiver Versicherter mit Geschlecht g , Deckung d , Wartefrist w und Alter x , welche während der Beobachtungsperiode unter Invaliditätsrisiko gestanden haben und mit $I_{g,d,w,x}$ die aus diesem Bestand hervorgegangenen Invaliden, so erhält man den Mittelwert der Stichprobe als

$$\tilde{i}_{g,d,w,x} = \frac{I_{g,d,w,x}}{B_{g,d,w,x}} \quad (4)$$

Diese Anpassung der Definition der rohen Invalidierungswahrscheinlichkeiten gegenüber Definition (1) ist aufgrund des vorhandenen Beobachtungsmaterials notwendig, sonst aber äquivalent zu (1).

Mit den Bezeichnungen

$$\begin{aligned} I_{g,d,w,\bullet} &= \sum_x I_{g,d,w,x} \\ B_{g,d,w,\bullet} &= \sum_x B_{g,d,w,x} \\ \tilde{i}_{g,d,w,\bullet} &= \frac{I_{g,d,w,\bullet}}{B_{g,d,w,\bullet}} \end{aligned}$$

gibt die Tabelle 1 einen Überblick über das Datenmaterial ($x = 15, 16, \dots, 64$):

² Diese Daten dienen nicht als alleinige Basis für den Tarif 1995 der Kollektivlebensversicherung [5], da in diesen die massive Zunahme der ausbezahlten Invaliditätsleistungen aus den Jahren 1990 bis 1994 noch nicht enthalten ist.

Wartefrist, Deckung	$B_{g,d,w,\bullet}$		$I_{g,d,w,\bullet}$		$\tilde{i}_{g,d,w,\bullet}$	
	Männer	Frauen	Männer	Frauen	Männer	Frauen
3 Monate, Volldeckung	3 015 916	1 292 824	21 383	4 978	7.090 ‰	3.850 ‰
6 Monate, Volldeckung	310 207	150 714	1 421	351	4.581 ‰	2.329 ‰
12 Monate, Volldeckung	104 523	39 982	529	120	5.061 ‰	3.001 ‰
24 Monate, Volldeckung	187 211	67 782	889	210	4.749 ‰	3.098 ‰
Total (Volldeckung)	3 617 857	1 551 302	24 222	5 659	6.695 ‰	3.648 ‰
12 Monate, Krankheitsdeckung	569 016	296 033	2 116	572	3.719 ‰	1.932 ‰
24 Monate, Krankheitsdeckung	2 598 153	1 065 388	7 280	1 697	2.802 ‰	1.593 ‰
Total (Krankheitsdeckung)	3 167 169	1 361 421	9 396	2 269	2.967 ‰	1.667 ‰
Total (deckungsunabhängig)	6 785 026	2 912 723	33 618	7 928	4.955 ‰	2.722 ‰

Tabelle 1: Datenmaterial

Aus diesen groben Masszahlen ist bereits ersichtlich, dass die Invalidierungshäufigkeiten für die Wartefrist 6 Monate unterhalb derjenigen mit Wartefrist 12 Monate liegen, was im Gegensatz zu (2) steht.

3 Bailey & Simon

3.1 Multiplikatives Modell

Das multiplikative Verfahren von Bailey & Simon geht vom Ansatz $i_{g,d,w,x} = \xi_g \cdot \zeta_d \cdot \gamma_w \cdot \lambda_x$ aus, wobei die einzelnen Faktoren durch Minimierung der Grösse Φ

$$\Phi = \sum_{g,d,w,x} \frac{B_{g,d,w,x} \cdot (\tilde{i}_{g,d,w,x} - \xi_g \cdot \zeta_d \cdot \gamma_w \cdot \lambda_x)^2}{\xi_g \cdot \zeta_d \cdot \gamma_w \cdot \lambda_x} \quad (5)$$

in Bezug auf $\xi_g, \zeta_d, \gamma_w, \lambda_x$ bestimmt werden.

Durch Nullsetzen der partiellen Ableitungen von Φ und anschliessendem Auflösen nach den gesuchten Faktoren gelangt man zu folgendem Gleichungssystem:

$$\xi_g = \left[\frac{\sum_{d,w,x} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{\zeta_d \cdot \gamma_w \cdot \lambda_x}}{\sum_{d,w,x} B_{g,d,w,x} \cdot \zeta_d \cdot \gamma_w \cdot \lambda_x} \right]^{1/2} \quad \text{für } g \in \{\text{Mann, Frau}\} \quad (6)$$

$$\zeta_d = \left[\frac{\sum_{g,w,x} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{\xi_g \cdot \gamma_w \cdot \lambda_x}}{\sum_{g,w,x} B_{g,d,w,x} \cdot \xi_g \cdot \gamma_w \cdot \lambda_x} \right]^{1/2} \quad \text{für } d \in \{K, U + K\} \quad (7)$$

$$\gamma_w = \left[\frac{\sum_{g,d,x} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{\xi_g \cdot \zeta_d \cdot \lambda_x}}{\sum_{g,d,x} B_{g,d,w,x} \cdot \xi_g \cdot \zeta_d \cdot \lambda_x} \right]^{1/2} \quad \text{für } w \in \{3, 6, 12, 24\} \quad (8)$$

$$\lambda_x = \left[\frac{\sum_{g,d,w} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{\xi_g \cdot \zeta_d \cdot \gamma_w}}{\sum_{g,d,w} B_{g,d,w,x} \cdot \xi_g \cdot \zeta_d \cdot \gamma_w} \right]^{1/2} \quad \text{für } x \in \{15, 16, \dots, 64\} \quad (9)$$

Das ergibt maximal 58 nichtlineare Gleichungen (2 Geschlechter, 2 Deckungen, 4 Wartefristen und 50 Alter) mit 58 Unbekannten. Da die Gleichungen nicht direkt lösbar sind, muss ein numerisches Iterationsverfahren zur Bestimmung eines Lösungsvektors angewandt werden:

- Zuerst werden die Startwerte festgelegt, indem die Randtotale der Invaliden durch die Randtotale der Bestände unter Risiko dividiert und als erste Schätzungen für ξ_g , ζ_d und γ_w festgelegt werden, also

$$\xi_g = \frac{\sum_{d,w,x} I_{g,d,w,x}}{\sum_{d,w,x} B_{g,d,w,x}}, \text{ etc.}$$
- Einsetzen dieser Werte in (9) ergibt dann eine erste Schätzung für λ_x .
- Mit dieser Schätzung für λ_x (zusammen mit den Schätzungen für ζ_d und γ_w) kann mittels Gleichung (6) eine nächste Schätzung für ξ_g berechnet werden. Analog werden mit den Gleichungen (7) und (8) neue Schätzungen für ζ_d resp. γ_w bestimmt.
- Dieser iterative Prozess wird solange fortgesetzt, bis sich die Differenz zweier Lösungsvektoren innerhalb einer vorgegebenen Genauigkeits-schranke befindet.

Bemerkung: ξ_g , ζ_d , γ_w und λ_x sind durch die Gleichungen (6) bis (9) nicht eindeutig bestimmt. Die Lösungen dieser Gleichungen sind nur bis auf multiplikative Konstanten gegeben. Hingegen ist das Produkt verschiedener Lösungsvektoren durch (6) bis (9) eindeutig bestimmt.

3.2 Additives Modell

Beim additiven Modell wird die Produktdarstellung von $i_{g,d,w,x}$ durch den additiven Ansatz $i_{g,d,w,x} = \xi_g + \zeta_d + \gamma_w + \lambda_x$ ersetzt:

$$\Phi = \sum_{g,d,w,x} \frac{B_{g,d,w,x} \cdot (\tilde{i}_{g,d,w,x} - (\xi_g + \zeta_d + \gamma_w + \lambda_x))^2}{\xi_g + \zeta_d + \gamma_w + \lambda_x} \quad (10)$$

Partielle Differentiation nach den gesuchten Grössen und anschliessendes Nullsetzen führt zu folgendem Gleichungssystem:

$$\sum_{d,w,x} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{(\xi_g + \zeta_d + \gamma_w + \lambda_x)^2} = \sum_{d,w,x} B_{g,d,w,x} \quad \text{für } g \in \{\text{Mann, Frau}\} \quad (11)$$

$$\sum_{g,w,x} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{(\xi_g + \zeta_d + \gamma_w + \lambda_x)^2} = \sum_{g,w,x} B_{g,d,w,x} \quad \text{für } d \in \{K, U + K\} \quad (12)$$

$$\sum_{g,d,x} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{(\xi_g + \zeta_d + \gamma_w + \lambda_x)^2} = \sum_{g,d,x} B_{g,d,w,x} \quad \text{für } w \in \{3, 6, 12, 24\} \quad (13)$$

$$\sum_{g,d,w} \frac{B_{g,d,w,x} \cdot \tilde{i}_{g,d,w,x}^2}{(\xi_g + \zeta_d + \gamma_w + \lambda_x)^2} = \sum_{g,d,w} B_{g,d,w,x} \quad \text{für } x \in \{15, 16, \dots, 64\} \quad (14)$$

Eine Lösung für dieses Gleichungssystems kann mit dem Newton-Verfahren durch einen iterativen Prozess gewonnen werden. Die Lösungen des Gleichungssystems sind allerdings nur bis auf additive Konstanten festgelegt. Die Summe verschiedener Lösungsvektoren ist dagegen durch (11) bis (14) eindeutig bestimmt.

3.3 Modellwahl

In der Praxis stellt sich natürlich die Frage, ob das additive oder das multiplikative Modell einem vorgegebenen Tarifierungsproblem besser gerecht wird. Diese Frage führt direkt zum Kern des Problems: “Welches ist die wahre (wenngleich Unbekannte) zugrundeliegende Risikostruktur?”. Obwohl sich schon verschiedene Arbeiten mit diesem Thema beschäftigt haben, existieren noch keine statistischen Tests, welche über die additive oder multiplikative Zusammensetzung der Daten Auskunft geben.

Für die Praxis bedeutet das, dass das günstigste Modell durch Ausprobieren bestimmt werden muss, falls nicht schon gewisse Zusammenhänge über den Aufbau der Risikodaten bekannt sind.

Anstelle der bereits behandelten multiplikativen und additiven Modelle kann man den Ansatz weiter verallgemeinern und erhält dann $i_{g,d,w,x} = f_{g,d,w,x}(\Theta)$ wobei die $f_{g,d,w,x}$ feste Funktionen sind und Θ ein Vektor von Parametern ist, welche geschätzt werden müssen.

Es gibt noch zwei verwandte Methoden, welche zur Schätzung der unbekannten Parameter in einem vorgegebenen Modell herangezogen werden können. Einerseits die Methode der Randtotale von Bailey und Jung und andererseits die Methode der kleinsten Quadrate von Sant (multiplikatives Modell) resp. Chang und Fairley (additives Modell). Nähere Angaben dazu sind in [4] p. 92ff zu finden. Für Schadenfrequenzen – wie in unserem Fall die Invalidierungshäufigkeiten – ist Bailey & Simon aus stochastischer Sicht zu bevorzugen (siehe ebenfalls in [4] p. 95ff). Für die Schätzung nicht binomialverteilter Zufallsvariablen, beispielsweise der mittleren Risikosumme im Invaliditätsfall, wären jedoch die anderen beiden erwähnten Methoden zu bevorzugen.

4 Diskussion der Ergebnisse

Zur besseren Interpretation der Resultate werden zuerst die wesentlichen Unterschiede des Bailey & Simon-Algorithmus gegenüber der heutzutage oft verwendeten kubischen Spline-Ausgleichung – unter besonderer Berücksichtigung der Aspekte der Lebensversicherung – dargestellt.

4.1 *Spline versus Bailey & Simon*

Im Gegensatz zur Nichtleben-Versicherung, wo man in der Regel mit diskreten Tarifklassen arbeitet, geht man in der Lebensversicherung üblicherweise von kontinuierlichen oder sogar stetig differenzierbaren Daten in Abhängigkeit des Alters, wie z.B. Sterblichkeit, Invalidierung, Wahrscheinlichkeit beim Tode verheiratet zu sein, etc., aus.

Dieser Aspekt wird bei der Approximation der Rohdaten durch die Spline-Ausgleichung optimal berücksichtigt, indem aus der Menge aller stetig differenzierbaren Funktionen über dem auszugleichenden Intervall diejenige Funktion bestimmt wird, welche einerseits den Abstand zu den Rohdaten minimiert und andererseits minimale Krümmung (im Sinne der zweiten Ableitung) aufweist. Dabei kann die Gewichtung zwischen minimaler Krümmung und minimalem Abstand durch geeignete Wahl der freien Parameter individuell bestimmt werden.

Im Unterschied dazu liefert das Bailey & Simon Verfahren Schätzungen für (diskrete) Tarifklassen. Das ist allerdings in der Praxis nicht weiter tragisch, da die Werte in der Regel in Tabellenwerken für ganzzahlige Alter festgehalten werden und Zwischenwerte durch Interpolation bestimmt werden.

Störender dagegen ist, dass die Berücksichtigung aller Messpunkte für die Schätzung der Parameter zwar einen “Glättungseffekt” auf die geschätzten Werte der einzelnen Tarifklassen zur Folge hat, dieser jedoch deutlich weniger ausgeprägt ist, als etwa bei der Spline-Ausgleichung. Das führt, insbesondere bei einjährigen Tarifen, zu zufälligen Schwankungen der Prämiensätze für benachbarte Alter.

Beide Effekte können beseitigt werden, indem der durch Bailey & Simon bestimmte Parameter für das Alter zusätzlich Spline ausgeglichen wird. Dadurch produziert man ausgeglichene stetig differenzierbare Kurven für alle Ausprägungen der Merkmale Geschlecht, Wartefrist und Deckung.

Ein charakteristischer Unterschied zwischen den beiden diskutierten Ausgleichsverfahren besteht auch bezüglich der Behandlung der verschiedenen Merkmale. Während bei der Bailey & Simon Ausgleichung ein mathematischer Zusammenhang (Modell) zwischen den einzelnen Merkmalen angenommen wird (im einfachsten Fall multiplikativer oder additiver Art), werden bei der Spline-Ausgleichung die Rohwerte jeweils für ein festes Tupel von Ausprägungen in Abhängigkeit des Alters ausgeglichen (ohne Berücksichtigung der anderen gemessenen Daten), d. h. es wird kein Zusammenhang zwischen den einzelnen Merkmalen vorausgesetzt oder angenommen. Als Folge des einfachen Modellansatzes beim Bailey & Simon-Verfahren lassen sich die ausgeglichenen Tarifwerte aus wenigen Faktoren resp. Summanden berechnen.

4.2 *Ergebnisse*

Um die Leistungsfähigkeit des Bailey & Simon-Algorithmus, aber auch seine Grenzen, zu demonstrieren, sind die vorhandenen Rohdaten im wesentlichen für alle Merkmale und deren Ausprägungen als Ganzes ausgeglichen worden. Einzig der Altersbereich ist aus Plausibilitätsgründen auf die Alter zwischen 22 und 58 eingeschränkt worden. Dagegen sind die Rohdaten für die Wartefrist 6 Monate in die Ausgleichung einbezogen worden, obwohl sie in Bezug auf das verwendete Invaliditätsmodell unplausibel sind.

4.2.1 Glattheit der Approximation

Als Mass zur quantitativen Erfassung der Glätte der Ausgleichskurven verwenden wir die Quadratsumme der doppelten Differenzen (die Schrittweite ist im vorliegenden Fall 1). Da die Differenzen höherer Ordnung bei einer glatten Kurve gewöhnlich klein sind, ist dieses Mass gut geeignet zur Beurteilung der Glattheit der Ausgleichskurven:

Art der Ausgleichung	Quadratsumme der doppelten Differenzen
Rohwerte (keine Ausgleichung)	$4.27 \cdot 10^{-4}$
Spline-Ausgleichung	$3.01 \cdot 10^{-6}$
Bailey & Simon-Ausgleichung	$2.50 \cdot 10^{-5}$
Bailey & Simon mit nachträglicher Spline-Ausgleichung	$1.03 \cdot 10^{-6}$

4.2.2 Güte der Approximation

Als messbare Kriterien für die Güte der Ausgleichung verwenden wir einerseits das Verhältnis V der gemessenen Invaliden zu den aufgrund der Ausgleichung erwarteten Invaliden (bei gleichem Bestand unter Risiko) pro Merkmal, andererseits die Summe Q der normierten Abstandsquadrate zwischen den Schätzungen und den Rohwerten. Mit den aufgrund des Ausgleichsverfahrens geschätzten Invalidierungshäufigkeiten $\hat{i}_{g,d,w,x}$ lässt sich V formelmässig, z.B. für die Ausprägung "Männer" des Merkmals Geschlecht, darstellen als:

$$V(\text{Mann}) = \frac{\sum_{\text{Mann},d,w,x} I_{g,d,w,x}}{\sum_{\text{Mann},d,w,x} (\hat{i}_{g,d,w,x} \cdot B_{g,d,w,x})} \quad (15)$$

Die Definition von V für andere Merkmalsausprägungen erfolgt analog. Q ist festgelegt als:

$$Q = \sum_{g,d,w,x} \frac{(I_{g,d,w,x} - \hat{i}_{g,d,w,x} \cdot B_{g,d,w,x})^2}{\hat{i}_{g,d,w,x} \cdot B_{g,d,w,x}} \quad (16)$$

Geschlecht	V	Deckung	V
Männer	98.7 %	Volldeckung	99.0 %
Frauen	97.4 %	Krankheitsdeckung	96.9 %

Wartefrist	V	Alter	V
3 Monate	99.3 %	25	94.2 %
6 Monate	98.3 %	30	96.7 %
12 Monate	97.1 %	35	97.6 %
24 Monate	96.6 %	40	98.5 %
		45	99.6 %
		50	99.6 %
		55	98.6 %

Im Total über alle Merkmalsausprägungen beträgt das Verhältnis $V = 98.4\%$ (zum Vergleich: $V = 99,2\%$ für die Spline-ausgeglichenen Invalidierungshäufigkeiten). Die Bailey & Simon Schätzungen, differenziert nach Merkmalsausprägungen, liegen immer zwischen 0 % bis 3.5 % zu hoch³; grössere Abweichungen ergeben sich einzig für junge Alter, z. B. 5.8 % im Alter 25. Aufgrund der Verhältniszahl V über das Total scheinen die Unterschiede zwischen der Spline-Ausgleichung und dem Bailey & Simon Algorithmus ziemlich gering zu sein. Die Masszahl Q dagegen zeigt gravierendere Abweichungen auf.

Ausgleichsmethode	Q
Spline-Ausgleichung	165
Bailey & Simon-Ausgleichung ⁴	1058
Bailey & Simon mit nachträglicher Spline-Ausgleichung	1073

³ Es kann gezeigt werden, dass die Bailey & Simon Schätzer je Merkmalsausprägung immer zu hoch liegen.

⁴ Die Ausgleichskurven für Männer, Wartefrist 24 Monate mit Krankheitsdeckung und Frauen, Wartefrist 3 Monate mit Volldeckung tragen einen wesentlichen Anteil an Q bei, denn

$$\sum_{\substack{\text{Mann, } K, 24, x \\ \text{Frau, } U + K, 3, x}} \frac{(I_{g,d,w,x} - \hat{i}_{g,d,w,x} \cdot B_{g,d,w,x})^2}{\hat{i}_{g,d,w,x} \cdot B_{g,d,w,x}} = 602$$

Die Diskrepanz zwischen V und Q kommt dadurch zustande, dass sich positive und negative Differenzen zwischen den gemessenen und den geschätzten Werten bei der Bildung von V aufheben können, während sie bei der Berechnung von Q mit dem Quadrat eingehen.

Um das Gesamtbild zu vervollständigen sind die Resultate zusätzlich grafisch dargestellt worden, wobei zur Beurteilung der Abweichungen der ausgeglichenen Invalidierungshäufigkeiten von den Rohwerten die 95 %-Konfidenzintervalle für die Rohwerte der Wartefristen 3 und 24 Monate angegeben sind (Grafik 1, p. 152/153):

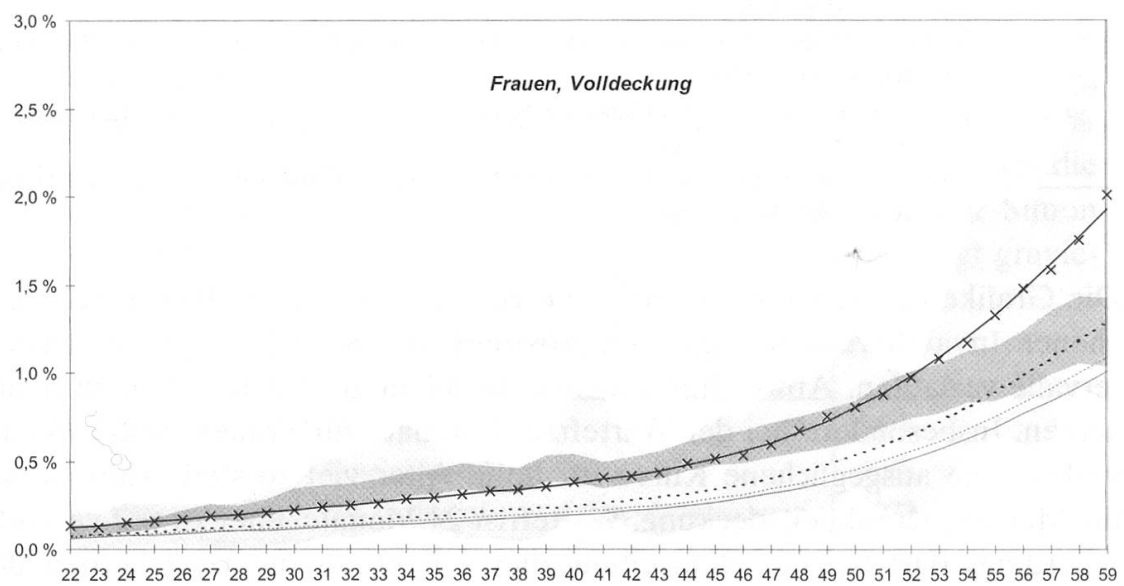
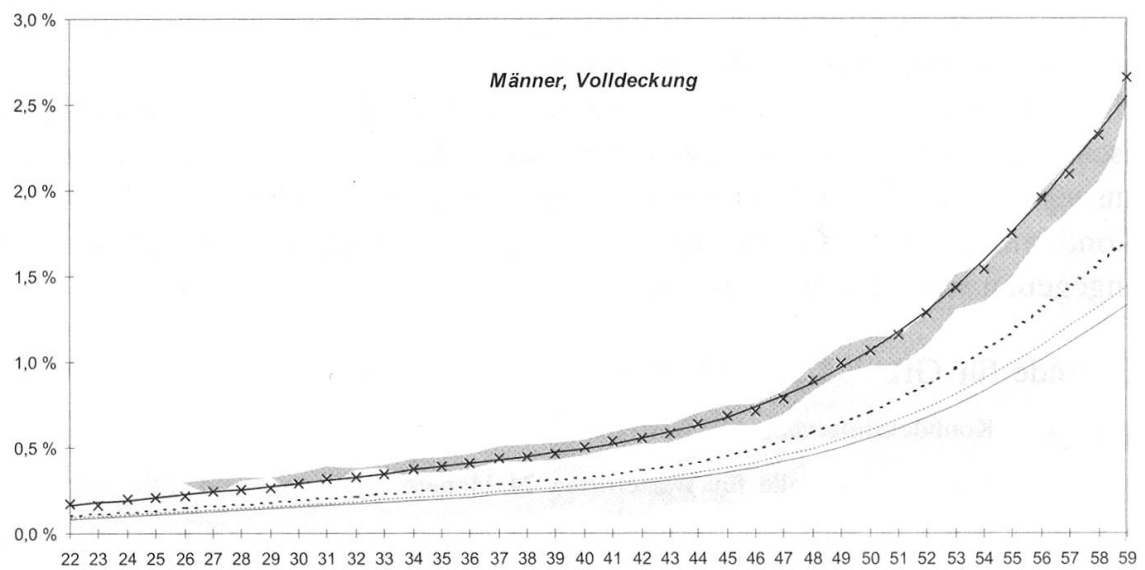
Legende für Grafik 1 (p. 152/153):

	Konfidenzintervalle für Wartefristen 3 Monate
	Konfidenzintervalle für Wartefristen 24 Monate
× × ×	Bailey & Simon ausgeglichene Invalidierungswahrscheinlichkeiten, Wartefrist 3 Monate
————	Bailey & Simon und anschliessend Spline ausgeglichene Invalidierungshäufigkeiten, Wartefrist 3 Monate
-----	Bailey & Simon und anschliessend Spline ausgeglichene Invalidierungshäufigkeiten, Wartefrist 6 Monate
.....	Bailey & Simon und anschliessend Spline ausgeglichene Invalidierungshäufigkeiten, Wartefrist 12 Monate
————	Bailey & Simon und anschliessend Spline ausgeglichene Invalidierungshäufigkeiten, Wartefrist 24 Monate

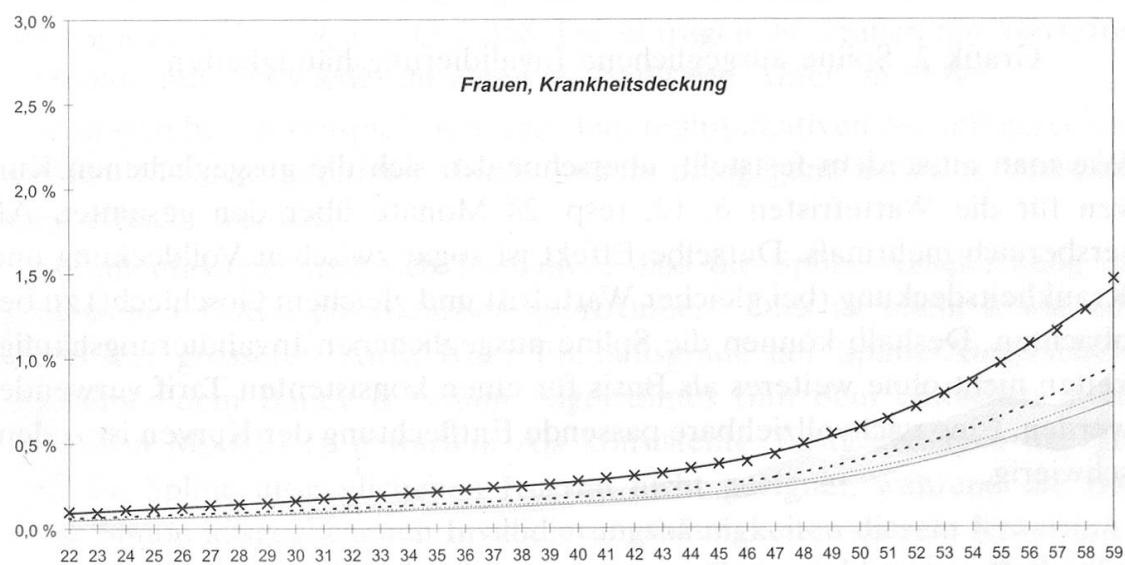
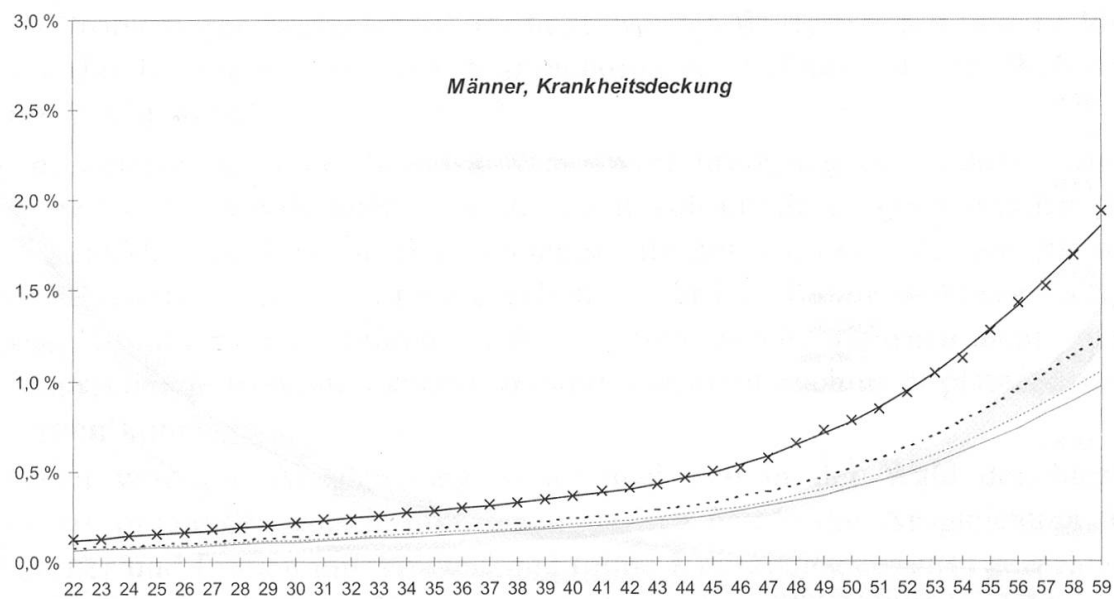
Die Grafiken zeigen, dass die mit dem Bailey & Simon Verfahren ausgeglichenen Invalidierungshäufigkeiten grösstenteils innerhalb der Konfidenzintervalle verlaufen. Abweichungen sind vor allem an den Rändern zu beobachten. Insbesondere bei der Wartefrist 3 Monate für Frauen, Volldeckung, verläuft die ausgeglichene Kurve für hohe Alter viel zu steil, während sie für Männer, Krankheitsdeckung, Wartefrist 24 Monate, eher zu tief verläuft. Allerdings darf bei der Beurteilung dieses Faktums die Unsicherheit des Datenmaterials in den Randbereichen nicht vernachlässigt werden.

Im Weiteren ist ersichtlich, dass die Invalidierungshäufigkeiten der Männer für Wartefrist 6 Monate mit Volldeckung unterhalb derjenigen mit Wartefrist 12 Monate verlaufen. Aufgrund der durchschnittlichen Invalidierungshäufigkeiten aus Tabelle 1 für den Altersbereich 15–64 ist das jedoch nicht überraschend.

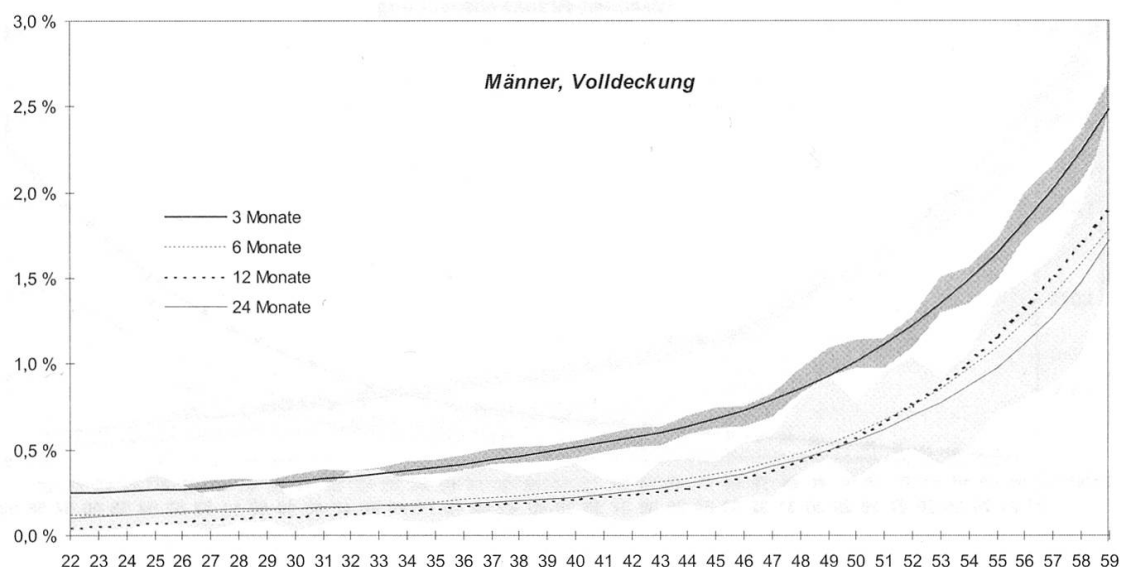
Grafik 1: Bailey & Simon und anschliessend Spline ausgeglichene Invalidierungshäufigkeiten



Grafik 1: Fortsetzung



Dasselbe Phänomen tritt ebenfalls – wenn auch nur im oberen Altersbereich – bei den Spline-ausgeglichenen Invalidierungshäufigkeiten auf:



Grafik 2: Spline ausgeglichene Invalidierungshäufigkeiten

Wie man ausserdem feststellt, überschneiden sich die ausgeglichenen Kurven für die Wartezeiten 6, 12, resp. 24 Monate über den gesamten Altersbereich mehrmals. Derselbe Effekt ist sogar zwischen Volldeckung und Krankheitsdeckung (bei gleicher Wartezeit und gleichem Geschlecht) zu beobachten. Deshalb können die Spline ausgeglichenen Invalidierungshäufigkeiten nicht ohne weiteres als Basis für einen konsistenten Tarif verwendet werden. Eine nachvollziehbare passende Entflechtung der Kurven ist zudem schwierig.

Beim Algorithmus von Bailey & Simon sind solche Überschneidungen schon aufgrund des Ansatzes von vornherein ausgeschlossen. Zudem kann das Problem der “verkehrten” Reihenfolge der Wartefristen 6 und 12 Monate durch einfache, passende Manipulation der Faktoren für die Wartefristen gelöst werden.

Eine weitere Stärke des Bailey & Simon Verfahrens liegt darin, dass es auch Werte für Merkmale liefert, welche nicht vollständig erhoben wurden. So ist Krankheitsdeckung in der Gemeinschaftsstatistik des VPL nur für die Wartefristen 12 und 24 Monate erhoben worden. Bailey & Simon liefert auch Resultate für die fehlenden Wartefristen, welche natürlich nicht durch entsprechende Rohdaten abgestützt sind und somit auch nicht plausibilisiert werden können.

Besser wird die Ausgleichung, wenn man sich in der Wahl der Merkmalsausprägungen einschränkt. Exemplarisch wurde die Ausgleichung für Männer und Frauen mit Volldeckung (ohne Krankheitsdeckung) gerechnet. Dadurch wurde die Ausgleichung – insbesondere für Wartefrist 3 Monate für Männer – doch um einiges genauer, was sich auch in den Masszahlen widerspiegelt: $V = 99,3\%$, $Q = 358$. Dabei tragen die Frauen mit Wartefrist 3 Monate mit 138 wiederum einen wesentlichen Anteil zu Q bei.

Alle angegebenen Beispiele sind mit dem multiplikativen Modell gerechnet worden. Mit dem additiven Modell konnte demgegenüber keine Verbesserung erreicht werden.

Zusammenfassend lässt sich festhalten, dass die Spline-Ausgleichung die gemessenen Daten etwas besser approximiert. Dies ist allein schon aufgrund der grösseren Anzahl der Freiheitsgrade der Spline-Ausgleichung gegenüber dem Bailey & Simon Algorithmus (mit dem einfachen multiplikativen Modell) zu erwarten. Als konsistente Tarifgrundlagen dagegen sind die Spline ausgeglichenen Kurven nicht geeignet, während die Bailey & Simon ausgeglichenen Invalidierungshäufigkeiten diesem Kriterium – abgesehen von der schon in den gemessenen Daten inhärenten Reihenfolgeproblematik – genügen.

5 Schlussbemerkungen

Da die Differenz $(\tilde{i}_{g,d,w,x} - \xi_g \cdot \zeta_d \cdot \gamma_w \cdot \lambda_x)$ im Ansatz für das Bailey & Simon Verfahren mit dem Quadrat in die zu minimierende Gleichung eingeht,

ist die Schätzung nicht allzu resistent gegen Ausreisser. Es ist deshalb zu empfehlen, die Rohdaten zuerst von Ausreissern zu bereinigen.

Wie die Resultate zeigen, wird die multiplikative Struktur dem untersuchten Datenmaterial gut gerecht. Dies muss keinesfalls immer zutreffen.

Abgesehen von der Problematik der Reihenfolge und der Überschneidungen liegen die rohen Invalidierungshäufigkeiten so nahe beisammen, dass sie sich durch 95 % Konfidenzintervalle nicht eindeutig trennen lassen. Insofern ist jede Ausgleichung dieser Daten problematisch.

Bibliographie

- [1] Bailey R. A. and Simon L. J. [1960]: Two studies in automobile insurance rate making, ASTIN Bulletin, Vol. 1, pp 192–217
- [2] Böhmer K. [1974]: Spline Funktionen, Teubner Verlag Stuttgart
- [3] Technische Kommission VPL [1993]: Cinquième rapport concernant la statistique 1986/90 – Bases techniques de deuxième ordre pour l'invalidité dans l'assurance collective, Zirkular Nr. 23/1993
- [4] Van Eeghen J, Greup E. K. and Nijssen J. A. [1983]: Rate Making, Surveys of actuarial studies no 2, Nationale-Nederlanden N. V., Rotterdam, pp 88–98
- [5] VPL [1994]: Tarif 1995, Kollektivlebensversicherung

Willi Thurnherr
Aktuariat
Basler Lebens-Versicherungs-Gesellschaft
Aeschengraben 21
CH-4002 Basel

Zusammenfassung

Mit dem Parameterschätzungsverfahren von Bailey & Simon werden rohe Invalidierungshäufigkeiten, welche von den Merkmalen Alter, Geschlecht, Wartefrist und Deckung abhängen, ausgeglichen. Die Resultate werden den Spline-ausgeglichenen Invalidierungshäufigkeiten gegenübergestellt. Die Ergebnisse der Ausgleichungen werden anhand verschiedener Kriterien (Konfidenzintervalle, Masszahlen zur Bestimmung der Glattheit und der Qualität der Approximation) unter Berücksichtigung der speziellen Aspekte der Lebensversicherung beurteilt, wobei auch deren Eignung als Basis für einen in sich geschlossenen (konsistenten) Invaliditätstarif mitberücksichtigt wird.

Résumé

La méthode d'estimation des paramètres de Bailey & Simon est appliquée pour ajuster les taux d'entrée en invalidité observés: ces taux dépendent des facteurs âge, sexe, délai d'attente et couverture. Les résultats sont comparés avec les taux obtenus en appliquant la méthode des splines. Les résultats des deux méthodes sont examinés à l'aide de différents critères (intervalles de confiance, mesures de lissage, qualité de l'approximation) en considérant les aspects spéciaux de l'assurance-vie et tout en gardant en mémoire que les taux d'entrée en invalidité ajustés doivent servir de base à un tarif en assurance invalidité consistant.

Summary

The method of Bailey & Simon for parameter estimation is applied to the observed disability rates depending on age, sex, waiting period and coverage. The estimations are compared with the results obtained by smoothing the same data with the spline algorithm. The effectiveness of the two methods is examined by using appropriate criterias like confidence intervalls, measures for smoothness and the quality of the approximation, taking into consideration the special aspects of life insurance and keeping in mind the use of the smoothed disability rates as a base for a consistent tariff.

