

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK

INSTITUT FÜR MATHEMATISCHE STOCHASTIK

MASTERARBEIT

Modellbildung und Trendschätzung bei Sterbetafeln

Autor:

Rolf Tobias Hajo Henrik

Henning Hause

Abgabetermin:

01.04.2020

Studiengang:

M. Sc. Mathematik

Erstgutachter:

Dr. Frank Werner

Zweitgutachter:

Prof. Dr. Stephan

Huckemann

Inhaltsverzeichnis

1	Motivation und Aufgabenstellung	3
1.1	Verwendung von Sterbetafeln	3
1.2	Ziele und Aufbau der Arbeit	4
2	Definitionen und historische Daten	6
2.1	Rohe Sterbehäufigkeiten und Sterbetafeln	6
2.2	Probleme beim Schätzen roher Sterbewahrscheinlichkeiten	9
2.3	Glättung von Sterbewahrscheinlichkeiten	10
2.3.1	Überblick über Glättungsverfahren	11
2.3.2	Das Verfahren von Whittaker-Henderson	14
2.4	Öffentlich zugängliche Sterblichkeitsdaten	16
3	Modellbildung	20
3.1	Ein einfaches Modell: Geburtsjahr als erklärende Variable	20
3.2	Ein einfaches Modell: Alter als erklärende Variable	22
3.3	Das Vorgehen der DAV	25
3.4	Das Lee-Carter-Modell	27
3.5	Verbesserung des einfachen Modells	35
4	Eigenschaften der Modelle und Simulationen	37
4.1	Das Theorem von Cramér	37
4.2	Eigenschaften des einfachen Modells	42
4.3	Eigenschaften des Lee-Carter-Modells	59
4.3.1	Der Schätzer für β_x	59
4.3.2	Schätzer für ν	64
4.4	Eigenschaften des einfachen Trendmodells	67
4.5	Qualitätsvergleich durch Simulation	68
4.5.1	Simulationen des einfachen Modells	69
4.5.2	Simulationen zum Lee-Carter-Modell	72
5	Zusammenfassung	81

Abbildungsverzeichnis

1	Veränderte Todeszahlen in der Bundesrepublik Deutschland 1190 vs. 2017.	4
2	Jahreszeitliche Schwankung von Sterblichkeiten.	10
3	Glättung von Daten mittels des Whittaker-Henderson-Verfahrens.	16
4	Geburtsjahre versus mittleres erreichtes Alter.	21
5	Eine Heatmap der Todesfälle der Generationstafel 1876 West.	22
6	Das Lee-Carter-Modell für die Generationstafel 1876 Westdeutschland. . .	23
7	Alter versus die Sterblichkeit bei dem einfachen Modell.	24
8	Lee-Carter $\exp(\alpha)$	31
9	Lee-Carter β	32
10	Lee-Carter γ	33
11	Heatmap für (γ_t, β_x)	34
12	$\gamma_t \cdot \beta_x$ mit $t \in \{1956, 2017\}$	35
13	Man sieht, $\hat{\mu}_t$ gegen die Geburtsjahre aufgetragen.	36
14	Die Funktion d/dm der Log-Likelihoodfunktion.	47
15	Die Funktion d/dm der Log-Likelihoodfunktion ohne die Randwerte. . . .	48
16	Die Funktion d/ds^2 der Log-Likelihoodfunktion.	50
17	Die Funktion d/ds^2 der Log-Likelihoodfunktion im Bereich 4 bis 10. . . .	51
18	Vorhersage des einfachen Modells für den historischen Datensatz.	69
19	Ein Beispiel der Simulationsdaten für das einfache Modell.	71
20	Fehler der Simulation im Detail.	72
21	Das Lee-Carter-Modell angewendet auf den historischen Datensatz.	73
22	Die logarithmierten Werte wie in Abbildung 21.	74
23	Parameter des Lee-Carter-Modells.	75
24	Das Lee-Carter-Modell mit Beispielparametern.	76
25	Histogramme der Fehler in der Lee-Carter Simulation.	77
26	Das Lee-Carter-Modell angewendet auf den simulierten Datensatz.	78
27	Die logarithmierten Werte wie in Abbildung 26.	79
28	Der γ -Wert aus dem historischen Datensatz.	80
29	Konfidenzintervall für ν	82

1 Motivation und Aufgabenstellung

Die Idee für diese Arbeit ist entstanden, während ich im Winter 2018/2019 ein Praktikum bei einer Lebensversicherung machte. Dort habe ich eine Programmbibliothek geschrieben, die aus einer Sterbetafel versicherungsmathematische Kennzahlen bestimmt. Da ich einige Statistikvorlesungen gehört habe und Sterblichkeiten zufällig sind, hat sich mir die Frage gestellt, wie man Sterbetafeln aufstellt. Diese Arbeit fasst übliche Methoden zusammen und berechnet einige ihrer Eigenschaften. Dazu beginnen wir in Abschnitt 1.1 damit den Begriff der Sterbetafel durch seine Anwendungsgebiete zu motivieren. Dann werden in Abschnitt 1.2 die Ziele dieser Arbeit genauer diskutiert und der Aufbau der Arbeit angegeben.

1.1 Verwendung von Sterbetafeln

Sterbetafeln werden vor allem in der Lebensversicherungswirtschaft und in der Demografieforschung verwendet. Wir betrachten beide Bereiche getrennt.

Beginnen wir mit der Verwendung von Sterbetafeln in der Lebensversicherungswirtschaft. Um ein Versicherungsprodukt in der Lebensversicherungswirtschaft zu konstruieren, benötigt man die drei Rechnungsgrundlagen Rechenzins, biometrische Risiken und Kostensätze. Rechenzins und Kostensätze sind normalerweise gegeben, während man die biometrischen Risiken schätzen muss. Es handelt sich um eine Reihe von Risiken, die normalerweise die körperliche Unversehrtheit beeinträchtigen. Beispiele hierfür sind vor allem Tod und Invalidität, aber man kann auch das Risiko zu heiraten zu diesen Risiken zählen. Man spricht dann von der Aussteuerversicherung. Ein einfaches Beispiel für ein Netto-Versicherungsprodukt, das heißt ein Versicherungsprodukt ohne Kosten, ist die Einmalprämie einer n -jährigen Erlebensfallversicherung mit Versicherungssumme eine Geldeinheit für eine Person mit Alter x . Der Wert dieses Produktes kann über die Gleichung (1) beschrieben werden. Dabei bezeichnet l_x die Anzahl der Lebenden im Alter x und i ist der Rechenzins.

$${}_nE_x = \frac{l_{x+n}}{l_x} \left(\frac{1}{1+i} \right)^n \quad (1)$$

Es sei noch erwähnt, dass man l_x normalerweise mittels $l_{x+1} = l_x \cdot p_x = l_x \cdot (1 - q_x)$ berechnet. Hier ist q_x die Wahrscheinlichkeit, dass eine Person, die bereits x Jahre alt geworden ist, vor ihrem $x + 1$ -ten Geburtstag verstirbt. Man nennt q_x eine 1-jährige Sterbewahrscheinlichkeit und wir werden in dieser Arbeit Methoden sehen, q_x zu schätzen. Eine Sterbetafel ist dann eine Tabelle, die 1-jährige Sterbewahrscheinlichkeiten q_x aufführt.

Betrachten wir jetzt noch kurz das andere Anwendungsgebiet für Sterbetafeln, die Demografieforschung. Laut der Website [Far13] beschäftigt sich die Demografieforschung zum Beispiel mit Problemen der Gesundheitspolitik, Beschäftigungspolitik, Bildungspolitik und Migrationspolitik. In all diesen Bereichen ist die Struktur einer Bevölkerung wichtig, um politische Entscheidungen zu treffen. Ein anderer Bereich, der auf der Website [Far13] nicht erwähnt wird, aber trotzdem relevant ist, ist die staatliche Rentenpolitik.

1.2 Ziele und Aufbau der Arbeit

Im Abschnitt 1.1 wurde bei der Kalkulation von Versicherungsprodukten die Größe q_x verwendet. Wenn die Zufallsvariable X das Todesalter eines zufälligen Menschen bezeichnet, ist q_x durch folgende Gleichung gegeben:

$$q_x = \mathbb{P}(X \leq x + 1 | X > x)$$

Die Größe q_x wird durch die sogenannte einfache beziehungsweise rohe Sterbehäufigkeit angenähert. Die notwendige Notation für dieses Vorgehen wird in Kapitel 2 in Abschnitt 2.1 angegeben. Da Sterbetafeln nicht losgelöst von der Anwendung stehen, werden in Abschnitt 2.2 Probleme bei ihrer Erstellung diskutiert. Ein erster Schritt zur Verbesserung der Vorhersagekraft von Sterbetafeln sind Glättungsverfahren, die in Kapitel 2.3 vorgestellt werden. Abschließend für dieses Kapitel werden in Abschnitt 2.4 historische Daten und ihre Qualität diskutiert. Die historischen Daten stammen von der Website [Mor18]. Für Kapitel 2 wird hauptsächlich das Buch [KLO+85] verwendet. Außerdem wurden [NS12] von Nocon und Scott, sowie [Loe94] von Loebus verwendet.

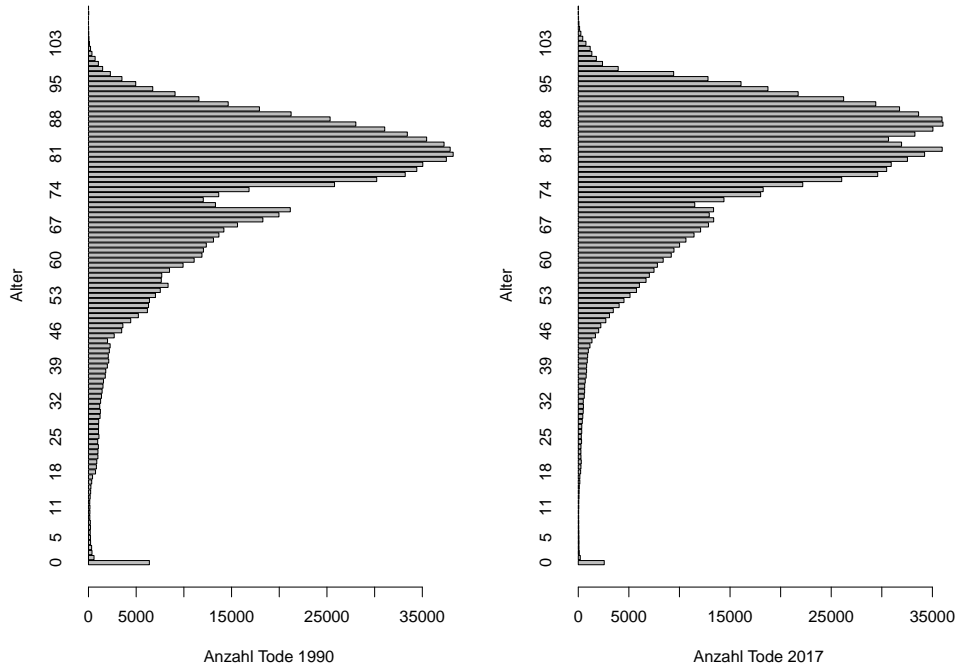


Abbildung 1: Anzahl der Tode in der Bundesrepublik Deutschland. Auf der linken Seite ist die Situation im Jahr 1990 und auf der rechten Seite die Situation im Jahr 2017 zu sehen.

Ein Problem bei der Erstellung von Sterbetafeln ist, dass sich die Lebenserwartung in den letzten Jahrzehnten erhöht hat. Um diesen Sachverhalt zu veranschaulichen sind in

Abbildung 1 je ein Säulendiagramme der Anzahl der Todesfällen der Jahre 1990 und 2017 in der Bundesrepublik Deutschland gegeben. Beim Vergleich der beiden Diagramme sieht man, dass im Jahr 2017 im Gegensatz zum Jahr 1990 das Diagramm für das Jahr 2017 nach oben, zu höheren Altern, verschoben ist. Daraus kann man folgern, dass das mittlere Todesalter steigt. Diese Erhöhung der Lebenserwartung nennt man Sterbetrends.

Um diese Sterblichkeitstrends in den Sterbetafeln abzubilden, werden Modelle der Form (2) angelegt. Hier bezeichnet f eine normalerweise bekannte Funktion, θ bezeichnet unbekannte Parameter und ϵ einen Fehler. Solche Modelle werden in Kapitel 3 betrachtet.

$$q_x = f_\theta(x) + \epsilon \quad (2)$$

Zuerst werden in 3.1 und 3.2 einfache Modelle eingeführt, die zum Vergleich mit dem üblicherweise verwendeten Lee-Carter-Modell dienen sollen. Anschließend wird in Abschnitt 3.3 diskutiert, wie die Deutsche Aktuarvereinigung (DAV) Sterbetafeln erstellt. Danach wird das Lee-Carter-Modell diskutiert. Für dieses Modell wird die Originalquelle [LC92] von Lee und Carter sowie [GK07] von Girosi und King verwendet. Am Ende des Kapitels 3 wird dann eine Verbesserung für die einfachen Modelle vorgeschlagen.

In Kapitel 4 wird in den Kapiteln 4.1, 4.2 und 4.3 für das einfache Modell und das Lee-Carter-Modell Identifizierbarkeit, Konsistenz und asymptotische Normalität gezeigt. Dies wird mit dem Satz von Cramér und einem aus dem Buch [Vaa98] von A. W. van der Vaart konstruierten Satz gezeigt. Konkret wird also für die von der Anzahl der Beobachtungen n abhängigen Schätzer $\hat{\theta}_n$ der Parameter θ_0 die folgende Eigenschaft gezeigt:

$$\begin{aligned} \hat{\theta}_n &\xrightarrow{P} \theta_0 \text{ für } n \rightarrow \infty \\ \sqrt{n}(\hat{\theta}_n - \theta_0) &\xrightarrow{D} Z \sim N(0, i(\theta_0)^{-1}) \text{ für } n \rightarrow \infty \end{aligned}$$

Hier bezeichnet $i(\theta_0)$ die Fisher-Informationsmatrix, die ebenfalls berechnet wird. Mit dieser Eigenschaft wird außerdem ein Konfidenzintervall für den Trendparameter ν berechnet:

$$1 - \alpha = \mathbb{P} \left(\nu \in \left[\hat{\nu} - z_{1-\alpha/2} \cdot \frac{\frac{(\hat{\gamma}_E - \hat{\gamma}_S)}{2\sigma_\xi^2}}{\sqrt{n}}, \hat{\nu} + z_{1-\alpha/2} \cdot \frac{\frac{(\hat{\gamma}_E - \hat{\gamma}_S)}{2\sigma_\xi^2}}{\sqrt{n}} \right] \right)$$

Abschließend wird in Abschnitt 4.5 der L_2 -Fehler der Modelle mittels Simulationen angenähert. Insgesamt erhält man beim Vergleich der Simulation das Ergebnis, dass es besser scheint, das Lee-Carter-Modell als das einfache Modell zu verwenden, da die Fehler geringer sind.

Die Simulationen in Kapitel 4 sowie fast alle Abbildungen in dieser Arbeit wurden in **R** erstellt und können auf meiner Github-Seite unter <https://github.com/fake1884/moRtRe> eingesehen werden. Der Name des Pakets “moRtRe” steht für “Mortality Trends” mit großen “R”s für die Programmiersprache **R**.

2 Definitionen und historische Daten

In diesem Kapitel wird in Abschnitt 2.1 der zentrale Begriff der Sterbetafel definiert. In dem nächsten Abschnitt 2.2 werden Probleme bei seiner Bestimmung angegeben. Diese ersten beiden Abschnitte orientieren sich an dem Buch [KLO+85]. Im daran anschließenden Kapitel 2.3 werden Verfahren zur Glättung von rohen Sterbehäufigkeiten vorgestellt. Im letzten Abschnitt 2.4 dieses Kapitels werden Daten vorgestellt, auf deren Grundlage Sterbetafeln erstellt werden. Insbesondere wird auf die öffentlich zugänglichen Sterbetafeln der Website [Mor18] eingegangen.

2.1 Rohe Sterbehäufigkeiten und Sterbetafeln

Da das Ziel dieser Arbeit das Schätzen von Sterbetafeln ist, müssen wir Notation einführen, um diesen Begriff definieren zu können. Der Startpunkt ist der Begriff des Kollektivs. Die folgende Definition ist [KLO+85, S. 12] entnommen. Es wurde nur der Begriff Personengesamtheit durch den kürzeren Begriff Kollektiv ersetzt, der in der Literatur ebenfalls Verwendung findet.

Definition 1 (Kollektive). Mit K bezeichnen wir eine Menge von Personen. Steht aufgrund der Definition dieser Menge K eindeutig fest, ob eine Person zu K gehört oder nicht, dann nennen wir K ein Kollektiv.

Die Definition des Kollektivs ist bewusst offen gehalten, da man dem Anwender die Möglichkeit lassen möchte beliebige Kollektive zu betrachten. In dieser Arbeit werden westdeutsche Personen betrachtet. Es werden keine gesamtdeutschen Daten betrachtet, da in der DDR Daten zum Teil anders erfasst wurden. Für weitere Informationen hierzu siehe Abschnitt 2.2. Es soll allerdings auch möglich sein mit den in dieser Arbeit vorgestellten Methoden Personen anderer Nationalität zu betrachten oder ganz andere Kollektive. Ein Beispiel hierfür ist die Sterblichkeit von Invaliden einer bestimmten Nationalität. Man kann Kollektive auch durch das Geschlecht der Personen definieren oder besonders riskante Gruppen wie Raucher betrachten.

Kollektive werden dann analog zu [KLO+85, S. 13] in bestimmte Risikoklassen K_\star unterteilt. Hier ist \star ein Platzhaltersymbol. Die Risikoklassen, die uns in dieser Arbeit interessieren sind das Alter x und das Geburtsjahr g der Personen sowie das Beobachtungsjahr t .

Aus dem Geburtsjahres g kann das Alter x in einem Beobachtungsjahr t bestimmen. Genauer gesagt ist das Alter das Beobachtungsjahr minus das Geburtsjahr. Dieses Vorgehen ist analog zu [KLO+85, S. 14]. Genauso kann man aus dem Todesjahr t und dem Geburtsjahr g direkt das Todesalter y bestimmen. Das Alter einer Person ändert sich somit zu Beginn eines neuen Beobachtungsjahres.

Betrachten wir jetzt konkrete Beispiele für Risikoklassen. In Abschnitt 3.1 wird nur die Risikoklasse Geburtsjahre verwendet. Das heißt, es ist $\star = g \in \{1878, \dots, 1987\}$, während in Abschnitt 3.2 nur die Risikoklasse Alter, das heißt $\star = x \in \{0, \dots, 95\}$, Verwendung

finden. In Abschnitt 3.4 wird das kartesische Produkt aus Alter $x \in \{0, \dots, 95\}$ und Beobachtungsjahr $t \in \{1956, \dots, 2017\}$ verwendet.

Die Risikoklassen sind immer derart gewählt, dass man das gesamte Kollektiv K als eine disjunkte Vereinigung über alle Risikoklassen erhält. Zum Beispiel fordert man für das kartesische Produkt aus Alter und Beobachtungsjahr $K = \cup_t \cup_x K_{x,t}$. Wir bezeichnen die Kardinalität einer Risikoklasse K_\star mit k_\star und die Kardinalität des Kollektivs K mit k .

Es kann für einige Anwendungen interessant sein, Risikoklassen erst ab einem bestimmten Alter beginnen zu lassen. Außerdem sind Daten nicht bis zu beliebig hohen Altern in guter Qualität vorhanden. Deshalb wird ein Start- und ein Endalter gesetzt. Für eine genauere Diskussion der Datenlage bei hohen Altern siehe Abschnitt 2.4 und insbesondere Tabelle 2.

Definition 2 (Start- und Endalter). Das Startalter S ist $S = 0$ und das Endalter E ist $E = 95$.

Aus einer Risikoklasse K_\star treten Personen aus verschiedenen Gründen ein und aus. Wir betrachten als Eintrittsgrund nur die Geburt, während Einwanderung vernachlässigt wird. Der einzige zugelassene Austrittsgrund ist durch Tod. Andere Austrittsgründe können Invalidität oder Auswanderung sein. Falls aus einem Kollektiv Personen nur durch Tod ausscheiden und Eintritte außer der Geburt nicht zugelassen sind, nennt man dieses Kollektiv geschlossen. Dieser Begriff ist [KLO+85, S. 16] entnommen. In dieser Arbeit wird davon ausgegangen, dass alle Kollektive geschlossen sind. Dies ist für Bevölkerungen sicher nicht der Fall. Der Fehler, der durch diese falsche Annahme entsteht, wird nicht korrigiert. Definieren wir jetzt die Teilmenge der Ausgeschiedenen analog zu [KLO+85, S. 13].

Definition 3 (Teilgesamtheit der Ausgeschiedenen). Es bezeichnet T_\star die Teilmenge der Verstorbenen aus K_\star und L_\star die Teilmenge der Überlebenden. Die Kardinalitäten werden mit $\#T_\star$ beziehungsweise $\#L_\star$ bezeichnet. Es gilt:

$$K_\star = T_\star \cup L_\star = \text{Verstorbene} \cup \text{Überlebende}$$

Bei mehreren Beobachtungsjahren werden, wenn \star zum Beispiel Alter ist, alle Personen aller Beobachtungsjahre zusammengefasst und in Altersgruppen unterteilt. Dann wird die Teilmenge aller Ausgeschiedenen ebenfalls nach Altern unterteilt und die Kardinalitäten aller betrachteten Gruppen bestimmt.

Aus den bisherigen Definitionen kann der Begriff der rohen Sterbehäufigkeit definiert werden. Hier ist Gleichung (3) analog zu [KLO+85, S. 13] und wird in Abschnitt 3.4 verwendet. Man beachtet, dass auch die Teilmenge der Überlebenden $\#L_\star$ verwendet wird, da man $\#K_\star = \#L_\star + \#T_\star$ zerlegen kann.

Definition 4 (Rohe Sterbehäufigkeit). Die rohe Sterbehäufigkeit \hat{q}_\star^{roh} für die Risikoklasse K_\star ist durch Gleichung (3) gegeben.

$$\hat{q}_\star^{roh} = \frac{\#T_\star}{\#K_\star} \quad (3)$$

Eine etwas simplere Sterbehäufigkeit, die sogenannte einfache rohe Sterbehäufigkeit, wird in Gleichung (4) gegeben. Diese wird im Abschnitt 3.2 verwendet. Sie ist unter der Annahme, dass sich das Kollektiv in der Zusammensetzung bezüglich \star nicht verändert sinnvoll, da dann $\#K/\#K_\star$ für jede Risikoklasse \star konstant ist und einfach mit diesem Wert multipliziert werden kann.

Definition 5 (einfache rohe Sterbehäufigkeit). Die einfache rohe Sterbehäufigkeit $\hat{q}_\star^{einfach}$ ist durch Gleichung (4) gegeben.

$$\hat{q}_\star^{einfach} = \frac{\#T_\star}{\#T} \quad (4)$$

Aus den Werten \hat{q}_\star^{roh} beziehungsweise $\hat{q}_\star^{einfach}$ werden, zum Beispiel durch Glättungsverfahren wie in Abschnitt 2.3 beschrieben, neue Werte berechnet. Diese Sterbehäufigkeiten \hat{q}_\star verlieren dann die Hochstellung. Fasst man Sterbehäufigkeiten \hat{q}_\star zusammen, erhält man Sterbetafeln.

An dieser Stelle sei noch bemerkt, dass in Kapitel 4 asymptotische Aussagen getätigt werden. Das heißt, es wird das Verhalten von \hat{q}_\star für $n \rightarrow \infty$ untersucht. Diese Aussage ist sinnvoll, da die Sterbehäufigkeiten Summen von Zufallsvariablen sind. Das heißt, man fasst Personen als Tupel der Zufallsvariablen Alter beziehungsweise Geburtsjahr und Todesalter beziehungsweise Todesjahr in einem bestimmten Beobachtungsjahr auf. Die Sterbehäufigkeiten sind dann das Verhältnis der Summe der Personen im Kollektiv mit passendem Todesalter durch die Summe aller Personen im Kollektiv.

Definition 6 (Sterbetafel). Eine Sterbetafel ST für Risikoklasse $\star = (\star_i, \star_j)_{i,j}$ ist von der Form wie Matrix 5.

$$ST = \begin{pmatrix} \hat{q}_{(\star_1, \star_1)} & \cdots & \hat{q}_{(\star_l, \star_1)} \\ \vdots & \vdots & \vdots \\ \hat{q}_{(\star_1, \star_k)} & \cdots & \hat{q}_{(\star_l, \star_k)} \end{pmatrix} \quad (5)$$

Definition 5 ist auch für Risikoklassen die nur eine Dimension durchlaufen sinnvoll, indem man die andere konstant hält. Zum Beispiel wird im Kapitel 3.2 eine Sterbetafel erstellt, die nur die Alter der Risikoklassen durchläuft. In diesem Fall ist $ST = (\hat{q}_S, \dots, \hat{q}_E)$. Das heißt, diese Sterbetafel gibt die Sterbehäufigkeit nur in einem bestimmten Beobachtungsjahr t wieder. Beinhalten Sterbetafeln Sterbehäufigkeiten für verschiedene Beobachtungsjahre t_i ermöglicht dies, sich entwickelnde Sterblichkeiten zu berücksichtigen. Diese Unterscheidung gibt Anlass zur Definition der Periodensterbetafel und der Generationensterbetafel. Laut [DAV04, S. 2] gibt eine Periodensterbetafel die Sterblichkeit für eine feste Beobachtungsperiode an, während Generationentafeln die Sterblichkeit pro Geburtsjahrgang inklusive der zukünftig erwarteten Sterblichkeitsveränderung enthalten.

Ein Beispiel für eine Sterbetafel ist die Sterbetafel 2004R. Dabei steht das R für Rente. Das heißt, diese Tafel wird für Versicherungsprodukte mit Erlebensfallcharakter, wie die Rente, konzipiert. Erlebensfallprodukte zeichnen sich dadurch aus, dass die versicherte Person eine Leistung erhält, wenn sie einen bestimmten Zeitraum n überlebt. Konkret werden zwei Arten von Sterbetafeln generiert: Zum einen wird versucht die Sterblichkeiten

möglichst genau zu schätzen, man spricht von einer Sterbetafel 2. Ordnung, zum anderen werden Risikoabschläge genommen, um sicherzugehen, dass die Versicherung ihren Verpflichtungen auf alle Fälle nachkommen kann. Hier spricht man von einer Sterbetafel 1. Ordnung. Für die Konstruktion von Verträgen wird die Tafel 1. Ordnung verwendet. In dieser Arbeit interessiert uns der statistische Aspekt der Versicherungsmathematik, weshalb uns nur die Herleitung der Tafel 2. Ordnung betrachtet wird.

An dieser Stelle sei noch kurz erwähnt, dass genau dieser Unterschied in den Sicherheitsabschlägen den Unterschied zwischen Tafeln mit Sterbe- und mit Rentencharakter ausmachen. Für Rentenversicherungen ist es sicherer, eine etwas zu geringe Sterblichkeit anzunehmen, während es für Risikolebensversicherungen sicherer ist, eine etwas zu hohe Sterblichkeit anzunehmen. Deshalb findet man auch in der DAV Tafel 2008T keine Schätzung von Sterbetrends. Es ist sicherer, für Versicherungen mit Todesfallcharakter keine Verbesserung der Sterblichkeit anzunehmen.

2.2 Probleme beim Schätzen roher Sterbewahrscheinlichkeiten

In diesem Kapitel werden Einflussgrößen diskutiert, die einen Einfluss auf die Erfassung der rohen Sterbewahrscheinlichkeiten haben. Es handelt sich um eine Aufzählung interessanter Punkte. Für diese Einflussgrößen werden keine Methoden zur Korrektur entwickelt:

- Jahreszeitliches Schwankungen: Wie man in der folgenden Abbildung 2 sieht, schwankt die Sterblichkeit je nach Jahreszeit. Erstellt man eine Sterbetafel, die nur auf wenigen Beobachtungsjahren basiert, hat je nachdem zu welchem Monat die Beobachtungsjahre beginnen und enden, eventuell einen Einfluss auf die Sterbetafel.
- Extremereignisse: Ein weiteres Problem ist, wie man mit Extremereignissen umgeht. Sowohl die DAV in der Dokumentation zur Sterbetafel 2004R, als auch Lee und Carter in ihrer Originalveröffentlichung [LC92] stellen die Frage, wie man mit Extremereignissen umgehen soll. Solche Ereignisse sind zum Beispiel Grippewellen, die durchaus wieder auftreten können.
- Änderung der Datenerfassung: Natürlich hat eine Änderung in der Datenerfassung einen Einfluss auf die Daten. Hier ist insbesondere die Wiedervereinigung Deutschlands 1989/90 problematisch. Zum Beispiel ist in Westdeutschland eine Lebendgeburt durch §31 Absatz 1 der Verordnung zur Ausführung des Personenstandsgesetzes wie folgt geregelt:

„Eine Lebendgeburt, für die die allgemeinen Bestimmungen über die Anzeige und die Eintragung von Geburten gelten, liegt vor, wenn bei einem Kinde nach der Scheidung vom Mutterleib entweder das Herz geschlagen oder die Nabelschnur pulsiert oder die natürliche Lungenatmung eingesetzt hat.“

Dagegen waren in der DDR mindestens zwei Merkmale erforderlich. Dies führt dazu, dass Säuglinge mit nur einem Lebenszeichen und somit schlechter Lebenserwartung

nicht als Lebendgeburt zählen und somit die Säuglingssterblichkeit besser geschätzt wird, als sie ist. Diese Information entstammt [Bel96].

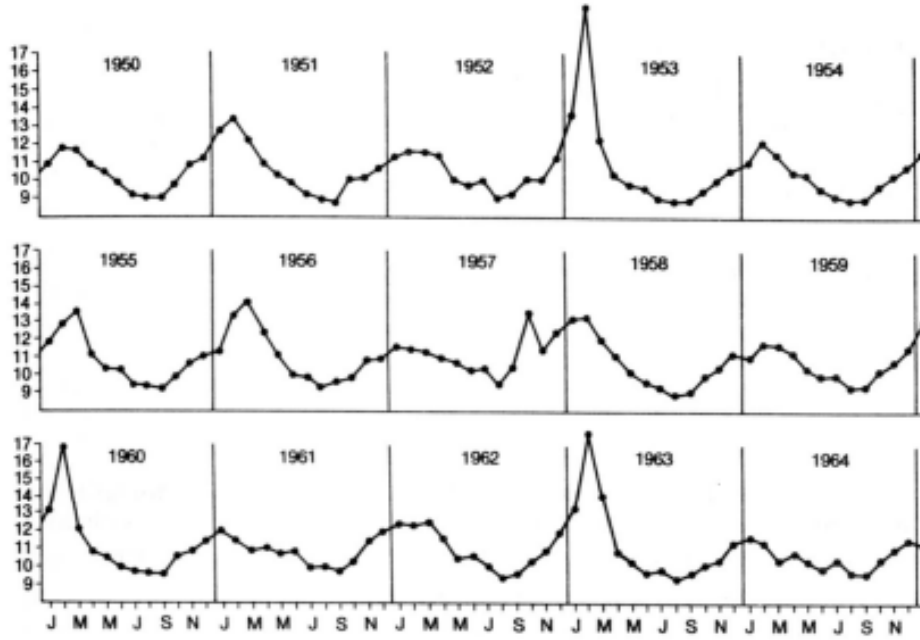


Abb. 10. Monatliche, auf ein Jahr normierte Sterbeziffern 1950–1964 in Prozent.

Abbildung 2: Diese Abbildung zeigt jahreszeitliche Schwankung der Sterblichkeit. Diese Abbildung ist Abbildung 10 in [KLO+85] Seite 4.2

2.3 Glättung von Sterbewahrscheinlichkeiten

In Abschnitt 2.1 wurde der Begriff der rohen Sterbehäufigkeit \hat{q}_x^{roh} beziehungsweise der einfachen Sterhäufigkeit $\hat{q}_x^{einfach}$ eingeführt. Das Ziel von Ausgleichsmethoden ist dann, aus diesen Werten neun Werte zu gewinnen, die näher an den wahren Sterbehäufigkeiten liegen. Früher wurde der Ausgleich mittels Millimeterpapier und Bleistift vorgenommen, doch da diese Art Daten auszugleichen sehr subjektiv ist, wurden objektivere Verfahren entwickelt. In diesem Kapitel wird ausgehend von einem Vektor $Q = (q_S, \dots, q_E)$ ein Ausgleichsverfahren als eine Abbildung $Q \ni q_x \mapsto q'_x \in Q'$ betrachten. Hier sind $Q' = (q'_S, \dots, q'_E)$ die ausgeglichenen Werte. Das heißt, man führt eine Funktion F ein mit:

$$q'_x = F(Q, x, a_k(Q, x)) \text{ für } k = 1, \dots, r \quad (6)$$

Bei der Definition der Funktion F , die auch als Ausgleichsfunktion bezeichnet wird, bezeichnet $x \in [S, E]$ das Alter, Q die Menge der rohen Sterbewahrscheinlichkeiten und a_k sind sogenannte Ausgleichsparameter. Die Ausgleichsparameter dienen dazu, Werte aus

Q verschieden zu gewichten. Sie müssen entweder im Vorfeld gesetzt oder aus den Daten berechnet werden.

Dieser Abschnitt ist in zwei Unterabschnitte unterteilt. Im ersten Unterabschnitt 2.3.1 wird eine Zusammenfassung von Kapitel 3 des Buches “Methodik von Sterblichkeitsuntersuchungen” [KLO+85] über verschiedene Glättungsverfahren gegeben. Die Informationen aus dieser Einleitung zu Glättungsverfahren stammen ebenfalls aus [KLO+85]. In Unterabschnitt 2.3.2 wird auf ein Glättungsverfahren, es handelt sich um das Whittaker-Henderson-Verfahren, genauer eingegangen. Dieses Verfahren schlägt die DAV in der Dokumentation zur Sterbetafel 2004R vor. Für das Whittaker-Henderson-Verfahren empfiehlt die DAV die Quelle [Loe94]. Diese ist nicht sehr tief gehend und deshalb wurde zusätzlich die Quelle [NS12] verwendet.

2.3.1 Überblick über Glättungsverfahren

Man unterscheidet zwei große Gruppen an Ausgleichsverfahren. Zum einen gibt es die Gruppe der mechanischen Methoden, zum anderen gibt es die Gruppe der analytischen Methoden. Die Zuordnung eines Verfahrens zu einer der beiden Gruppen ist nicht immer eindeutig.

Bei der mechanischen Ausgleichsrechnung ist die Funktion F aus Gleichung (6) eine lineare Funktion der Form:

$$q'_x = \sum_{k=-r}^{+s} a_k \cdot q_{x+k} \text{ für } r, s \in \mathbb{N}$$

Die a_k sind unabhängig von x und von q_x . Man beachte, dass r kleinere und s größere Werte den auszugleichenden Wert beeinflussen. Das heißt, die Werte an den Rändern können nicht voll ausgeglichen werden. Diese Werte an den Rändern nennt man Randwerte. Die zugrunde liegende Idee ist, Fehler durch die Zufallsschwankungen auf mehrere Werte zu verteilen und dadurch die Fehler zu verringern. Dadurch, dass die Fehler im Allgemeinen verschiedene Vorzeichen haben, hofft man, dass sich einige Fehler gegenseitig aufheben. Die mechanische Ausgleichsrechnung hat eine glättende Wirkung.

Drei bekannte Verfahren der mechanische Ausgleichsrechnung sind die 9-Punkte-Formel von Schärtlin (7), die 15-Punkte-Formel von Spencer (8) und das Whittaker-Henderson-Verfahren. Die Formel von Schärtlin und von Spencer sind nach [KLO+85, S. 87] durch die folgenden Gleichungen gegeben, während das Whittaker-Henderson-Verfahren in Unterabschnitt (2.3.2) vorgestellt wird:

$$q'_x = \frac{1}{27} (9q_x + 8q_{x+1} + 2q_{x+2} - q_{x+4}) \quad (7)$$

$$q'_x = \frac{1}{320} (74q_x + 67q_{x+1} + 46q_{x+2} + 21q_{x+3} + 3q_{x+4} - 5q_{x+5} - 6q_{x+6} - 3q_{x+7}) \quad (8)$$

Betrachten wir als nächstes die Verfahren der analytischen Ausgleichsrechnung. Bei diesen Verfahren ist die Funktion F aus Gleichung (6) eine nichtlineare Funktion in x und a_k .

Die Parameter a_k können von x und q_x abhängig sein. Die allgemeine Darstellung ist genau wie in Gleichung (6) und vereinfacht sich nicht. Dies bedeutet, dass wir die Parameter a_k für jeden Altersbereich und jede Wertereihe neu bestimmen müssen. Die Ausgleichsfunktion geben wir explizit vor, zum Beispiel durch eine Exponentialfunktion der Form $F(x, a_k(x, Q)) = \exp(a_2 x^2 + a_x x + a_0)$. Dadurch postulieren wir ein sogenanntes Sterbege-
setz, von dem wir annehmen, dass es den Verlauf der tatsächlichen Sterbewahrscheinlichkeit wiedergibt. Die Parameter a_k bestimmt man mit Hilfe zusätzlicher Forderungen, die man an die Ausgleichsfunktion F stellt.

Betrachten wir nun weitere Beispiele für Ausgleichsfunktionen. Eine einfache Klasse analytischer Ausgleichsverfahren basiert darauf, dass man die Anzahl der Lebenden l_x im Alter x als Folge betrachtet und eine Vorschrift angibt, wie man weitere Folgenglieder berechnen kann. Zwei Möglichkeiten sind die Sterbegeetze von de Moivre (1724) und Lambert (1765), die in [KLO+85, S. 104] beziehungsweise [For06, S. 1] angegeben sind. Die Formel von de Moivre ist sehr sinnvoll, da keine negativen Werte auftreten und bei Erreichen des Endalters E die Anzahl der Lebenden gleich Null ist. Bei der Formel von Lambert sind die Parameter a, b, c und d so zu bestimmen, dass keine negativen Werte auftreten und die Anzahl der Lebenden bei Erreichen des Endalters Null erreicht.

$$\begin{aligned} l_x &= l_0 \left(1 - \frac{x}{E}\right) \text{ de Moivre} \\ l_x &= l_0 \left[\left(\frac{a-x}{a}\right)^2 - b \left(e^{\frac{x}{c}} - e^{\frac{x}{d}}\right) \right] \text{ Lambert} \end{aligned}$$

Die nächsten Sterbegeetze, die entwickelt wurden, sind die von Gompertz (1824) und Makeham (1860). Wir werden diese in Abschnitt 3.3 genauer besprechen. Diese sind nach [KLO+85, S. 105] von der folgenden Form, wobei die Parameter $a_k \in \{h, s, g, c\}$ geschätzt werden müssen:

$$l'(x) = h \cdot s^x \cdot g^{c^x}$$

Im 20. Jahrhundert wurden laut [KLO+85, S. 105] Sterbegeetze der folgenden Form entwickelt:

$$\begin{aligned} q'_x = F(Q, x, a_k(Q, x)) &:= \exp \left(\sum_{k=1}^r a_k x^{k-1} \right) \\ q'_x = F(Q, x, a_k(Q, x)) &:= \exp \left(-\exp \left\{ \sum_{k=1}^r a_k x^{k-1} \right\} \right) \end{aligned}$$

Nach der Wahl eines Sterbegesetzes müssen als nächstes die Parameter a_k für $k = 1, \dots, r$ bestimmt werden. Eine Möglichkeit ist, eine Gleichung anzugeben, die minimiert werden soll. Das bekannteste Verfahren dieser Art ist das Verfahren der kleinsten Quadrate, bei dem nach [KLO+85, S. 106] folgende Bedingung (9) angesetzt wird. Dabei wird davon

ausgegangen, dass F und g zwei 2-fach stetig differenzierbare Funktionen sind. Mögliche Wahlen von g sind $g = 1$ oder $g = l_x/q_x$. Dabei ist l_x/q_x eine sinnvolle Wahl, wenn man davon ausgeht, dass Todesfälle poissonverteilt sind.

$$\sum_{x=S}^E g(x, a_k) (F(x, a_k) - q_x)^2 = \text{Minimum}_{a_k} \quad (9)$$

Ein weiteres Verfahren, bei dem eine Gleichung minimiert wird, ist das Verfahren der Spline-Funktionen nach Reinsch. Dieses ist nach [KLO+85, S. 107] durch die folgenden Gleichungen (10) und die Nebenbedingung (11) gegeben. Hier sind g_i und $A \in \mathbb{R}^+$ fest vorgegeben und man geht davon aus, dass $F(x, a_k)$ zweimal stetig differenzierbar ist.

$$\int_S^E \left(\frac{d^2}{dx^2} F(x, a_k) \right)^2 dx = \text{Minimum}_{a_k} \quad (10)$$

$$\sum_{i=S}^E \frac{(F(x_0 + i, a_k) - q_{S+i})^2}{g_i} \leq A \quad (11)$$

Eine andere Art, die r Parameter a_k für $k = 1, \dots, r$ zu bestimmen, ist r Gleichungen aufzustellen. Würde man mehr oder weniger Gleichungen aufstellen, wäre das lineare Gleichungssystem über- oder unterbestimmt. Überbestimmte Gleichungssysteme sind oft unlösbar. Eines dieser Verfahren ist das Verfahren der Ordinatenmomente (12). Es ist nach [KLO+85, S. 106] durch die folgenden Gleichungen gegeben:

$$\sum_{x=S}^E x^v F(x, a_k) = \sum_{x=S}^E x^v q_x \text{ für } v = 0, \dots, r-1 \quad (12)$$

Betrachten wir, um diesen Unterabschnitt abzuschließen, noch den Zusammenhang zwischen den mechanischen und den analytischen Ausgleichsverfahren. Dieser Zusammenhang wird deutlich, wenn man von dem folgenden Modell (13) ausgeht:

$$q_x = f(x) + \epsilon \quad (13)$$

Dabei bezeichnet ϵ zufällige Fehler und f ist eine unbekannte Funktion. Man geht also davon aus, dass es ein Sterbegesetz gibt, welches mit f bezeichnet wird. Wendet man nun ein mechanisches Ausgleichsverfahren auf q_x an, erhält man:

$$q'_x = \sum_{k=-r}^{+s} a_k \cdot q_{x+k} = \sum_{k=-r}^{+s} a_k \cdot f(x+k) + \sum_{k=-r}^{+s} a_k \cdot \epsilon_{x+k}$$

Man fordert von einem mechanischem Ausgleichsverfahren, dass es bezüglich f invariant ist und die Fehler verringert:

$$\sum_{k=-r}^{+s} a_k \cdot f(x+k) = f(x) \text{ (Invarianz)}$$

$$\sum_x \left(\sum_{k=-r}^{+s} a_k \cdot \epsilon_{x+k} \right)^2 < \sum_x \epsilon_x \text{ (Anpassung)}$$

Das heißt, in einem mechanischen Ausgleichsverfahren werden die Fehler minimiert, ohne dass ein Sterbegezet angenommen werden muss. In einem analytischen Ausgleichsverfahren wird dagegen ein Sterbegezet angenommen und danach die Fehler minimiert. Man kann also die mechanischen Ausgleichsverfahren als allgemeiner ansehen.

2.3.2 Das Verfahren von Whittaker-Henderson

Wie bereits oben erwähnt, wird im Folgenden das Verfahren von Whittaker-Henderson betrachtet. Die Idee dieser Konstruktionsmethode ist, die Bestimmungsgleichungen für die a_k aus einer Optimierungsaufgabe herzuleiten. Seien dazu das Bestimmungsmaß (14) und das Glättemaß (15) wie in [KLO+85, S. 92] definiert:

$$\sum_{k=S}^E (q'_{S+k} - q_{S+k})^2 \quad (14)$$

$$\sum_{k=S}^{E-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{S+k+v} \right)^2, \text{ für } s \in \mathbb{N} \quad (15)$$

Das Glättemaß (15) wird in der Literatur auch mit $\Delta^s q'_{S+k}$ bezeichnet. Insgesamt erhält man, wenn man Gleichung (14) und (15) kombiniert, folgende Optimierungsaufgabe (16):

$$\sum_{k=S}^E w_k (q'_{S+k} - q_{S+k})^2 + g \cdot \sum_{k=S}^{E-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{S+k+v} \right)^2 = \text{Minimum}_{q'} \quad (16)$$

Hier sind w_k zusätzliche positive Gewichte und g eine positive Konstante, die das Maß der Glättung vorgibt. Für $g \rightarrow 0$ erhält man die originalen Daten, während für $g \rightarrow \infty$ die Daten mehr und mehr geglättet werden. Im Grenzfall erhält man $q'_x = 0$ für alle $x \in \{E, \dots, S\}$, also maximal geglättete Daten.

Zur Lösung des Optimierungsproblems (16) bietet es sich an, das Problem in Matrixschreibweise zu formulieren. Dieses Reformulierung orientiert sich an [NS12]. Sei K die $(n-p) \times n$ Matrix mit den Einträgen $K_{i,j}$:

$$K_{i,j} = \begin{cases} (-1)^{s+j-i} \binom{s}{j-i} & \text{für } i = 1, 2, \dots, s \text{ und } j = i, i+1, \dots, i+s \\ 0 & \text{andernfalls} \end{cases}$$

Damit lässt sich das Optimierungsproblem (16) schreiben als:

$$\begin{aligned}
& \sum_{k=S}^E w_k (q'_{S+k} - q_{S+k})^2 + g \cdot \sum_{k=S}^{E-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{S+k+v} \right)^2 \\
&= (Q' - Q)^t W (Q' - Q) + g \cdot \sum_{k=S}^{E-s} \left(\sum_{v=0}^s K_{k,v} q'_{S+k+v} \right)^2 \\
&= (Q' - Q)^t W (Q' - Q) + g (Kq')^t (Kq') = \text{Minimum}_{Q'}
\end{aligned}$$

Setzt man $H = g \cdot I$ und $w_k = 1$ für alle k , erhält man ein Problem der Form:

$$(u - y)^t (u - y) + (Ku)^t H (Ku) = \text{Minimum}_u \quad (17)$$

Für Probleme der Form (17) zeigt Greville [Gre81, S. 54], dass eine Lösung durch die folgende Gleichung gegeben ist:

$$(I + K^t H K) u = y$$

Da die Matrix $(I + gK^t K)$ positiv definit und damit invertierbar ist erhält man durch Invertieren der Matrix $(I + gK^t K)$ eine Lösung:

$$Q' = (I + gK^t K)^{-1} Q$$

Es bleibt, sich Gedanken über die Auswahl des Glättungsparameters g zu machen. Wir werden hier die Generalised-Cross-Validation-Methode (GCV) benutzen, die in [NS12] Kapitel 4 vorgeschlagen wird. Dabei wird g als Minimum des folgenden Terms gewählt:

$$\text{GCV} = \frac{(q - q')^t (q - q')}{(1 - \text{Tr}((I + gK^t K)^{-1}))^2 / n}$$

Betrachten wir jetzt ein Beispiel. Es wird ein Datensatz aus dem Regressionsmodell (18) aus Kapitel 3.2 erzeugt. Dazu wird wie folgt vorgegangen:

1. Starte mit einem Vektor X , der die möglichen Alter enthält $X = (0, \dots, 95)$ und berechne die fehlerfreien Y Werte $f(X)$ gemäß (18). Dabei wird $m = 70$ und $s = 15$ benutzt.
2. Erzeuge zufällige Fehler $\epsilon \sim N(0, \sigma_\epsilon^2)$ mit $\sigma_\epsilon^2 = 4$
3. Berechne $Y = f(X) + \epsilon$

$$Y = f(X) + \epsilon = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp \left(-\frac{(m - X)^2}{2 \cdot s^2} \right) + \epsilon \quad (18)$$

Die Parameter für m und s sind so gewählt, dass sie nah an den geschätzten Parametern aus den Daten von [Mor18] sind. Der Glättungsparameter g wird aus $\{2^r : r = 0, \dots, 10\}$

ausgewählt. Das heißt, für $g \in \{2^r : r = 0, \dots, 10\}$ wird der GCV-Wert bestimmt und das g mit dem kleinsten Wert wird verwendet.

Das Ergebnis der Glättung ist in Abbildung 3 zu sehen. Die durchgezogene Linie sind Werte von Y und die gestrichelte Linie sind die mittels Whittaker geglätteten Daten. Die gepunkteten Linien sind die Werte von $f(X)$. Man sieht, dass das geglättete Modell dem wahren Modell stark ähnelt.

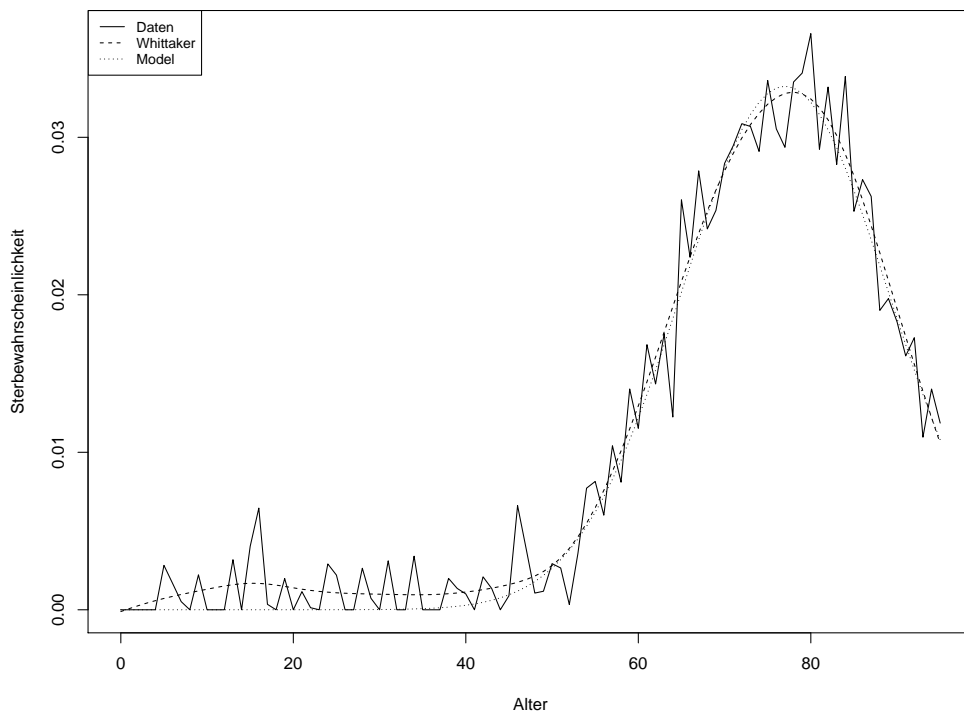


Abbildung 3: Glättung von Daten mittels des Whittaker-Henderson-Verfahrens.

2.4 Öffentlich zugängliche Sterblichkeitsdaten

Die Sterbetafel 2004R basiert laut [DAV04, S. 6] auf den folgenden Daten:

- Versichertendaten für die Jahre 1995 bis 2002, die von der Münchner Rück und der GenRe ausgewertet wurden,
- Abgekürzte Bevölkerungssterbetafeln des Statistischen Bundesamtes für das frühere Bundesgebiet ab 1971/73 bis 1998/2000 (für 1986/88 stattdessen die Allgemeine Deutsche Sterbetafel 1986/88) und
- Daten der gesetzlichen Rentenversicherung (GRV) für das frühere Bundesgebiet von 1986 bis 2002 für die Alter von 66 bis 98. Man beachte, dass die DAV $E = 110$ verwendet.

Die Daten der Münchner Rück und der Gen Re sind Privateigentum der jeweiligen Unternehmen und deshalb nicht öffentlich zugänglich. Auch die Daten des Statistischen Bundesamts beziehungsweise der gesetzlichen Rentenversicherung sind auf ihren jeweiligen Webseiten nicht ohne weiteres öffentlich zugänglich. Allerdings gibt es auf der Internetseite Mortality.org Daten über die Sterblichkeit für viele Staaten.

Werfen wir nun einen Blick auf einen Ausschnitt aus “Brief Summary of the Methods Protocol for the Human Mortality Database”, um zu sehen, welche Daten vorliegen. Dieses Dokument kann auf der Website von Mortality.org gefunden werden. Ich habe die Punkte etwas gekürzt.

- “Births: Annual counts of live births by sex and by month are collected for each population over the longest time period available. At a minimum, a complete series of birth counts by sex is needed for the time period over which mortality rates and period life tables are computed.”
- “Deaths: Death counts are collected by sex, completed age, year of birth, and year of death if available. Before making subsequent calculations, deaths of unknown age may be distributed proportionately across the age range.”
- “Population size: Below age 80, estimates of population size on January 1st of each year are either obtained from another source or derived using intercensal survival. When necessary, persons of unknown age are distributed proportionately across age. Above age 80, population estimates are used. For non-extinct cohorts aged 80 to 90 at the end of the observation period, population estimates are obtained either from another source or by applying the method of intercensal survival.”
- “Exposure-to-risk: Estimates of the population exposed to the risk of death during some age-time interval are based on annual population estimates, with a small correction that reflects the timing of deaths during the interval and variation in cohort’s birthdays by month.”
- “Death rates: For both periods and cohorts, death rates are simply the ratio of deaths to exposure-to-risk in matched intervals of age and time.”

Diese Daten sind von 1990 bis 2017 für das ganze Bundesgebiet verfügbar und für West-beziehungsweise Ostdeutschland separat für die Jahre 1956 bis 2017. Die folgende Tabelle soll einen kurzen Einblick in die Struktur der Daten geben:

Year	Age	Female	Male	Total
1949	0	.	.	.
1949	1	.	.	.
1949	2	.	.	.
1949	3	.	.	.
1949	4	.	.	.
1949	5	.	.	.
1949	6	.	.	.
1949	7	0.000412	0.000714	0.000566
1949	8	0.000363	0.000632	0.000501
1949	9	0.000335	0.000539	0.000440
1949	10	0.000260	0.000484	0.000375
1949	11	0.000298	0.000446	0.000374
1949	12	0.000332	0.000445	0.000390
1949	13	0.000282	0.000432	0.000359
1949	14	0.000315	0.000542	0.000432
1949	15	0.000384	0.000667	0.000529
1949	16	0.000467	0.001051	0.000767
1949	17	0.000500	0.001344	0.000933
1949	18	0.000621	0.001643	0.001145
1949	19	0.000562	0.001676	0.001133
1949	20	0.000571	0.001785	0.001190

Tabelle 1: Geburtsjahrgang 1949 der Generationstafel 1876 Westdeutschland

Man sieht anhand Tabelle 1, dass zum Beispiel für den Jahrgang 1949 Daten erst ab einem Alter von 7 Jahren gibt. Es gibt dann Daten bis zu einem Alter von 67 Jahren. Das Endalter bei diesen Tabellen ist 110. Alle Personen, die 110 Jahre oder mehr alt sind, werden in eine Kategorie, genannt 110+, gruppiert.

Man erhält nur einen Ausschnitt, da die Generationstafel aus einer Periodentafel bestimmt wird: Wenn man die Anzahl an Toden von Personen mit Alter und aktueller Jahreszahl hat, kann man zurückrechnen, wie viele Personen mit einem bestimmten Geburtsjahr in welchem Alter verstorben sind. Dies kann man auch dem oben bereits erwähnten Dokument “Brief Summary of the Methods Protocol for the Human Mortality Database” entnehmen:

“Raw data are usually obtained in a period format (i.e., by the year of occurrence). In the HMD, most data are presented in a period format, but we also provide death rates and life tables in a cohort format (i.e., by year of birth) if the observation period is sufficiently long to justify such a presentation.”

Man beachte weiterhin die Spalte “Year”. Diese Spalte kann sich entweder auf das Beobachtungsjahr oder auf das Geburtsjahr beziehen. Auf Mortality.org unter dem Link <https://www.mortality.org/Public/ExplanatoryNotes.php> findet man die folgende Erklärung:

“Period data are indexed by calendar year, whereas cohort data (if available) are indexed by year of birth.”

Im *R*-Paket “moRtRe”, in dem die Abbildungen und Simulationen dieser Arbeit ausgeführt wurden, sind deshalb die Daten im Format “Beobachtungsjahr Alter Total” beziehungsweise “Geburtsjahr Alter Total” in /data abgelegt.

Man beachte weiterhin, dass die Daten gerade für hohe Alter qualitativ recht fragwürdig sind. Ein Problem ist, dass manchmal Beobachtungen nicht zuzuordnen sind. In diesen Fällen werden die Beobachtungen proportional auf die verschiedenen Kategorien verteilt. Dies kann dann dazu führen, dass man bei Todeszahlen keine ganzen Zahlen bekommt, sondern gebrochene. Ein anderes Problem sind hohe Alter. Zum Beispiel betrug laut Mortality.org die Sterbewahrscheinlichkeit für 109-jährige im Kalenderjahr 1987 ganze 400 %. Diese hohe Prozentzahl entsteht durch zwei Tode, die sich auf 0.5 Personen unter Todesrisiko aufteilen. Hier die entsprechenden Zeilen von der Website:

Year	Age	Female	Male	Total	
1987	109	2.00	0.00	2.00	Todesfälle

Year	Age	Female	Male	Total	
1987	109	0.50	0.00	0.50	Individuen

Tabelle 2: Ein Beispiel für die schlechte Datenlage bei hohen Altern. Zu sehen ist die Altersgruppe der 109-jährigen im Beobachtungsjahr 1987 der Periodentafel 1956 Westdeutschland. Tode in dem oberen Teil und Personen unter Risiko in dem unteren.

Betrachtet man die zugehörigen rohen Daten, so sieht man, dass ein Problem darin liegt, dass für Alter 95+ keine Daten vorliegen und für höhere Alter bereits approximiert wird. Außerdem wird nicht die ganzzahlige Bevölkerung, sondern die Bevölkerung unter Risiko verwendet, welche nicht notwendigerweise ganzzahlig ist. Um diesen Problemen zu begegnen, werden in dieser Arbeit Alter größer 95 nicht beachtet.

3 Modellbildung

In Kapitel 2 wurden die zentralen Begriffe der Sterbehäufigkeit und der Sterbetafel definiert, Verfahren zum Schätzen von geglätteten Sterbetafeln diskutiert und Quellen für historische Daten angegeben. In diesem Kapitel beginnt der schließende Teil dieser Arbeit. Das heißt, es werden Regressionsmodelle eingeführt, um die Sterbehäufigkeiten zu schätzen. Wir beginnen in den Abschnitten 3.1 und 3.2 mit einfachen Modellen, die an den Datensatz Generationstafel 1876 Westdeutschland angepasst werden. Diese Modelle sind einfach, da nur eine Variable, entweder das Geburtsjahr oder das Alter, verwendet werden. Eine Trendschätzung wird nicht durchgeführt. Diese Modelle dienen als Basisfall und sollten beim Vergleich mit den komplexeren Modellen schlechter abschneiden.

Danach werden wir in Abschnitt 3.3 sehen, wie die DAV ihre Sterbetafeln schätzt. In Abschnitt 1 wurde bereits gesagt, dass die Herleitung der Sterbetafel 2004R der Ansatzpunkt dieser Arbeit war und in diesem Kapitel wird diese Sterbetafel diskutiert. Zwei Methoden, die in der Herleitung dieser Sterbetafel erwähnt werden, sind das Whittaker-Henderson-Verfahren zum Glätten von Sterbewahrscheinlichkeiten und das Lee-Carter-Modell zum Trendschätzen. Im Abschnitt 2.3.2 haben wir uns bereits mit dem Whittaker-Henderson-Verfahren befasst. In Abschnitt 3.4 wird das Lee-Carter-Modell diskutiert.

Am Schluss dieses Kapitels wird in Abschnitt 3.5 das einfache Modell aus Abschnitt 3.2 um eine Trendkomponente erweitert.

3.1 Ein einfaches Modell: Geburtsjahr als erklärende Variable

Beginnen wir damit ein Modell zu entwickeln, indem das Todesjahr t beziehungsweise das Todesalter y in Abhängigkeit des Geburtsjahres g vorausgesagt werden soll. In diesem Modell sind sowohl das Todesalter beziehungsweise Todesjahr als auch das Geburtsjahr zufällig. Seien also $Y = (Y_1, \dots, Y_n)$ ein Zufallsvektor von Todesaltern, $\mathfrak{T} = (\mathfrak{T}_1, \dots, \mathfrak{T}_n)$ ein Zufallsvektor von Todesjahren, $g = (g_1, \dots, g_n)$ ein Zufallsvektor von Geburtsjahren und $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ ein Zufallsvektor von Fehlern. Zu diesem Zeitpunkt machen wir keine Annahmen über die Fehler. Für das erste Modell gehen wir von einem einfachen, funktionalen Zusammenhang der folgenden Art aus:

$$Y = \mathfrak{T} - g = f(g) - g + \epsilon$$

Dieses Modell hat den Vorteil, dass man Veränderungen in den Sterblichkeiten im Laufe der Zeit, das heißt für verschiedene Geburtsjahrgänge, direkt abbilden kann. Man berücksichtigt also Sterblichkeitstrends. Es hat allerdings das große Problem, dass es bei Anwendung auf historische Daten problematische Ergebnisse liefert. Genauer gesagt führt dieses Modell dazu, dass eine Person, je später sie geboren ist, umso jünger stirbt. Dies ist in Abbildung 4 veranschaulicht:

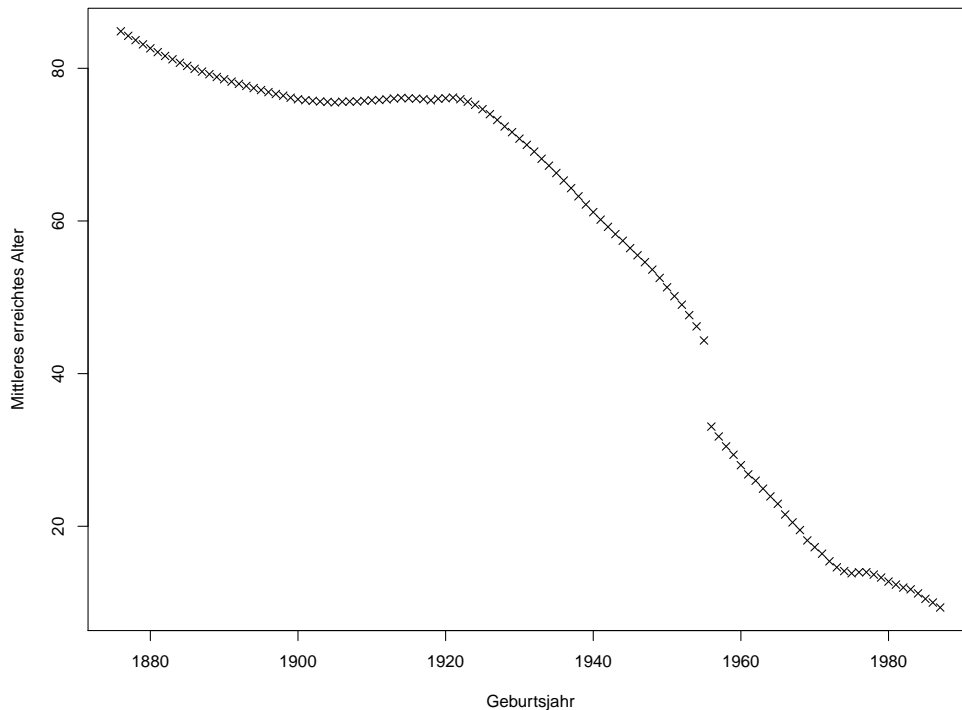


Abbildung 4: Man sieht die beobachteten Werte der Tupel (Geburtsjahr, mittleres erreichtes Alter) bei der Generationstafel 1876 Westdeutschland.

Dies macht auch Sinn, wenn man einen Moment darüber nachdenkt: Das mittlere Alter in dem man Personen sterben sieht, die 1950 geboren sind, ist natürlich höher als für Personen, die 2017 geboren sind. Diese fehlenden Information über Personen die heute geboren, aber noch nicht gestorben sind bezeichnet man als Zensur der Daten. Zum Abschluss dieses Abschnitts betrachten wir noch eine Heatmap in Abbildung 5, die die Anzahl an Todesfällen in Abhängigkeit von Geburtsjahr und Alter beschreibt. Die wichtige Aussage dieser Abbildung ist, dass man eine Diagonale von links oben, hohe Alter und frühe Geburtsjahre, nach unten links, junge Alter und späte Geburtsjahre hat. Außerhalb der Diagonale gibt es keine Beobachtungen. Deshalb eignet sich das Geburtsjahr nicht als einzige erklärende Variable.

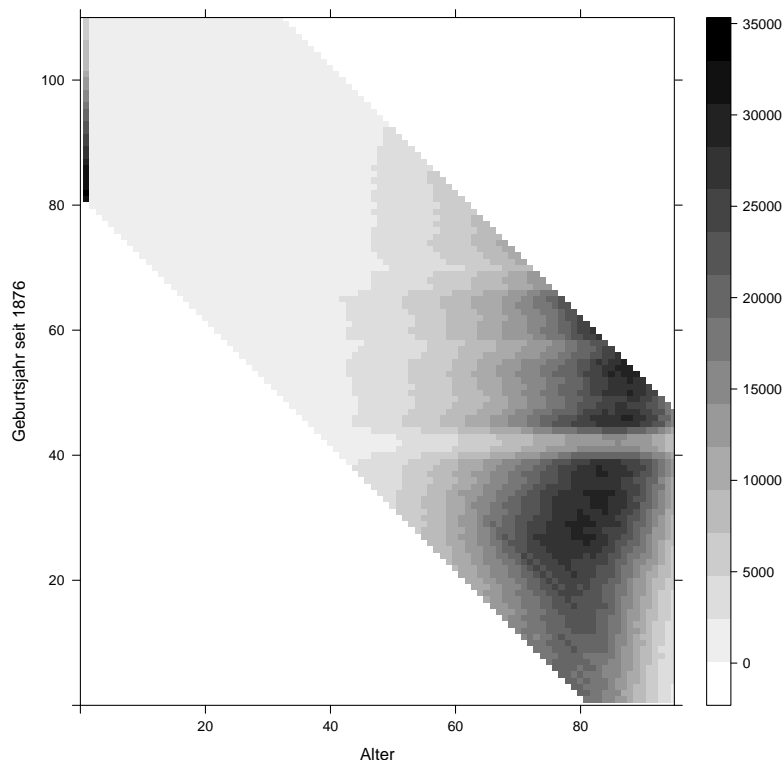


Abbildung 5: Eine Heatmap der Todesfälle der Generationstafel 1876 West. Die x-Achse zeigt das Geburtsjahr und die y-Achse das Alter der Personen.

3.2 Ein einfaches Modell: Alter als erklärende Variable

In Abschnitt 3.1 haben wir ein Modell gesehen, welches Sterblichkeitstrends berücksichtigt, aber leider bei der Anwendung auf historische Daten versagt. Als nächstes wird ein weiteres einfaches Modell betrachtet, doch diesmal wird nur das Alter als erklärende Variable benutzt. Dies führt dazu, dass Sterblichkeitstrends nicht berücksichtigt werden, da Sterblichkeiten für ein Alter für verschieden Beobachtungsjahre als konstant angenommen werden.

Konkret werden diesmal Risikoklassen nach Alter $x = \mathfrak{b} - \mathfrak{g}$ gebildet. Für gegebene Risikoklassen K_S, \dots, K_E wird die Sterbewahrscheinlichkeit in jeder Klasse als der relative Anteil an Toden in dieser Altersklasse im Verhältnis zu allen Toden geschätzt. Dieser Sachverhalt ist in der folgenden Formel fixiert:

$$q_x^{einfach} = \frac{\#T_x}{\#T} \quad (19)$$

Es wird wieder ein Regressionsmodell verwendet, bei dem der Vektor $Q = (q_S, \dots, q_E)$ die Zielvariable beschreibt. Bezeichne weiterhin $X = (S, \dots, E)$ den Vektor mit Altern. Sei

schlussendlich $\epsilon = (\epsilon_S, \dots, \epsilon_E)$ ein Vektor mit Fehlern. Dann wird der folgende funktionale Zusammenhang unterstellt:

$$Q = h(X) + \epsilon \text{ beziehungsweise } q_x = h(x) + \epsilon_x \text{ für } x \in \{S, \dots, E\} \quad (20)$$

Die Frage ist nun, von welcher Art der funktionale Zusammenhang h ist. Ein Modell, welches oft verwendet wird, ist das Lee-Carter-Modell, das in Abschnitt 3.4 genauer besprochen wird. Im Lee-Carter-Modell werden zuerst die Sterbehäufigkeiten logarithmiert. Das dadurch entstehende Modell ist dann linear. Logarithmiert man die Sterbewahrscheinlichkeiten des Datensatzes Generationstafel 1876 Westdeutschland und passt dann ein lineares Modell an, so erhält man Abbildung 6. Man sieht, dass der Ansatz die Daten zu logarithmieren, sicher nicht zu einer Linearisierung im ausreichenden Maß führt. Die Frage ist also, was für ein anderer Zusammenhang sinnvoll ist. Da die einfachen Sterbehäufigkeiten in Summe 1 ergeben, sprich $\sum_{x=S}^E q_x^{einfach} = 1$, liegt es nahe, dass die Funktion h diese Eigenschaft ebenfalls erfüllt. Es liegt also nahe für den funktionalen Zusammenhang h eine Dichte anzusetzen.

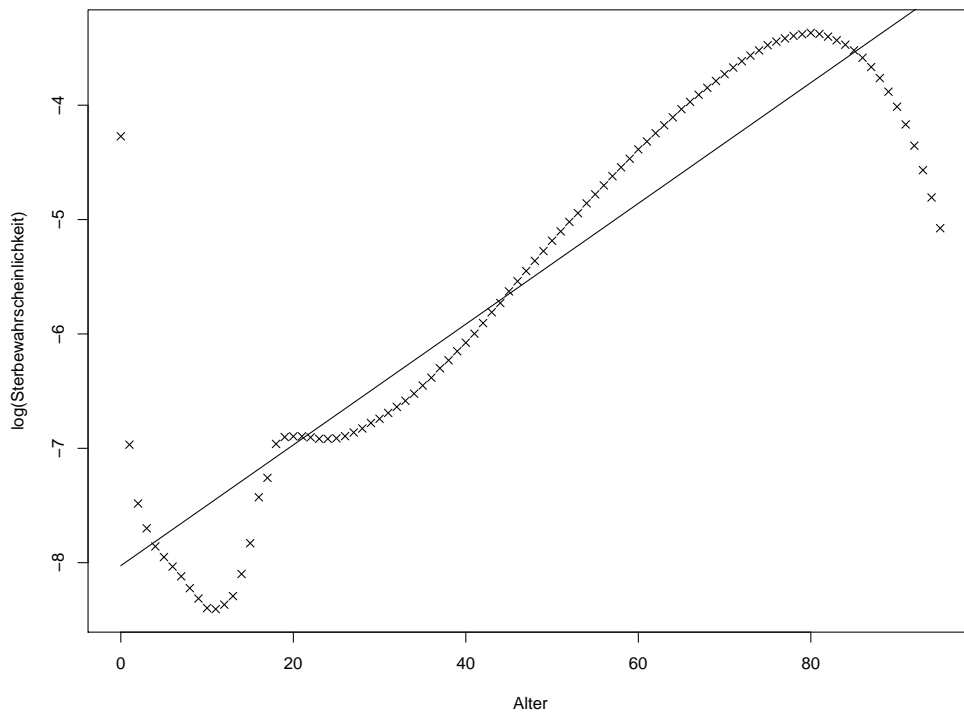


Abbildung 6: Alter versus logarithmierte relative Anzahl an Sterbefällen pro Alter bei der Generationstafel 1876 Westdeutschland. Außerdem ist eine Regressionsgerade eingezeichnet, bei der mit der einfachen kleinste-Quadrate-Methode (OLS) $\log(Q)$ auf X regressiert wurde.

Um das Modell einfach zu halten, wird in dieser Arbeit die Normalverteilung verwendet. Das heißt, es wird das Regressionsmodell (21) benutzt. Diese Wahl hat sicherlich einige Nachteile. So hat die Normalverteilung beispielsweise Masse im negativen Bereich. Das heißt, die Wahrscheinlichkeit mit einem negativen Alter zu sterben, ist größer Null. Außerdem sind die Daten linksschief, während die Normalverteilung symmetrisch ist. Um diesen Problemen zu begegnen, wird, soweit möglich, in Abschnitt 4.2 mit dem allgemeineren Modell (22) gearbeitet. In diesem Modell kann die Funktion h so gewählt werden, dass es keine Masse im negativen Bereich gibt und h insgesamt linksschief ist.

$$Q = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon \quad (21)$$

$$Q = \frac{1}{s} \cdot h\left(\frac{(m - X)}{s^2}\right) + \epsilon \quad (22)$$

Bei Modell (21) sind alle Operationen vektorwertig zu verstehen. Das heißt, $Q, X, \epsilon \in \mathbb{R}^{E-S+1}$. In Abbildung 7 sieht man die Daten aus der Generationstafel 1876 Westdeutschland und zusätzlich der geschätzte, funktionale Zusammenhang h .

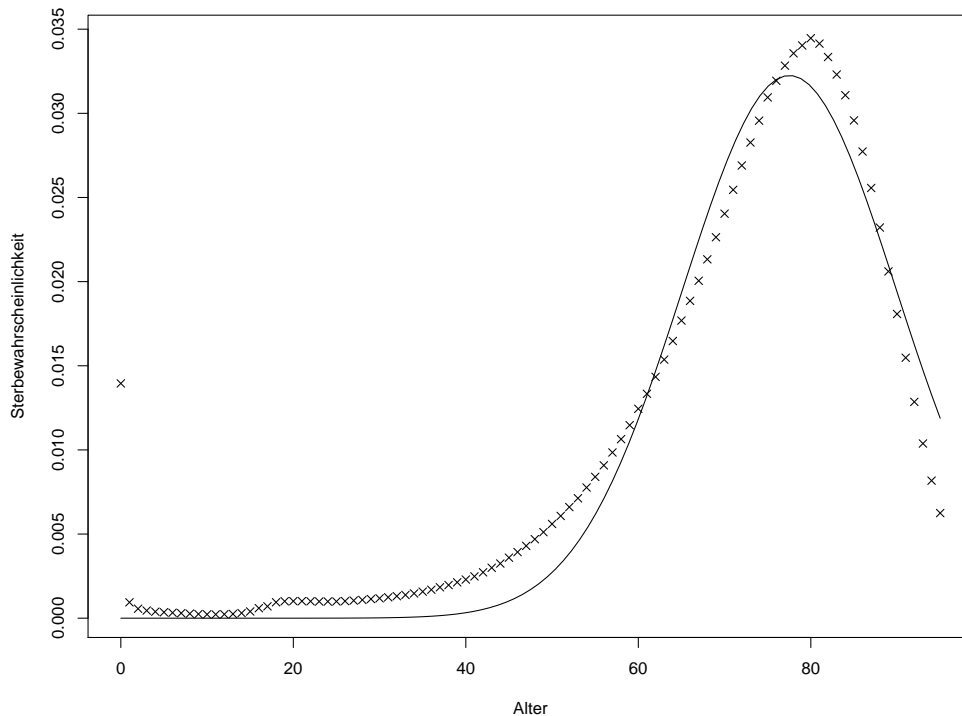


Abbildung 7: Zu sehen ist Alter versus die Sterblichkeit bei der Generationstafel 1876 Westdeutschland. Die durchgezogene Linie ist ein Schätzer für das Modell.

Die Interpretation der Parameter in Modell (21) ist recht einfach: Der Parameter m ist der Mittelwert der erreichten Lebensalter und der Parameter s ist die Standardabweichung von diesem Wert.

3.3 Das Vorgehen der DAV

Die DAV hat die Sterbetafel 2004R in zwei Schritten erstellt: Im ersten wird eine Basistafel geschätzt, im zweiten wird auf Grundlage der Basistafel eine Trendfunktion geschätzt. Diese Trendfunktion soll Verbesserungen in der Sterblichkeit modellieren. Bei einer Basistafel handelt es sich um eine Periodensterbetafel bei der ein bestimmtes Kalenderjahr als Basisjahr zugrunde gelegt wird.

Bei der Erstellung der Basistafel wird in 4 Schritten vorgegangen: Zuerst muss ein Basisjahr für die Tafel festgelegt werden. Das heißt, es muss ein Jahr bestimmt werden, für das die Tafel gelten soll. Danach werden die rohen Sterbewahrscheinlichkeiten bestimmt und im nächsten Schritt geglättet. Zum Schluss werden Sterbewahrscheinlichkeiten für hohe Alter extrapoliert. Die folgende Beschreibung orientiert sich an [DAV04, S. 9]. Man beachte, dass die Sterbetafel 2004R aus einer Tafel für die Rentenbezugszeit und aus einer Tafel für die Aufschubszeit besteht. Da es sich bei der Tafel zur Rentenbezugszeit um eine sogenannte Selektionstafel handelt, wird deren Herleitung hier nicht betrachtet. Selektionstafeln sind nach [Wol] besondere Tafeln die neben dem Alter und dem Geschlecht zusätzlich noch den Zeitraum seit Abschluss eines Lebensversicherungsvertrags als weiteres Risikomerkmal (Risikofaktoren) zur Bestimmung der Sterbewahrscheinlichkeit berücksichtigen.

Die Auswahl des Basisjahrs wird in Anhang 4 der Herleitung der Sterbetafel 2004R erklärt:

“Der mittlere Zeitpunkt des Beobachtungsmaterials wird definiert durch Gewichtung der Verweildauer der Jahre 1995 bis 2002, also als

$$\frac{\sum_{t=1995}^{2002} t \cdot VWD(t)}{\sum_{t=1995}^{2002} VWD(t)} \quad (23)$$

wobei $VWD(t)$ die gesamte nach Rentenhöhen gewichtete Verweildauer des Beobachtungsmaterials im Jahr t bezeichnet.”

Die DAV schlägt also vor, als Bezugsjahr der Basistafel das mittlere Jahr des Beobachtungszeitraums nach Rentenhöhe zu nehmen. Dies wird nicht weiter motiviert, allerdings spricht dafür, dass im mittleren Jahr der Effekt der Sterblichkeitsverbesserung zu den Altern an den Rändern des Beobachtungszeitraums am geringsten sei dürfte.

Als nächstes muss man für das Basisjahr rohe Sterbewahrscheinlichkeiten bestimmen. Wie bereits in Abschnitt 2.4 beschrieben, werden zur Bestimmung der rohen Sterbewahrscheinlichkeiten die beiden Kennzahlen $\#L_x$ und $\#T_x$ für ein Kollektiv bestimmt. Die DAV nutzt allerdings nicht $\#L_x$ direkt, sondern die Größe i_x die anzahlgewichtete Verweildauer der Lebenden des Alters x . Das heißt, eine Person, die ein halbes Jahr nach ihrem x -ten

Geburtstag verstirbt, trägt $1/2$ zu i_x bei. Auf der anderen Seite bezeichnet $\#T_x$ weiterhin die Anzahl an Todesfällen von Personen im Alter x . Die rohen Sterbehäufigkeiten sind dann der Bruch dieser beiden Größen.

$$q_x^{\text{roh}} = \frac{\#T_x}{i_x} \quad (24)$$

Dieses Verfahren nennt man Verweildauerverfahren. Es findet auch in der Sterbetafel 2008T Anwendung.

Als nächstes wird die Basistafel geglättet. Diese Methoden wurde in Abschnitt 2.3 im Detail betrachten. Es stehen eine Vielzahl von Methoden zur Verfügung.

Nachdem man eine Basistafel geschätzt und diese geglättet hat, schlägt die DAV vor, hohe Alter zu extrapolieren. Die Sterbetafel 2004R enthält in der Basistafel nur Alter bis 99. Die Alter 100 bis 120 werden mittels der Methodik aus [Ham00] extrapoliert. Dort werden eine Reihe von Modellen angegeben und ihre Güte verglichen. Hier eine Auswahl von Modellen, die [Ham00] vorschlägt und dann vergleicht:

- Gompertz-Modell: Das Gompertz-Modell wurde 1825 vorgeschlagen und ist das einfachste Modell dieser Liste. Die Sterbehäufigkeiten werden nach Gleichung (25) extrapoliert. Die Parameter α und β müssen aus den Altern Null bis 99 geschätzt werden.

$$q_x = \alpha \cdot \exp(\beta x) \quad (25)$$

Im Gompertz-Modell geht man davon aus, dass die Sterblichkeit eine einfache exponentielle Funktion des Alters ist. Gompertz entdeckte dieses Sterblichkeitsgesetz, indem er die Sterbewahrscheinlichkeiten der ihm zur Verfügung stehenden Sterbetafeln analysierte. Trotz seiner Einfachheit hat es in den vergangenen Jahrhunderten gute Vorhersagen geliefert.

- Makehams Gesetz: Makehams Gesetz (26) ist eine einfache Modifikation des Gompertz-Modells. Es wird ein konstanter, additiver Term γ auf das Gompertz-Modell addiert, um Risiken, die nicht vom Alter abhängen, zu berücksichtigen:

$$q_x = \gamma + \alpha \exp(\beta x) \quad (26)$$

- Logistisches Modell: Das logistische Modell (27) wurde 1923 von Perks vorgeschlagen und ist für $\alpha' = 0$ genau Makehams Gesetz:

$$q_x = \gamma + \frac{\alpha \exp(\beta x)}{1 + \alpha' \exp(\beta x)} \quad (27)$$

- Quadratisches Modell: Das Quadratische Modell (28) von Coale und Kisker (1990) geht von einem quadratischem Zusammenhang der logarithmierten Sterbewahrscheinlichkeiten aus:

$$\ln(q_x) = \alpha + \beta x + \gamma x^2 \quad (28)$$

Betrachten wir noch, welche Methoden zum Schätzen von Sterblichkeitstrends vorgeschlagen werden. Die DAV schlägt in der Dokumentation zur Sterbetafel 2004R [DAV04] im vierten Kapitel die folgenden drei Modelle vor: Das traditionelle Modell, das Kohortenmodell und das Synthesemodell. Hier die drei Modelle:

- traditionelles Modell:

$$\frac{q_{x,t+1}}{q_{x,t}} = \exp(-F(x))$$

- Kohortenmodell:

$$\frac{q_{x,t+1}}{q_{x,t}} = \exp(-G(t+1-x))$$

- Synthesenmodell:

$$\frac{q_{x,t+1}}{q_{x,t}} = \exp(-F(x) - G(t+1-x))$$

Laut Anhang 11 werden F beziehungsweise G mittels einer Methode der kleinsten Quadrate durch lineare Regression aus den Werten für $\ln(q_{x,t})$ für $t_0 \leq t \leq t_1$ geschätzt. Die Ergebnisse der Trendschätzung können in [DAV04, S. 49] gefunden werden. Als Sekundärliteratur zu diesen Modellen wird dort [Wil99] empfohlen. Da diese Quelle im Internet nicht zu finden ist, betrachten wir im folgenden ein Modell, das besser bekannt ist, oder zu dem zumindest einfacher Quellen zu finden sind. Es handelt sich um das Lee-Carter-Modell, welches wir in Abschnitt 3.4 genauer betrachten werden.

Die DAV schätzt dann alle Parameter für die Basistafel und dann wird auf Grund von Gütekriterien das beste Modell ausgewählt. Einige dieser Kriterien sind der Vergleich von gemäß extrapolierten Sterbewahrscheinlichkeiten erwarteten Totenzahlen mit tatsächlichen Totenzahlen, der Vergleich der Werte der Loglikelihoodfunktionen (ohne konstante Summanden) und der Vergleich der Werte der χ^2 -Statistik für die verschiedene Modelle. Für genauere Informationen zu den Gütekriterien siehe zum Beispiel [DAV04, S. 78]

Bei den Glättungsverfahren wird das Whittaker-Henderson-Verfahren empfohlen. Bezüglich der Extrapolation von hohen Altern wird das logistische Modell als das beste für die vorliegenden Daten ausgewählt. Für die Sterblichkeitstrends empfiehlt die DAV das Synthesenmodell [DAV04, S. 22]

3.4 Das Lee-Carter-Modell

Wir haben bereits in der Motivation in Kapitel 1 gesehen, dass die Sterblichkeit in der Gesamtbevölkerung in Deutschland in den letzten Jahren zurückgegangen ist. Deshalb wird in diesem Abschnitt ein statistisches Modell angegeben, um Sterblichkeiten und unterliegende Trends zu schätzen. Betrachten wir dazu zuerst das Lee-Carter-Modell. Dieses Modell wurde von Lee und Carter zur Modellierung und Vorhersage von Sterblichkeiten in den U.S.A. entwickelt. Laut [GK07] ist es mittlerweile zum “leading statistical model of mortality [forecasting] in the demographic literature” geworden. Dieses Zitat hat [GK07] aus [DP01]

entnommen. Für weitere Informationen, für welche Zwecke genau das Lee-Carter-Modell verwendet wurde, siehe die Einleitung zu [GK07].

In diesem Abschnitt 3.4 werden rohe Sterbehäufigkeiten der Art Gleichung (3) benutzt, die dann abgeändert werden.

$$\hat{q}_{x,t}^{roh} = \frac{\#T_{x,t}}{\#K_{x,t}} \quad (29)$$

Ein dem Lee-Carter-Modell sehr ähnliches Modell benutzt Bomsdorf [Bom93], um deutsche Sterblichkeiten zu schätzen. Sein Ansatz ist in der folgenden Gleichung angegeben:

$$q_x(t) = e^{a_x + b_x t}, \text{ für } x = 0, 1, \dots, 100$$

Bei diesem Ansatz bezeichnet x ein Alter, t eine Jahreszahl und q_x die einjährige Sterbewahrscheinlichkeit einer x -jährigen Person im Jahr t . Bei a_x und b_x handelt es sich um Parameter, die zu schätzen sind.

Die folgenden Beschreibungen zum Lee-Carter-Modell orientieren sich an [GK07]:

- Das Modell: Seien $m_{x,t}$ die logarithmierten Sterblichkeitsraten in den Altersgruppen $x \in \{S, S+1, \dots, E\}$ und zu den Beobachtungsjahren $t = 1, \dots, T$. Die Sterblichkeitsraten $m_{x,t}$ werden durch folgenden Ansatz modelliert:

$$m_{x,t} = \alpha_x + \beta_x \gamma_t + \epsilon_{x,t} \quad (30)$$

Hier sind α_x, β_x und γ_t Parameter, die geschätzt werden müssen. Bei $\epsilon_{x,t}$ handelt es sich um zufällige Störungen. Die Parametrisierung (30) ist nicht eindeutig, da sie unter den folgenden beiden Transformationen invariant ist:

- $\beta_x \rightarrow c\beta_x$ und $\gamma_t \rightarrow \frac{1}{c}\gamma_t$ für $c \in \mathbb{R}, c \neq 0$
- $\alpha_x \rightarrow \alpha_x - \beta_x c$ und $\gamma_t \rightarrow \gamma_t + c$ für $c \in \mathbb{R}$

Dies führt dazu, dass die Likelihood-Funktion, die zu Modell (30) gehört, mehr als ein Maximum hat. Deshalb werden die zwei Nebenbedingungen (31) und (32) eingeführt.

$$\sum_t \gamma_t = 0 \quad (31)$$

$$\sum_x \beta_x = 1 \quad (32)$$

Stellt man das Modell (30) nach γ_t um, erhält man den folgenden Ausdruck:

$$\gamma_t = \frac{m_{x,t} - \alpha_x - \epsilon_{x,t}}{\beta_x}$$

Setzt man diese Gleichung für γ_t in die Nebenbedingung (31) ein und benutzt man, dass die Fehler in den Altersgruppen x über die Zeit t im Erwartungswert Null sind und deshalb in der Rechnung als Null angenommen werden, erhält man:

$$\begin{aligned}
0 &= \sum_t \gamma_t = \sum_t \frac{m_{x,t} - \alpha_x - \epsilon_{x,t}}{\beta_x} \\
\iff 0 &= \sum_t (m_{x,t} - \alpha_x - \epsilon_{x,t}) \\
\iff \sum_t \alpha_x &= \sum_t m_{x,t} - \sum_t \epsilon_{x,t} \\
\iff \bar{m}_x &:= \frac{1}{t} \sum_t m_{x,t} = \alpha_x
\end{aligned}$$

Dadurch lässt sich das Modell (30) neu schreiben, indem man $\alpha_x = \bar{m}_x$ setzt und diesen Wert auf beiden Seiten abzieht. Mit der Definition $\tilde{m}_{x,t} = m_{x,t} - \bar{m}_x$ erhält man:

$$\tilde{m}_{x,t} = \beta_x \gamma_t + \epsilon_{x,t}$$

Weiterhin kann man Nebenbedingung (32) nach γ_t umstellen. Dazu stellt man zuerst das Modell (30) nach β_x um und erhält:

$$\beta_x = \frac{m_{x,t} - \alpha_x - \epsilon_{x,t}}{\gamma_t}$$

setzt man diese Charakterisierung für β_x nun in Nebenbedingung (32) ein, erhält man:

$$\begin{aligned}
1 &= \sum_x \beta_x = \sum_x \frac{m_{x,t} - \alpha_x - \epsilon_{x,t}}{\gamma_t} \\
\iff \gamma_t &= \sum_x (m_{x,t} - \alpha_x) = \sum_x (m_{x,t} - \bar{m}_x)
\end{aligned}$$

- Schätzung der Parameter: Wir haben bereits Charakterisierungen für α und für γ gesehen. Es bleibt, sich Gedanken über β zu machen. In [GK07] wird der folgende Ansatz vorgeschlagen: Zuerst bestimmt man die Matrix \tilde{M} für die Alter $\{0, \dots, 95\}$ und die Beobachtungsjahre $\{1956, \dots, 2017\}$ wie folgt:

$$\tilde{M} = \begin{pmatrix} \tilde{m}_{0,1956} & \dots & \tilde{m}_{0,2017} \\ \vdots & \vdots & \vdots \\ \tilde{m}_{95,1956} & \dots & \tilde{m}_{95,2017} \end{pmatrix} \quad (33)$$

Als nächstes macht man eine Singulärwertzerlegung der Matrix $\tilde{M} = BLU'$ und schätzt β als die erste Spalte von B . Hier muss man darauf achten, dass die Spalten der Matrix B Länge eins haben und die Singulärwerte in L absteigend sortiert sind. Ansonsten muss man β normieren und die Spalte aus B nehmen, die dem größten Singulärwert entspricht. Später wird in Abschnitt 4.3.1 der Maximum-Likelihood Schätzer $\hat{\beta}$ als folgende Identität bestimmt:

$$\hat{\beta}_x = \left(\sum_{t=1}^n \tilde{m}_{x,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1}$$

- Vorhersagen: Im Lee-Carter-Modell geht man davon aus, dass $\alpha = (\alpha_S, \dots, \alpha_E)$ und $\beta = (\beta_S, \dots, \beta_E)$ über die Zeit konstant bleiben. Das heißt, es genügt, γ_t für t außerhalb der Beobachtungsjahre zu schätzen.

Um aus diesem Modell zukünftige Werte für γ_t zu schätzen, schlagen Lee und Carter ein Random-Walk-with-Drift-Modell (RWD) vor:

$$\hat{\gamma}_t = \hat{\gamma}_{t-1} + \nu + \xi_t \quad (34)$$

Hier ist ν ein unbekannter Drift Parameter, der geschätzt werden muss, und $\xi_t \sim \mathcal{N}(0, \sigma_{\text{rw}}^2)$. Der Maximum-Likelihood-Schätzer für ν wird in Abschnitt 4.3.2 als folgende Identität bestimmt:

$$\hat{\nu} = (\hat{\gamma}_E - \hat{\gamma}_S) / (E - S)$$

Hier sind E, S das End-und Startalter der Sterbetafel.

- Bedeutung der Parameter: Die Parameter α, β, γ und ν werden einzeln besprochen.
 - α : Der α -Parameter ist auf Grund der Nebenbedingungen der Mittelwert der Sterblichkeiten über die Beobachtungsjahre. Die Abbildung 8 zeigt die Werte von α für das oben genannte Beispiel. Man kann diesen Parameter als die Sterblichkeit ohne Trends verstehen.

Betrachtet man Abbildung 8, sieht man, dass die Sterbewahrscheinlichkeit mit den Altern zuerst langsam und dann stark ansteigt. Die Sterbewahrscheinlichkeit für das Endalter beträgt 1.0886442632 und ist damit über 100 %. Insgesamt ist der Verlauf dieser Sterblichkeitsfunktion recht sinnvoll: Man betrachtet Periodendaten, das heißt, man betrachtet nur ein Beobachtungsjahr und in diesem Beobachtungsjahr die Verteilung der Tode. Man sieht, dass von den älteren Personen relativ mehr sterben als von den jungen Personen. Dies ist genau das, was man erwarten würde.

Außerdem veranschaulicht die Abbildung 8, dass ein höherer Wert von α zu höheren Sterbewahrscheinlichkeiten führt und ein niedrigerer Wert zu niedrigeren. Dies wird bei der Diskussion der Sterblichkeitsverbesserung wichtig.

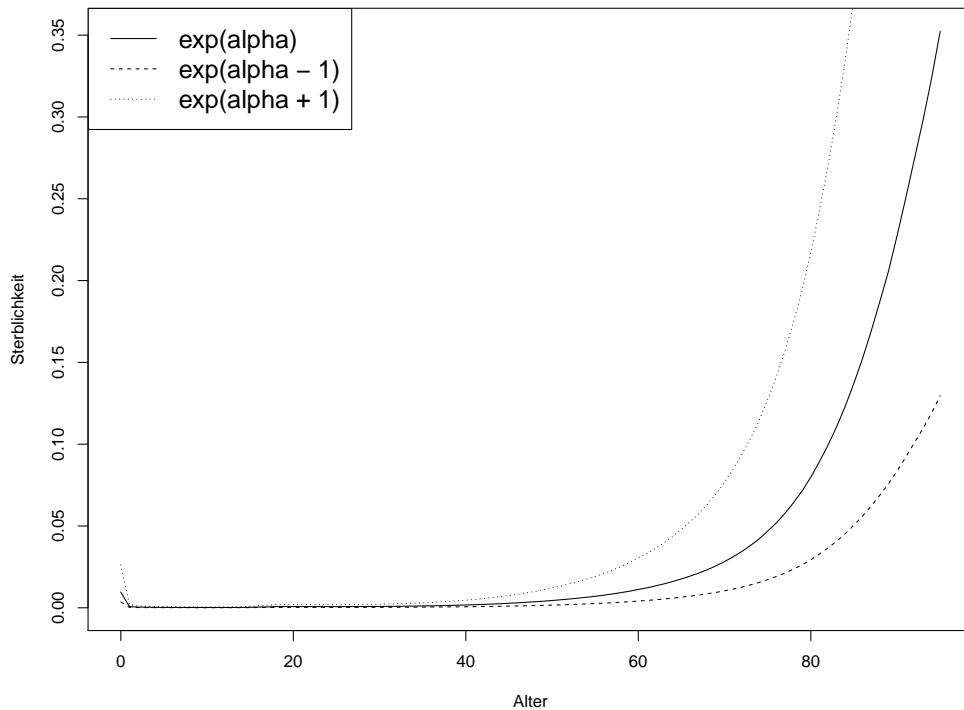


Abbildung 8: Lee-Carter-Schätzung für $\exp(\alpha)$ für westdeutsche Periodendaten 1956 bis 2017.

- β : Der β -Parameter gibt an, wie stark der Einfluss der Sterblichkeitsverbesserung ist, indem er den β -Parameter gewichtet. Die Gewichtungen, die man bei den Westdeutschen Periodendaten bekommt, ist wie in Abbildung 9 gezeigt. Man sieht, dass der Einfluss der Sterblichkeitsverbesserung zuerst stark ist, sich dann abschwächt, dann wieder steigt und schließlich wieder absinkt. Um den Einfluss auf die Sterblichkeitsverbesserung zu verstehen, muss man Beta und Gamma gemeinsam betrachten. Dies passiert später in Abbildung 11.

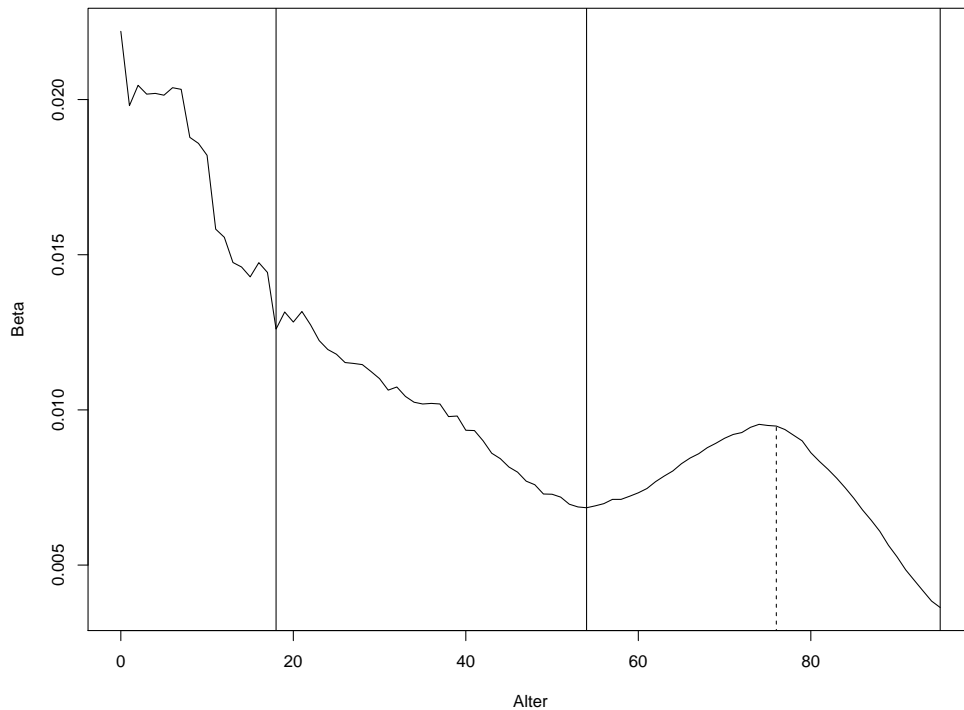


Abbildung 9: Lee-Carter-Schätzung für β für Westdeutsche Periodendaten 1956 bis 2017. Als optische Hilfe sind 4 Bereiche verschiedener Sterblichkeitsverbesserung eingetragen.

- γ : Der γ -Parameter gibt die Verbesserung beziehungsweise die Verschlechterung der Sterblichkeit an. Beides ist grundsätzlich möglich. In Abbildung 10 sieht man, wie sich γ_t im Laufe der Jahre entwickelt hat.

Man sieht, dass die Sterblichkeitsverbesserung ungefähr linear verläuft und von einem Wert von +60 zu einem Wert von -60 läuft. Da der β -Parameter immer positiv ist, gibt der γ -Parameter das Vorzeichen vor.

An dieser Stelle kann man versuchen, den etwas unintuitiven Verlauf der Beta-Alter-Funktion zu verstehen: Intuitiv würde man denken, dass die Sterblichkeitsverbesserung insbesondere Neugeborene und alte Personen beeinflusst.

Hier liegt allerdings eine sehr starke Verbesserung bei den unter 19-jährigen dann eine etwas langsamere Verbesserung bis zum Alter 54, dann einen Buckel bei einem Alter von 76 vor. Die Sterblichkeit der unter 19-jährigen lässt sich gut durch einen Rückgang der Kindersterblichkeit erklären. So sank laut [MD19] “die Kindersterblichkeit weltweit [...] von 19 % im Jahr 1960 auf 4 % im Jahr 2017”. Der Buckel zwischen 54 und 95 lässt sich damit erklären, dass in diesem Altersbereich der Großteil der sichtbaren Sterblichkeitsverbesserung passiert: Das mittlere Alter von Personen ist in diesem Bereich angestiegen und das deckt sich mit der Alltagserfahrung. Personen die früher nur 70 geworden wären werden

heute 80. Für den Bereich zwischen 19 und 54 bietet sich keine Erklärung direkt an. Ich vermute, dass es sich hierbei um Verbesserungen der Arbeitsbedingungen (Arbeitsunfälle) und bessere Verkehrssicherheit (Anschnallpflicht) handelt.

Es sei noch bemerkt, dass die Annahme $\sum_t \gamma_t = 0$ dazu führt, dass es ein Referenzjahr gibt. Das heißt, es gibt ein Jahr, in dem $\gamma_t \approx 0$ und in diesem Jahr gibt es weder eine Sterblichkeitsverbesserung noch eine Sterblichkeitsverschlechterung ($\gamma_t < 0$ beziehungsweise $\gamma_t > 0$).



Abbildung 10: Lee-Carter-Schätzung für γ für Westdeutsche Periodendaten 1956 bis 2017. Zur optischen Hilfe ist eine Linie auf Höhe der Null geplottet.

- ν : Der ν -Parameter wird nur für Vorhersagen gebraucht und gibt wieder, wie sich der γ -Parameter verändert. Schätzt man ν im Beispiel, erhält man den Wert -1.908261. Das heißt, von einem Jahr zum nächsten ändert sich Gamma um -1.908261. Berechnen wir also beispielsweise das Verhältnis der Sterblichkeit von einem Jahr zum nächsten und benutzen dabei den Mittelwert der Betas von 0.009009, so erhält man den folgenden Ausdruck:

$$\frac{e^{\alpha+\beta(\gamma-1.908261)}}{e^{\alpha+\beta\gamma}} = e^{\alpha-\alpha+\beta\gamma-\beta\gamma-\beta 1.908261} = e^{-1.908261\beta} = e^{-1.908261 \cdot 0.009009} \approx 0.98$$

Das heißt, die Sterblichkeit verringert sich pro Jahr im Mittel um zwei Prozent.

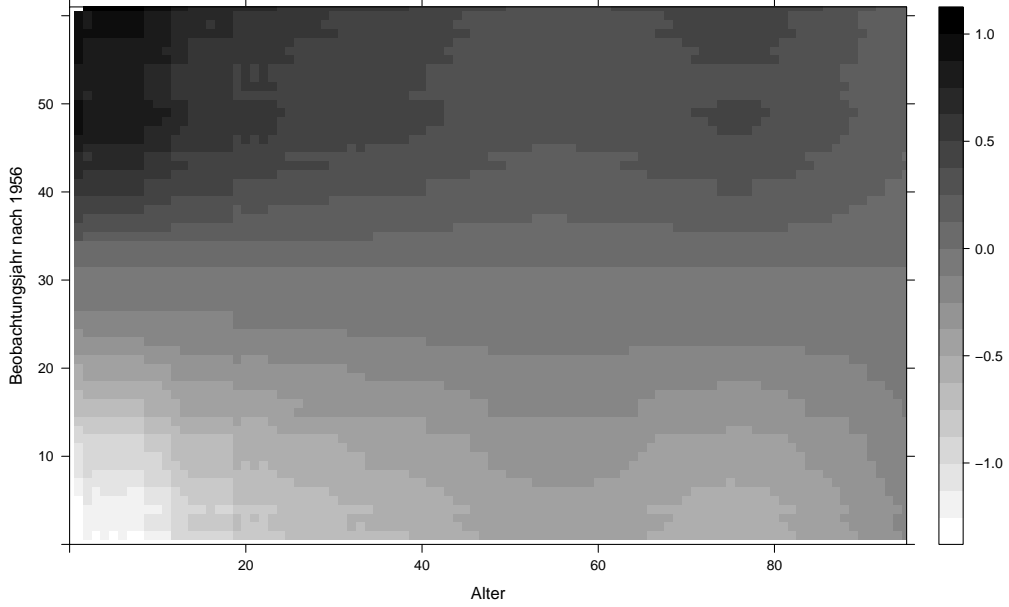


Abbildung 11: Heatmap für (γ_t, β_x)

In Abbildung 11 sieht man die Werte der Produkte der β -Parameter und der γ -Parameter. Das heißt, man sieht die Werte der Matrix (35):

$$\begin{pmatrix} \beta_0 \cdot \gamma_S & \dots & \beta_{95} \cdot \gamma_S \\ \vdots & \vdots & \vdots \\ \beta_0 \cdot \gamma_E & \dots & \beta_{95} \cdot \gamma_E \end{pmatrix} \quad (35)$$

Auf den ersten Blick ist die Abbildung 11 nicht sehr vielsagend. Man sieht, dass die Werte links unten klein und negativ sind und links oben groß und positiv. Außerdem sieht man bei einem Alter von 76 in der oberen Hälfte der Beobachtungsjahre einen positiven Buckel und bei den selben Altern in der unteren Hälfte einen negativen Buckel. Der Bereich in der Mitte ist wenig strukturiert. Um Abbildung 11 zu verstehen, hilft Abbildung (12).

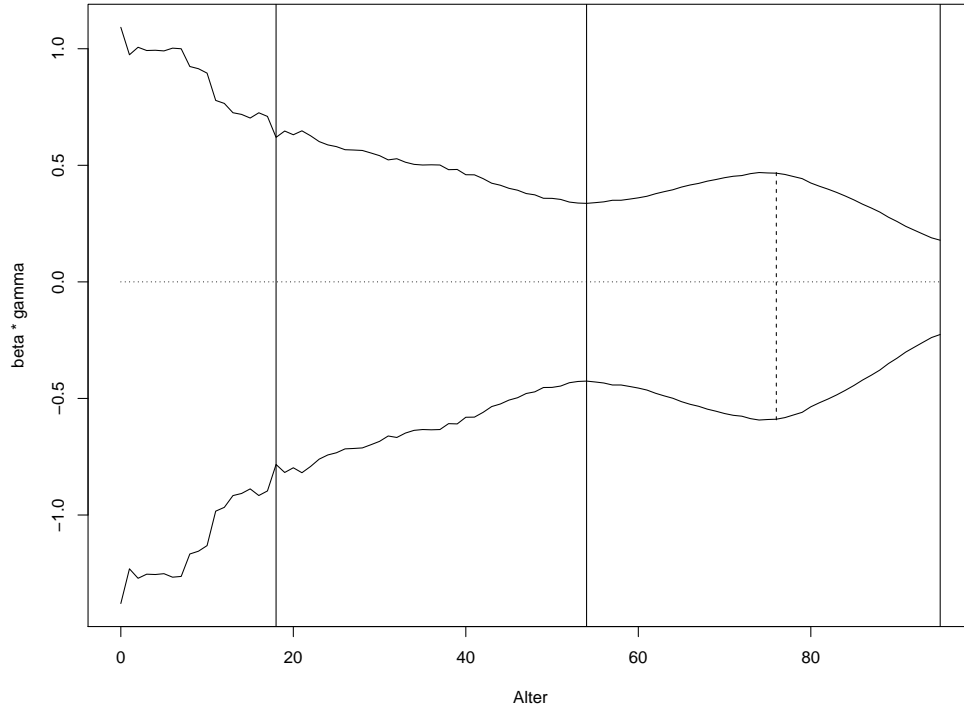


Abbildung 12: Obere Linie ist $\gamma_t \cdot \beta_x$ mit $t = 1956$ und die untere Linie ist $\gamma_t \cdot \beta_x$ mit $t = 2017$. Außerdem sind die Hilfslinien wie in Abbildung 9 mit eingetragen.

In Abbildung (12) sieht man die erste und die letzte Zeile aus der Matrix (35). Man sieht also $\gamma_t \cdot \beta_x$ mit $t \in \{1956, 2017\}$ und außerdem die Hilfslinien wie in Abbildung 9. Das heißt, man sieht die durchgezogene Linie bei den Altern 19, 54 und 95, sowie die Hilfslinie bei dem Buckel bei einem Alter von 76. Insgesamt sieht man, dass die obere und die untere Linie, die die Graphen von $\beta \cdot \gamma_{1956}$ und $\beta \cdot \gamma_{2017}$ zeigen, fast genau an der x -Achse gespiegelt sind. Konkret ist jeweils der erste Wert 1.0919484 beziehungsweise -1.3795982. Entscheidend ist, dass γ große positive bis große negative Werte durchläuft. Deshalb sieht man in Abbildung 11 oben und unten Struktur, aber in der Mitte nicht. Dadurch, dass γ dort Werte nahe bei Null annimmt, liegt die Struktur, die man an den Rändern sieht, in der selben Größenordnung und wird mit dem selben Grauton wiedergegeben. An den Rändern dagegen sind die Werte extremer und die Struktur aus Abbildung 12 beziehungsweise Abbildung 9 ist sichtbar.

3.5 Verbesserung des einfachen Modells

In Abschnitt 3.2 haben wir ein einfaches Modell gesehen, in dem die Sterblichkeit nur von dem Alter einer Person abhing. In Abschnitt 3.4 haben wir gesehen, wie das Lee-Carter-Modell sich verändernde Sterblichkeiten berücksichtigt. Es wird ein Random-Walk-with-

Drift-Modell (RWD) der Form (34) vorgeschlagen. Es liegt also nahe, dass man das einfache Modell aus Abschnitt 3.2 durch einen RWD erweitert, um sich verändernde Sterblichkeiten zu berücksichtigen. Das heißt, es wird von dem Modell (36), (37) ausgegangen.

$$Q = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m_t - X)^2}{2 \cdot s^2}\right) + \epsilon \quad (36)$$

$$m_t = m_{t-1} + \nu + \xi_t \quad (37)$$

In Modell (37) ist ν der unbekannte Driftparameter und ξ_t ein Fehler. Schätzt man die μ_t für den Datensatz der Generationssterbetafel erhält man Abbildung 13. Man sieht, wie in Abschnitt 3.1, dass sich das mittlere Todesalter nicht so entwickelt, wie man es erwarten würde. Man würde erwarten, dass es im Laufe der Jahre steigt, aber stattdessen sinkt es. Dies ist wieder ein Problem, welches durch die Zensur der Daten erklärbar ist. Man hat die späten Todesfälle von Personen, die erst vor kurzem geboren sind, noch nicht beobachtet. Entsprechend kann man keine Aussage über ihr erreichtes Alter machen. Das heißt, um dieses Kapitel zusammen zu fassen, der Random Walk with Drift, den man im Lee-Carter-Modell benutzen kann, um die Sterblichkeit in der Zukunft vorherzusagen, ist im einfachen Modell nicht anwendbar, was wieder an der Zensur der Daten liegt.



Abbildung 13: Man sieht, $\hat{\mu}_t$ gegen die Geburtsjahre aufgetragen.

4 Eigenschaften der Modelle und Simulationen

Das Ziel dieses Kapitels ist es, die folgenden Fragen für die vorgestellten Modelle zu beantworten:

- Ist das Modell identifizierbar?
- Ist der Schätzer konsistent?
- Was ist die asymptotische Verteilung?

Alle diese Eigenschaften kann man mit dem Theorem von Cramér zeigen. Leider werden wir im ersten Abschnitt 4.1 dieses Kapitels sehen, dass man das Theorem von Cramér nur schwer auf das einfache Modell anwenden kann. Deshalb werden andere Voraussetzungen als die von dem Theorem von Cramér zusammengestellt. Diese werden dann für das einfache Modell in Abschnitt 4.2 gezeigt. Weiterhin kann man für das Whittaker-Henderson-Verfahren wenig Eigenschaften zeigen, da es sich nicht um ein statistisches Modell, sondern um ein numerisches Verfahren handelt. Dieses Modell wird deshalb nicht näher betrachtet. Das Lee-Carter-Modell macht wenig Probleme und man kann das Theorem von Cramér direkt anwenden. Die Rechnungen für das Lee-Carter-Modell werden in Abschnitt 4.3 ausgeführt. Die Verbesserung des einfachen Modells hat keine guten Ergebnisse in der Modellbildung geliefert und wird deshalb nur kurz in Abschnitt 4.4 besprochen. Am Ende des Kapitels werden in Abschnitt 4.5 noch einige Simulationen vorgestellt, mit denen die Modelle verglichen werden.

4.1 Das Theorem von Cramér

Wir wollen von den bisher betrachteten Modellen unter anderem Identifizierbarkeit zeigen. Diese Eigenschaft ist wie folgt definiert:

Definition 7 (Identifizierbarkeit). Sei $\Theta \subset \mathbb{R}^n$. Wir nennen Θ einen Parameterraum. Sei P_θ ein Wahrscheinlichkeitsmaß, dass von θ abhängt. Wir bezeichnen eine Menge von P_θ mit $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ und nennen \mathcal{P} ein statistisches Modell. Ein statistisches Modell \mathcal{P} heißt identifizierbar genau dann, wenn \mathcal{P} folgende Eigenschaft besitzt:

$$P_{\theta_1} = P_{\theta_2} \implies \theta_1 = \theta_2 \quad (38)$$

Die anderen beiden Eigenschaften, die uns interessieren sind Konsistenz und die asymptotische Verteilung. Das Theorem von Cramér gibt Bedingungen für diese Eigenschaften:

Theorem 1 (Cramér). Sei $\mathbf{X} = (X_1, \dots, X_n)$ mit X_1, \dots, X_n unabhängig und identisch verteilt mit einer Dichte $f(x|\theta)$ und sei θ_0 der wahre Parameter. Eine Realisierung von \mathbf{X} ist dann ein Sample aus einer parametrischen Familie $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$, wobei P_θ die Dichte $f(x|\theta) := f_\theta$ bezüglich des Lebesguemaßes hat. Seien weiterhin die folgenden fünf Eigenschaften erfüllt:

1. Die Menge Θ ist eine offene Teilmenge des \mathbb{R}^n .
2. Das Modell ist identifizierbar.
3. Die zweiten Ableitungen von f_θ nach θ_i existieren, sind stetig für alle $x \in \{S, \dots, E\}$ und können unter das Integral $\int f_\theta(x) dx$ geschoben werden.
4. Es existiert eine integrierbare Funktion $K(x)$, sodass $\mathbb{E}_{\theta_0} K(x) < \infty$ und jede Komponente von $|\frac{d^2}{d\theta^2} \log(f_\theta(x))| \leq K(x)$ für alle θ mit $|\theta - \theta_0| < \epsilon$.
5. Die Fisher-Informationsmatrix $i(\theta_0) := -\mathbb{E}(\frac{d^2}{d\theta^2} \log(f_{\theta_0}(x)))$ ist positiv definit.

Dann existiert eine stark konsistente Folge $\hat{\theta}_n$ von Lösungen der Gleichungen (39) mit der Eigenschaft $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0_k, i(\theta_0)^{-1})$, die als asymptotische Normalität bezeichnet wird.

$$n^{-1} \sum_{i=1}^n \frac{d}{d\theta_i} \log(f_\theta(x)) = 0 \quad (39)$$

Beweis. Für den vollständigen Beweis siehe [Fer02], Theorem 18. □

Man beachte, dass das Theorem von Cramér nur aussagt, dass ein kritischer Wert der Likelihoodfunktion stark konsistent und asymptotisch normalverteilt ist. Deshalb wird in dem Kapitel 4.3, in dem das Theorem von Cramér benutzt wird, gezeigt, dass es nur eine Nullstelle gibt.

Das größere Problem ist, dieses Theorem überhaupt erst einmal zu zeigen. Die Eigenschaften 1 kann man durch eine passende Definition von Θ gewährleisten. Eigenschaft 2 folgt für die betrachteten Modelle durch direkte Rechnung. Die Glattheit in Eigenschaft 3 ist auch kein Problem, da wir von normalverteilten Fehlern ausgehen. Das Problem ist die Vertauschbarkeit von Integration und Differenzierbarkeit für das einfache Modell zu zeigen. Für diese Eigenschaft gibt es bei Van der Vaart, [Vaa98], das folgende Theorem:

Theorem 2. Die Funktion $\theta \mapsto \int h(x) e^{\theta^T t(x)} d\mu(x)$ ist analytisch auf der Menge $\{\theta \in \mathbb{C}^k : \text{Re}(\theta) \in \Theta^\circ\}$. Die Ableitungen dieser Funktion kann man bestimmen, indem man unter dem Integralzeichen ableitet:

$$\frac{d^p \int h(x) e^{\theta^T t(x)} d\mu(x)}{d\theta_1^{i_1} \dots d\theta_k^{i_k}} = \int h(x) t_1(x)^{i_1} \dots t_k(x)^{i_k} e^{\theta^T t(x)} d\mu(x),$$

für $p \in \mathbb{N}$ und $i_1 + \dots + i_k = p$

Beweis. Dieses Theorem ist Lemma 4.5 in [Vaa98, S. 45]. □

Man beachte, dass die Funktion $\theta \mapsto \int h(x) e^{\theta^T t(x)} d\mu(x)$ genau der Dichte einer Exponentialfamilie in natürlicher Parametrisierung entspricht. Später zeigt Van der Vaart, dass das selbe Ergebnis immer noch gilt, wenn man statt θ eine Transformation $Q(\theta)$ hat. Allerdings muss Q eine bijektive Abbildung sein. Definieren wir als nächstes den Begriff einer natürlichen Exponentialfamilie, um zu untersuchen, ob das einfache Modell und beziehungsweise oder das Lee-Carter-Modell dieser Definition entspricht:

Definition 8 (Exponentialfamilie). Eine Familie von Wahrscheinlichkeitsmaßen $(P_\theta)_{\theta \in \Theta}$ auf einem Messraum (X, A) mit $\Theta \subset \mathbb{R}^k$ heißt eine k -parametrische Exponentialfamilie, wenn es ein σ -endliches Maß μ gibt, so dass alle P_θ eine Dichte der Form (40) bezüglich μ besitzen.

$$f(y, \theta) = h(y)A(\theta)\exp\left(\eta(\theta)^t T(y)\right) \quad (40)$$

Dabei sind h und T messbare Funktionen und $A_i, \eta_i : \mathbb{R}^k \supset \Theta \mapsto \mathbb{R}$ Funktionen des k -dimensionalen Parameters θ . Weiterhin spricht man von einer natürlichen Exponentialfamilie, wenn $\eta_i(\theta) = \theta_i$. Man beachte, dass der Parameter der Dichte y und nicht wie üblich x ist, da in dem Zusammenhang mit Regression x ein fester Vektor ist und y der zufällige Wert. Die Notation mit y statt x soll dies unterstreichen.

Betrachten wir also das einfache Modell. Das Ziel ist, die Dichte des Modells in die Dichte einer Exponentialfamilie umzuformen:

$$\begin{aligned} f_{m,s}(y) &= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - y\right)^2}{2\sigma_\epsilon^2}\right) \\ &=: \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\frac{1}{s}h\left(\frac{m-x}{s}\right)\right)^2 - 2\frac{1}{s}h\left(\frac{m-x}{s}\right)y + y^2}{2\sigma_\epsilon^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{y^2}{2\sigma_\epsilon^2}\right) \cdot \exp\left(\frac{\frac{1}{s}h\left(\frac{m-x}{s}\right)y}{\sigma_\epsilon^2}\right) \cdot \exp\left(-\frac{\left(\frac{1}{s}h\left(\frac{m-x}{s}\right)\right)^2}{2\sigma_\epsilon^2}\right) \\ &=: h(y) \cdot \exp(\eta(m, s)T(y)) \cdot A(m, s) \end{aligned}$$

Man sieht, dass es sich in der Tat um eine Exponentialfamilie handelt. Es handelt sich nicht um eine natürliche Exponentialfamilie, da $\eta(\theta) \in \mathbb{R}$, während $\theta \in \mathbb{R}^2$. Wäre $\eta_i(\theta) = \theta_i$ für $i = 1, 2$, dann wäre $\theta_1 = \theta_2$. Im Allgemeinen ist jedoch $\theta_1 \neq \theta_2$ und deshalb erhält man einen Widerspruch zur Eigenschaft der natürlichen Exponentialfamilie. Das einfache Modell ist also keine natürliche Exponentialfamilie. Untersuchen wir als nächstes die beiden Teile des Lee-Carter-Modells. Wir beginnen mit dem Schätzer für β :

$$\begin{aligned} f(\tilde{m}_{x,t} | \beta_x, \sigma_{\epsilon_{x,t}}) &= \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t)^2 - 2\beta_x \hat{\gamma}_t \tilde{m}_{x,t} + \tilde{m}_{x,t}^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t)^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \cdot \exp\left(\frac{\beta_x \hat{\gamma}_t \tilde{m}_{x,t}}{\sigma_{\epsilon_{x,t}}^2}\right) \cdot \exp\left(-\frac{\tilde{m}_{x,t}^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\ &=: h(\tilde{m}_{x,t})A(\beta)\exp(\eta(\beta)T(\tilde{m}_{x,t})) \end{aligned}$$

Man erhält ohne Probleme eine parametrische Familie, die bei passender Wahl von T auch natürlich ist. Das gleiche Ergebnis erhält man für die Schätzung von ν :

$$\begin{aligned}
f(\hat{\gamma}_t, \nu, \sigma_{\xi_t}^2) &= \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(-\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1} - \nu)^2}{2\sigma_{\xi_t}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(-\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1})^2 - 2(\hat{\gamma}_t - \hat{\gamma}_{t-1})\nu + \nu^2}{2\sigma_{\xi_t}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(-\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1})^2}{2\sigma_{\xi_t}^2}\right) \cdot \exp\left(-\frac{\nu^2}{2\sigma_{\xi_t}^2}\right) \cdot \exp\left(\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1})\nu}{\sigma_{\xi_t}^2}\right) \\
&= h(\hat{\gamma}_t)A(\nu)\exp(\eta(\nu)T(\hat{\gamma}_t))
\end{aligned}$$

Wir werden in Abschnitt 4.2 zeigen, dass auch die anderen Eigenschaften des Theorems von Cramér ohne Probleme zu zeigen sind. Gehen wir also wieder zu dem einfachen Modell zurück. Ausgehend von der Repräsentation als Exponentialfamilie liegt es nahe, durch konkretes Nachrechnen die Vertauschbarkeit von Integration und Differentiation der partiellen Ableitungen zu zeigen. Wir wollen also unter anderem die folgende Gleichung zeigen:

$$\int \frac{d^2}{dm^2} f(y|\theta) dy = \frac{d^2}{dm^2} \int f(y|\theta) dy$$

Betrachten wir zuerst die rechte Seite und benutzen, dass $f(y|\theta) = h(y)A(\theta)\exp(\eta(\theta)y)$:

$$\begin{aligned}
\int f(y|\theta) dy &= \int h(y)A(\theta)\exp(\eta(\theta)y) dy \\
&= A(\theta) \int \exp\left(-\frac{y^2}{2\sigma_\epsilon^2}\right) \exp(\eta(\theta)y) dy \\
&=: A(\theta) \int \exp(ay - by^2) dy
\end{aligned}$$

Leider gibt es für Funktionen der Form $\exp(ay - by^2)$ keine allgemeine Stammfunktion, weshalb dieser Ansatz nicht weiter verfolgt wird. Allerdings gibt es die Möglichkeit, den Maximum-Likelihood Schätzer für (m, s) als Zero-Schätzer zu betrachten. Zero-Schätzer sind definiert als Nullstellen von Schätzgleichungen. Genauer gesagt nennt man einen Schätzer $\hat{\theta}$ Zero-Schätzer, wenn er die folgende Gleichung löst:

$$\Psi_n(\theta) = \frac{1}{n} \sum_i^n \psi_\theta(y_i) = 0 \quad (41)$$

In Van der Vaart [Vaa98] sind eine Reihe von Theoremen angegeben, mit denen man Konsistenz und asymptotische Normalität von Zero-Schätzern zeigen kann. Das Problem ist, dass viele Theoreme für Konsistenz fordern, dass der Parameterraum kompakt ist, während viele Theoreme für asymptotische Normalität fordern, dass das Modell quadratic

mean differentiable (q.d.m.) ist. Die Eigenschaft q.d.m. zu sein, zeigt wird in Theorem 12.2.2 [LR05, S. 498] gezeigt. Eine Voraussetzung für Theorem 12.2.2 ist, dass der Parameterraum offen ist. Deshalb werden Lemma 5.10 (S.64) und Theorem 5.39 (S.82) aus Van der Vaart verwendet. Für die Konsistenz wird eine Variante von Lemma 5.10 in Verbindung mit Lemma 2.2 aus [EM86, S. 14] benutzt. Insgesamt erhält man folgendes Theorem, bei dem die Bedingungen 1 bis 6 für die Konsistenz verwendet werden, Bedingungen 7,8,9 für q.m.d. und Bedingungen 9,10 für die asymptotische Normalität:

Theorem 3. Bezeichne θ_0 den wahren Parameter und mögen die folgenden Eigenschaften erfüllt sein:

1. Das Modell ist identifizierbar.
2. Es ist $\mathbb{E}(|\ln(f(x|\theta))|) < \infty$.
3. Der Parameterraum $\Theta \subset \mathbb{R}^n$ ist ein möglicherweise offenes Rechteck.
4. Die Ableitungen der Log-Likelihoodfunktion konvergieren in Wahrscheinlichkeit gegen ihren Erwartungswert.
5. Die Ableitungen der Log-Likelihoodfunktion sind stetig in θ .
6. Der Punkt θ_0 erfüllt $\frac{d}{d\theta} l(\theta_0 - \epsilon) < 0 < \frac{d}{d\theta} l(\theta_0 + \epsilon)$ für alle $\epsilon > 0$ und $\theta = \theta_1, \theta_2, \dots$
7. Der Parameterraum Θ ist offen.
8. Das Modell hat ein Maß P_θ mit Dichte p_θ , die stetig differenzierbar in θ und in x ist.
9. Die Fisher-Informationsmatrix $i(\theta)$ existiert, ist nicht singulär und ist komponentenweise stetig in θ .
10. Es existiert eine Funktion g mit $\mathbb{E}(g^2) < \infty$ und

$$|l_{m_1, s_1}(y) - l_{m_2, s_2}(y)| \leq g(y) \|(m_1, s_1) - (m_2, s_2)\|$$

Dann ist das Modell konsistent und asymptotisch normal. Das heißt, es gelten die folgenden beiden Eigenschaften:

$$\begin{aligned} \hat{\theta}_n &\xrightarrow{P} \theta_0 \text{ für } n \rightarrow \infty \\ \sqrt{n}(\hat{\theta}_n - \theta_0) &\xrightarrow{D} Z \sim N(0, i(\theta_0)^{-1}) \text{ für } n \rightarrow \infty \end{aligned}$$

Beweis. Für die Konsistenz siehe [EM86] Theorem 2.7 (S.23) und Lemma 2.2 (S.14). Für die asymptotische Normalität siehe [Vaa98] Theorem 5.39 (S.82) in Verbindung mit [LR05] Theorem 12.2.2 (S.498). \square

4.2 Eigenschaften des einfachen Modells

Das Ziel dieses Abschnitts ist, die Voraussetzungen für das Theorem 3 für das einfache Modell mit dem Alter als erklärende Variable zu zeigen. Dazu müssen wir noch eine Annahme bezüglich des Fehlers machen. Wir gehen davon aus, dass $\epsilon_i \sim N(0, \sigma_\epsilon)$ für $i = S, \dots, E$ unabhängig und identisch (u.i.v.) verteilt sind. Weiterhin hängt für diese asymptotischen Betrachtungen q_x von der Anzahl beobachteter Personen ab. Diese Sterbehäufigkeiten kann man als Summe von Indikatorfunktionen auffassen, die so gewählt sind, dass man einen Zero-Schätzer erhält.

Zuerst wird die Likelihood-Funktion aufgestellt und die Maximum-Likelihood Schätzer \hat{m}, \hat{s} für m, s bestimmt. Danach werden die Voraussetzungen für das Theorem 3 gezeigt. Betrachten wir zuerst das einfache Modell mit einer allgemeinen Funktion h . Dieses Modell ist durch (42) gegeben. Die zugehörige Dichte ist (43) und die zugehörige Log-Likelihoodfunktion ist (44).

$$Q \sim N\left(\frac{1}{s} \cdot h\left(\frac{m-X}{s}\right), \sigma_\epsilon^2\right) \quad (42)$$

$$f_{(m,s)}(q_S, \dots, q_E) = \prod_{x=S}^E \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \cdot \exp\left(-\frac{\left(\frac{1}{s} \cdot h\left(\frac{m-x}{s}\right) - q_x\right)^2}{2\sigma_\epsilon^2}\right) \quad (43)$$

$$l_{(q_S, \dots, q_E)}(m, s) = \sum_{x=S}^E \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{s} \cdot h\left(\frac{m-x}{s}\right) - q_x\right)^2}{2\sigma_\epsilon^2} \right\} \quad (44)$$

Die ersten Ableitungen der Log-Likelihoodfunktion (44) sind durch die folgenden Gleichungen gegeben.

$$\begin{aligned} \frac{d}{dm} l_{(q_S, \dots, q_E)}(m, s) &= - \sum_{x=S}^E \frac{2 \left(\frac{1}{s} \cdot h\left(\frac{m-x}{s}\right) - q_x \right) \cdot \frac{1}{s^2} h'\left(\frac{m-x}{s}\right)}{2\sigma_\epsilon^2} \\ \frac{d}{ds} l_{(q_S, \dots, q_E)}(m, s) &= - \sum_{x=S}^E \frac{2 \left(\frac{1}{s} \cdot h\left(\frac{m-x}{s}\right) - q_x \right) \left\{ -\frac{1}{s^2} \cdot h\left(\frac{m-x}{s}\right) - \frac{1}{s} \cdot h'\left(\frac{m-x}{s}\right) \frac{m-x}{s^2} \right\}}{2\sigma_\epsilon^2} \end{aligned}$$

Die Berechnung der zweiten Ableitungen führt zu längeren Ausdrücken. In den folgenden drei Gleichungssystemen wird der Zähler der Summanden berechnet berechnet.

$$\begin{aligned}
\frac{d^2}{dm^2} \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right)^2 &= \frac{d}{dm} \left\{ 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) \right\} \\
&= \left\{ \frac{d}{dm} 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \right\} \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) + \dots \\
&\quad \dots 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \left\{ \frac{d}{dm} \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) \right\} \\
&= 2 \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) + \dots \\
&\quad \dots 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \frac{1}{s^3} h'' \left(\frac{m-x}{s} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{ds dm} \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right)^2 &= \frac{d}{ds} \left\{ 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) \right\} \\
&= \left\{ \frac{d}{ds} 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \right\} \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) + \dots \\
&\quad \dots 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \left\{ \frac{d}{ds} \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) \right\} \\
&= 2 \left\{ -\frac{1}{s^2} h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} h' \left(\frac{m-x}{s} \right) \right\} \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) + \dots \\
&\quad \dots 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \dots \\
&\quad \dots \left\{ -\frac{2}{s^3} h' \left(\frac{m-x}{s} \right) - \frac{m-x}{s^4} h'' \left(\frac{m-x}{s} \right) \right\}
\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{ds^2} \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right)^2 &= \frac{d}{ds} \left[2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \right. \\
&\quad \cdot \left. \left\{ -\frac{1}{s^2} \cdot h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} \cdot h' \left(\frac{m-x}{s} \right) \right\} \right] \\
&= \left\{ \frac{d}{ds} 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \right\} \cdot \dots \\
&\quad \dots \left\{ -\frac{1}{s^2} \cdot h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} \cdot h' \left(\frac{m-x}{s} \right) \right\} + \dots \\
&\quad \dots 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \dots \\
&\quad \dots \frac{d}{ds} \left\{ -\frac{1}{s^2} \cdot h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} \cdot h' \left(\frac{m-x}{s} \right) \right\} \\
&= \left[2 \left\{ -\frac{1}{s^2} h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} h' \left(\frac{m-x}{s} \right) \right\} \cdot \dots \right. \\
&\quad \left. \dots \left\{ -\frac{1}{s^2} \cdot h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} \cdot h' \left(\frac{m-x}{s} \right) \right\} \right] + \dots \\
&\quad \dots \left[\left\{ 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \right\} \cdot \dots \right. \\
&\quad \dots \left\{ \frac{2}{s^3} h \left(\frac{m-x}{s} \right) + \frac{m-x}{s^4} h' \left(\frac{m-x}{s} \right) + \dots \right. \\
&\quad \left. \left. \dots \frac{3(m-x)}{s^4} h' \left(\frac{m-x}{s} \right) + \frac{(m-x)^2}{s^5} h'' \left(\frac{m-x}{s} \right) \right\} \right]
\end{aligned}$$

Man erhält also insgesamt die folgenden drei zweiten Ableitungen:

$$\begin{aligned}
\frac{d^2}{dm^2} l_{(q_S, \dots, q_E)}(m, s) &= - \sum_{x=S}^E \frac{2 \left\{ \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) \right\}^2 + 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \frac{1}{s^3} h'' \left(\frac{m-x}{s} \right)}{2\sigma_\epsilon^2} \\
\frac{d^2}{dm ds} l_{(q_S, \dots, q_E)}(m, s) &= - \frac{1}{2\sigma_\epsilon^2} \sum_{x=S}^E \left[2 \left\{ -\frac{1}{s^2} h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^3} h' \left(\frac{m-x}{s} \right) \right\} \cdot \frac{1}{s^2} h' \left(\frac{m-x}{s} \right) + \dots \right. \\
&\quad \dots 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \cdot \dots \\
&\quad \left. \dots \left\{ -\frac{2}{s^3} h' \left(\frac{m-x}{s} \right) - \frac{m-x}{s^4} h'' \left(\frac{m-x}{s} \right) \right\} \right] \\
\frac{d^2}{ds^2} l_{(q_S, \dots, q_E)}(m, s) &= - \frac{1}{2\sigma_\epsilon^2} \sum_{x=S}^E \left[2 \left\{ -\frac{1}{s^2} h \left(\frac{m-x}{s} \right) - \frac{m-x}{s^2} h' \left(\frac{m-x}{s} \right) \right\}^2 + \dots \right. \\
&\quad \dots \left\{ 2 \left(\frac{1}{s} \cdot h \left(\frac{m-x}{s} \right) - q_x \right) \right\} \cdot \dots \\
&\quad \left. \dots \left\{ \frac{2}{s^3} h \left(\frac{m-x}{s} \right) + \frac{4(m-x)}{s^4} h' \left(\frac{m-x}{s} \right) + \frac{(m-x)^2}{s^5} h'' \left(\frac{m-x}{s} \right) \right\} \right]
\end{aligned}$$

Betrachten wir jetzt wieder das konkrete einfache Modell mit h wie in Gleichung (45). Hier gibt es zwei Möglichkeiten, um die Ableitungen des konkreten Modells (48) zu bestimmen. Entweder kann man die Ableitungen direkt ausrechnen oder die obigen Ableitungen benutzen und die entsprechenden Ableitungen (45), (46) und (47) von h einsetzen.

$$h(x) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) \quad (45)$$

$$\frac{d}{dx} h(x) = -\frac{x}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) \quad (46)$$

$$\frac{d^2}{dx^2} h(x) = \frac{(x^2 - 1)}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) \quad (47)$$

Hier wird der erste Ansatz verfolgt, während der zweite nur zur Kontrolle dient. Das Modell ist durch die Gleichung (48) gegeben und besitzt die Dichte (49) und Log-Likelihoodfunktion (50).

$$Q \sim N \left(\frac{1}{\sqrt{2\pi s^2}} \cdot \exp \left(-\frac{(m-X)^2}{2 \cdot s^2} \right), \sigma_\epsilon^2 \right) \quad (48)$$

$$f_{(m,s)}(q_S, \dots, q_E) = \prod_{x=S}^E \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp \left(-\frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp \left(-\frac{(m-x)^2}{2s^2} \right) - q_x \right)^2}{2\sigma_\epsilon^2} \right) \quad (49)$$

$$l_{(q_S, \dots, q_E)}(m, s) = \sum_{x=S}^E \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp \left(-\frac{(m-x)^2}{2s^2} \right) - q_x \right)^2}{2\sigma_\epsilon^2} \right\} \quad (50)$$

Um das Theorem von Cramér zu zeigen, brauchen wir die zweite Ableitung der Log-Likelihoodfunktion. Beginnen wir damit, die ersten Ableitungen der Log-Likelihoodfunktion zu bestimmen. Berechnen wir zuerst die Ableitung in m -Richtung:

$$\begin{aligned}
& \frac{d}{dm} l_{(q_S, \dots, q_E)}(m, s) \\
&= \frac{d}{dm} \sum_{x=S}^E \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right)^2}{2\sigma_\epsilon^2} \right\} \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{x=S}^E \frac{d}{dm} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right)^2 \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{x=S}^E 2 \cdot \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \frac{d}{dm} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \\
&= -\frac{1}{2\sigma_\epsilon^2} \cdot \frac{2}{\sqrt{2\pi s^2}} \sum_{x=S}^E \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \frac{d}{dm} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi s^2}} \sum_{x=S}^E \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \frac{d}{dm} \left(-\frac{(m-x)^2}{2s^2} \right) \\
&= \frac{1}{\sigma_\epsilon^2 2s^2 \sqrt{2\pi s^2}} \sum_{x=S}^E \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot 2(m-x) \\
&= \frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}} \sum_{x=S}^E \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x)
\end{aligned}$$

Es ist \hat{m} das Maximum der Likelihoodfunktion. Das heißt, \hat{m} ist ein kritischer Wert der Likelihoodfunktion und somit eine Nullstelle der ersten Ableitung der Log-Likelihoodfunktion in m -Richtung. Diese Nullstellen sind analytisch nicht zu bestimmen. Weiterhin sind die Nullstellen invariant unter Multiplikation mit einer positiven Konstante. Plotten wir also die erste Ableitung der Log-Likelihoodfunktion in m -Richtung ohne den vorstehenden Faktor $\frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}}$ und mit $s = 1$, um zu sehen, ob es Nullstellen gibt. Diese Funktion ist in Abbildung 14 zu sehen.

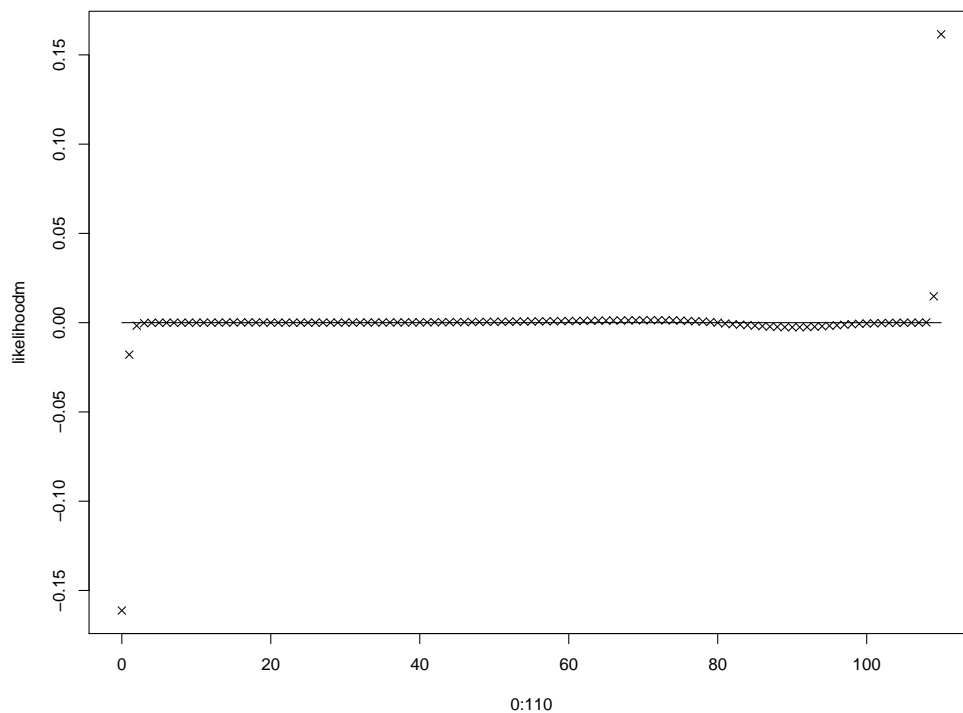


Abbildung 14: Die erste Ableitung der Log-Likelihoodfunktion in m -Richtung. Auf der x -Achse mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht, dass die Werte an den Rändern sehr groß sind. Deshalb ist in Abbildung 15 dieselbe Funktion noch einmal geplottet, diesmal allerdings ohne die Ränder:

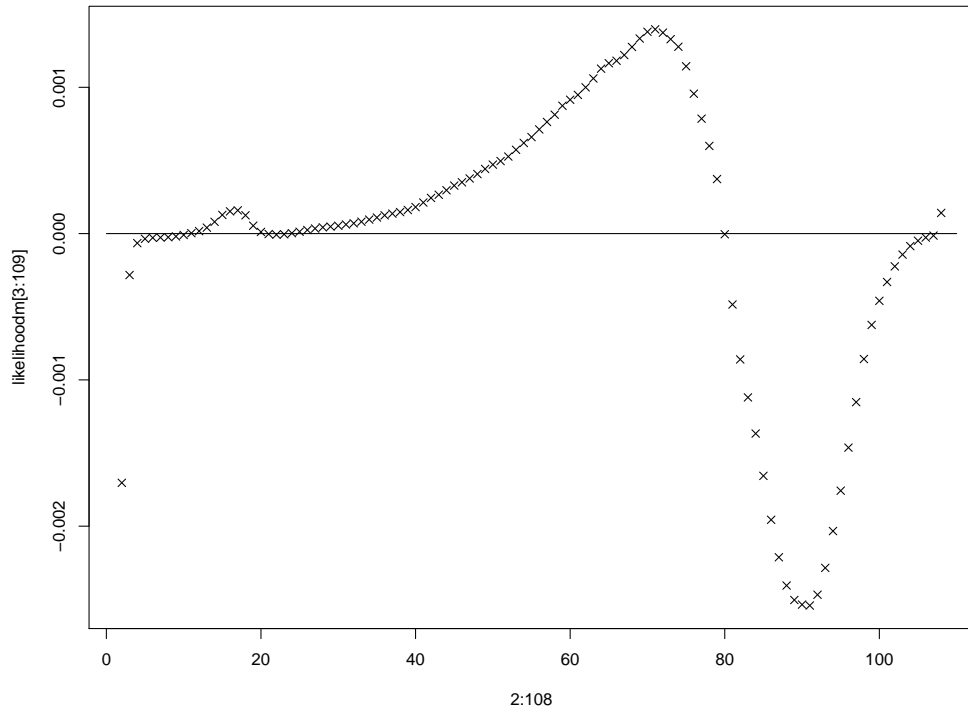


Abbildung 15: Die erste Ableitung der Log-Likelihoodfunktion in m -Richtung ohne die Randwerte. Auf der x -Achse sind wieder mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht vier mögliche Nullstellen. Die Nullstelle bei 80 scheint am sinnvollsten, wenn man sich die Daten aus Abbildung 7 anschaut. Berechnen wir jetzt die erste Ableitung in s^2 -Richtung:

$$\begin{aligned}
& \frac{d}{ds^2} l(q_S, \dots, q_E)(m, s) \\
&= \frac{d}{ds^2} \sum_{x=S}^E -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x\right)^2}{2\sigma_\epsilon^2} \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{x=S}^E \frac{d}{ds^2} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right)^2 \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{x=S}^E 2 \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \frac{d}{ds^2} \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \right\} \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ \left(\frac{d}{ds^2} s^{-1} \right) \exp\left(-\frac{(m-x)^2}{2s^2}\right) + s^{-1} \frac{d}{ds^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ (-s^{-2}) \exp\left(-\frac{(m-x)^2}{2s^2}\right) + s^{-1} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \frac{d}{ds^2} \left(-\frac{(m-x)^2}{2s^2}\right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ (-s^{-2}) \exp\left(-\frac{(m-x)^2}{2s^2}\right) + s^{-1} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(-\frac{(m-x)^2}{2}\right) \cdot (-2s^{-3}) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ (-s^{-2}) \exp\left(-\frac{(m-x)^2}{2s^2}\right) + \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x)^2 \cdot (s^{-4}) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right]
\end{aligned}$$

Wir wollen wieder die erste Ableitung der Log-Likelihoodfunktion plotten. Diesmal in s^2 -Richtung, aber wieder ohne störenden Vorfaktor und mit $m = 80$, der Schätzer für m aus der Abbildung 15. Man erhält die Abbildung 16:

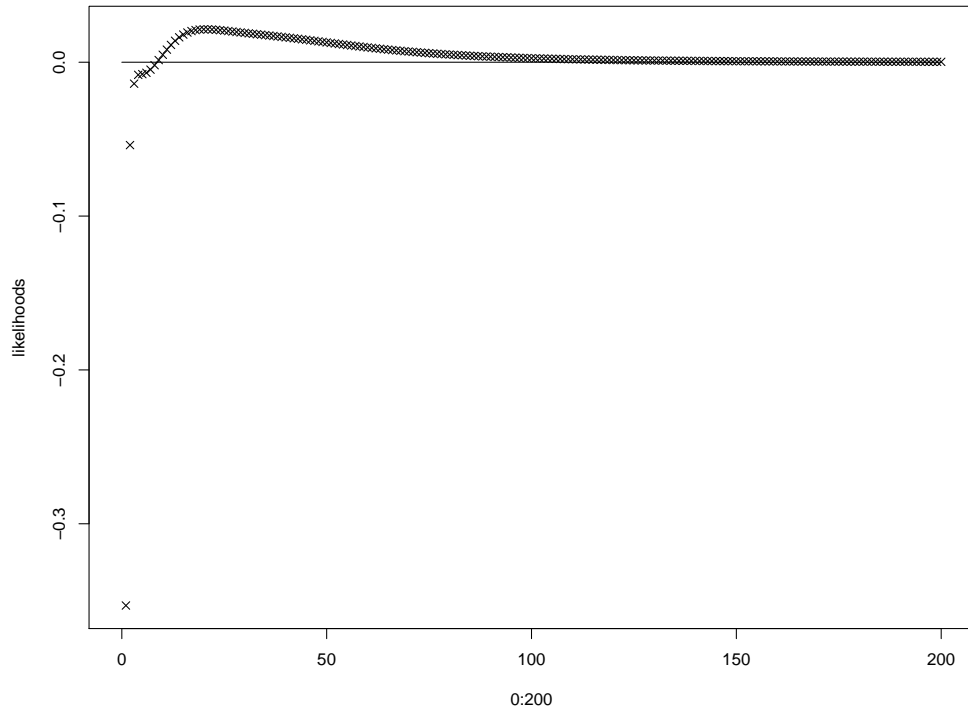


Abbildung 16: Die erste Ableitung der Log-Likelihoodfunktion in s^2 -Richtung. Auf der x -Achse mögliche Werte für \hat{s}^2 und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht, dass die ersten paar s^2 -Werte stark von den späteren abweichen. Man sieht weiterhin, dass die s^2 -Werte für $s^2 \rightarrow \infty$ positiv bleiben. Zumindest ist dies in einem sinnvollen Bereich von $s^2 < 200$ der Fall. Betrachten wir also Abbildung 17, in der nur der Bereich von 4 bis 10 in 0.1-Inkrementen zu sehen ist:

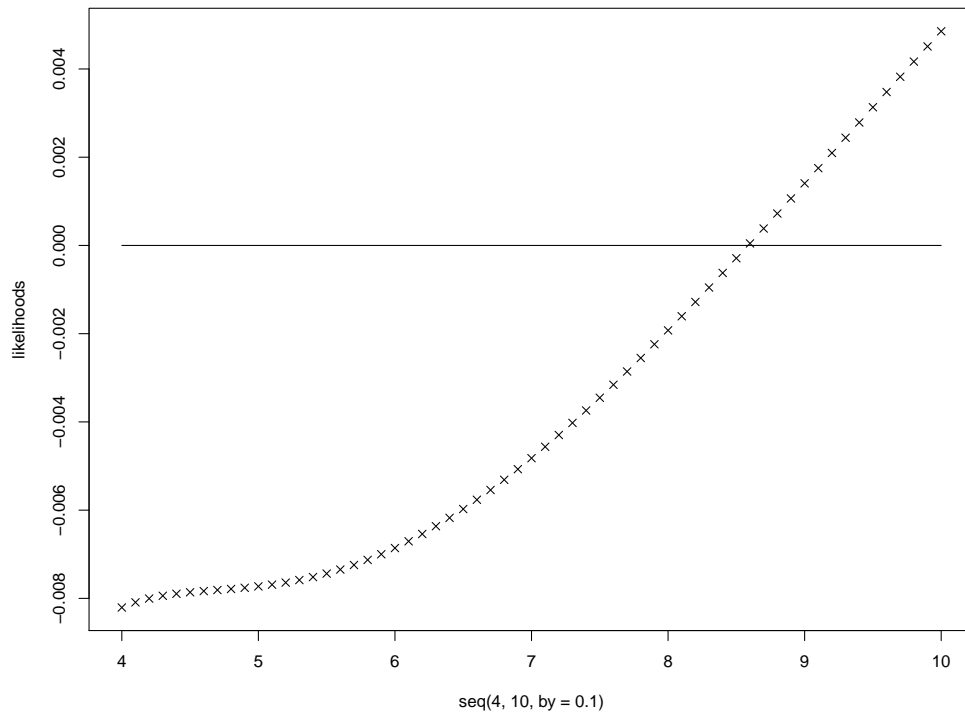


Abbildung 17: Die erste Ableitung der Log-Likelihoodfunktion in s^2 -Richtung im Bereich 4 bis 10 in 0.1-Inkrementen. Auf der x -Achse sind wieder mögliche Werte für \hat{s}^2 und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht in Abbildung 17 genau eine Nullstelle bei ungefähr 8,7. Berechnen wir als Nächstes die zweiten Ableitungen der Log-Likelihoodfunktion. Beginnen wir mit der zweiten Ableitung in m -Richtung:

$$\begin{aligned}
& \frac{d^2}{dm^2} l_{(q_S, \dots, q_E)}(m, s) \\
&= \frac{d}{dm} \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{x=S}^E \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \\
&= \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{x=S}^E \frac{d}{dm} \left[\left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \right] \\
&= \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{x=S}^E \left[\frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \frac{-(m-x)}{s^2} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) + \dots \right. \\
&\quad \dots \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \frac{-(m-x)}{s^2} \cdot (m-x) + \dots \\
&\quad \left. \dots \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \right] \\
&= \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{x=S}^E \left[\exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \dots \right. \\
&\quad \dots \left\langle -\frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \frac{(m-x)^2}{s^2} - \dots \right. \\
&\quad \left. \dots \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \frac{(m-x)^2}{s^2} + \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \right\rangle \left. \right] \\
&=: c \cdot \sum_{x=S}^E \left[a \cdot \left(-\frac{1}{\sqrt{2\pi} s^2} \cdot a \cdot b - (\mathbb{E}(q_x) - q_x) \cdot b + \mathbb{E}(q_x) - q_x \right) \right]
\end{aligned}$$

Hier ist $a := \exp\left(-\frac{(m-x)^2}{2s^2}\right)$ und $b := \frac{(m-x)^2}{s^2}$. Mit diesen Definitionen kann man den Erwartungswert direkt ablesen:

$$\mathbb{E} \left(\frac{d^2}{dm^2} l_{(q_S, \dots, q_E)}(m, s) \right) = -\frac{1}{2\pi s^4 \sigma_\epsilon^2} \sum_{x=S}^E a^2 b \quad (51)$$

Betrachten wir als Nächstes die Ableitung in m -und s -Richtung. Dabei wird die folgende Identität benutzt:

$$\begin{aligned}
\frac{d}{ds^2} \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} &= \frac{d}{ds^2} \left\{ \frac{1}{\sqrt{2\pi} s^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \right\} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \left(\frac{(m-x)^2}{2s^2} - \frac{1}{s^2} \right) \\
&=: \frac{1}{\sqrt{2\pi}} d_s
\end{aligned}$$

Außerdem ist $\mathbb{E}(q_x) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right)$. Damit können wir die gemischte Ableitung gut berechnen:

$$\begin{aligned}
& \frac{d}{ds^2} \frac{d}{dm} l_{(q_S, \dots, q_E)}(m, s) \\
&= \frac{d}{ds^2} \frac{1}{\sqrt{2\pi s^3 \sigma_\epsilon^2}} \sum_{x=S}^E \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \\
&= \frac{1}{\sqrt{2\pi \sigma_\epsilon^2}} \left\{ \frac{d}{ds^2} s^{-3} \right\} \sum_{x=S}^E \left[\left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \right] + \dots \\
&\quad \dots \frac{1}{\sqrt{2\pi s^3 \sigma_\epsilon^2}} \sum_{x=S}^E \frac{d}{ds^2} \left[\left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \right] \\
&= -\frac{3}{\sqrt{2\pi \sigma_\epsilon^2} s^4} \sum_{x=S}^E \left[\left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \right] + \dots \\
&\quad \dots \frac{1}{\sqrt{2\pi s^3 \sigma_\epsilon^2}} \sum_{x=S}^E \left[\frac{d}{ds^2} \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) + \dots \right. \\
&\quad \left. \dots \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \frac{d}{ds^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) \right] \\
&= -\frac{3}{\sqrt{2\pi \sigma_\epsilon^2} s^4} \sum_{x=S}^E \left[\{\mathbb{E}(q_x) - q_x\} \cdot a \cdot (m-x) \right] + \frac{1}{\sqrt{2\pi s^3 \sigma_\epsilon^2}} \cdot \dots \\
&\quad \dots \sum_{x=S}^E \left[d_s \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot (m-x) + \dots \right. \\
&\quad \left. \dots \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right\} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \left(\frac{2(m-x)^2}{2s^3} \right) \cdot (m-x) \right] \\
&= -\frac{3}{\sqrt{2\pi \sigma_\epsilon^2} s^4} \sum_{x=S}^E \left[\{\mathbb{E}(q_x) - q_x\} \cdot a \cdot (m-x) \right] - \dots \\
&\quad \dots \frac{1}{\sqrt{2\pi s^3 \sigma_\epsilon^2}} \sum_{x=S}^E \left[d_s \cdot a \cdot (m-x) + \{\mathbb{E}(q_x) - q_x\} \cdot a \cdot \left(2 \cdot \frac{(m-x)^3}{s^3} \right) \right]
\end{aligned}$$

Man kann den Erwartungswert direkt ablesen:

$$\mathbb{E} \left(\frac{d}{ds^2} \frac{d}{dm} l_{(q_S, \dots, q_E)}(m, s) \right) = -\frac{1}{2\pi s^2 \sigma_\epsilon^2} \sum_{x=S}^E d_s a \sqrt{b} \quad (52)$$

Berechnen wir als Nächstes die zweifache Ableitung in s -Richtung:

$$\begin{aligned}
& \frac{d^2}{ds^4} l_{(q_S, \dots, q_E)}(s) \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \frac{d}{ds^2} \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \dots \\
&\quad \dots \left[\frac{d}{ds^2} \left(\frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \left\{ \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right\} + \dots \right. \\
&\quad \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \frac{d}{ds^2} \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) + \dots \\
&\quad \left. \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \frac{d}{ds^2} \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \dots \\
&\quad \left[\left\{ -\frac{1}{s^2 \sqrt{2\pi}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) + \frac{1}{\sqrt{2\pi s^2}} \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{s^3} \right) \right\} \cdot \dots \right. \\
&\quad \dots \left\{ \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right\} + \dots \\
&\quad \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \frac{(m-x)^2}{s^3} \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) + \dots \\
&\quad \left. \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right) \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{2}{s^3} - \frac{4(m-x)^2}{s^5} \right) \right] \\
&=: -\frac{1}{\sigma_\epsilon^2 2\pi} (A + B + C)
\end{aligned}$$

Hierbei sind A, B und C die folgenden Ausdrücke:

$$\begin{aligned}
A &= a^2 \cdot \frac{1}{\sqrt{2\pi}} \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right)^2 = d_s^2 \\
B &= (\mathbb{E}(q_x) - q_x) \cdot a \cdot \left(\frac{2}{s^3} - \frac{4(m-x)^2}{s^5} \right) \\
C &= (\mathbb{E}(q_x) - q_x) \cdot a \cdot \frac{b}{s} \cdot \left(\frac{b}{s^2} - \frac{1}{s^2} \right)
\end{aligned}$$

Man sieht direkt, dass A als Produkt von positiven Zahlen positiv ist und die Erwartungswerte von B und C Null sind:

$$\mathbb{E}(B) = 0 = \mathbb{E}(C)$$

Insgesamt erhält man den folgenden Erwartungswert:

$$\mathbb{E} \left(\frac{d^2}{ds^4} l_{(q_S, \dots, q_E)}(s) \right) = -\frac{1}{\sigma_\epsilon^2 2\pi} \sum_{x=S}^E d_s^2 \quad (53)$$

Zeigen wir jetzt, dass die Voraussetzungen von Theorem 3 aus Abschnitt 4.1 erfüllt sind:

1. Das Modell ist identifizierbar: Zeigen wir jetzt Identifizierbarkeit. Dazu fassen wir das einfache Modell (48) als eine Menge normalverteilter Maße auf, die aus dem folgenden Modell generiert werden:

$$P_\theta = P_{m,s} : Q = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp \left(-\frac{(m - X)^2}{2 \cdot s^2} \right) + \epsilon$$

Die Maße P_θ sind dann Transformationen von ϵ um $f(\theta, x)$ wobei f die Dichte einer Standardnormalverteilung ist. Setzen wir für dieses Modell die Definition der Identifizierbarkeit an:

$$\begin{aligned} & P_{\theta_1} = P_{\theta_2} \\ \iff & \frac{1}{\sqrt{2 \cdot \pi \cdot s_1^2}} \cdot \exp \left(-\frac{(m_1 - X)^2}{2 \cdot s_1^2} \right) + \epsilon = \frac{1}{\sqrt{2 \cdot \pi \cdot s_2^2}} \cdot \exp \left(-\frac{(m_2 - X)^2}{2 \cdot s_2^2} \right) + \epsilon \\ \iff & -\frac{\ln(2\pi s_1^2)}{2} - \frac{(m_1 - X)^2}{2 \cdot s_1^2} = -\frac{\ln(2\pi s_2^2)}{2} - \frac{(m_2 - X)^2}{2 \cdot s_2^2} \\ \iff & \ln(s_1^2) + \frac{(m_1 - X)^2}{s_1^2} = \ln(s_2^2) + \frac{(m_2 - X)^2}{s_2^2} \\ \iff & x^2 \left(\frac{1}{s_1^2} - \frac{1}{s_2^2} \right) - 2x \left(\frac{m_1}{s_1^2} - \frac{m_2}{s_2^2} \right) + \left(\frac{m_1^2}{s_1^2} - \frac{m_2^2}{s_2^2} + \ln(s_1) - \ln(s_2) \right) = 0 \end{aligned}$$

Diese letzte Gleichung gilt genau dann, wenn die Koeffizienten Null sind. Das heißt, es muss gelten $m_1 = m_2$ und $|s_1| = |s_2|$. Da wir vorausgesetzt haben, dass $s_1, s_2 \in \mathbb{R}^+$ folgt $\theta_1 = \theta_2$ und somit ist das Modell identifizierbar.

2. Es ist $\mathbb{E}(|\ln(f(x|\theta))|) < \infty$: Durch Herausziehen von Minus Eins und Anwenden der Dreiecksungleichung erhält man:

$$\begin{aligned}
\mathbb{E}(|\ln(f(q_x|\theta))|) &= \mathbb{E} \left(\left| -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi}}\frac{1}{s}h\left(\frac{m-x}{s}\right) - q_x\right)^2}{2\sigma_\epsilon^2} \right| \right) \\
&\leq \frac{|\ln(2\pi\sigma_\epsilon^2)|}{2} + \frac{\mathbb{E}(\mathbb{E}(q_x) - q_x)^2}{2\sigma_\epsilon^2} \\
&= \frac{|\ln(2\pi\sigma_\epsilon^2)|}{2} + \frac{\text{Var}(q_x)}{2\sigma_\epsilon^2} \\
&= \frac{|\ln(2\pi\sigma_\epsilon^2)|}{2} + \frac{1}{2} < \infty
\end{aligned}$$

3. Der Parameterraum $\Theta \subset \mathbb{R}^n$ ist ein möglicherweise offenes Rechteck: Bei der Wahl eines passenden Parameterraums muss man beachten, dass auch Voraussetzungen 6 und 10 erfüllbar sind. Sei deshalb $\Theta = (30, 85) \times (4, 10)$. Diese Menge ist genau so gewählt, dass man ein offenes Rechteck in \mathbb{R}^2 erhält.
4. Die Ableitungen der Log-Likelihoodfunktion konvergieren in Wahrscheinlichkeit gegen ihren Erwartungswert: Diese Eigenschaft gilt nach dem Gesetz der großen Zahlen, wenn der Erwartungswert existiert und die Zufallsvariablen unabhängig und identisch verteilt (u.i.v.) sind. Die Eigenschaft u.i.v. zu sein ist Teil der Modelldefinition. Die Erwartungswerte sind beide Null:

$$\begin{aligned}
\mathbb{E} \left\{ \frac{d}{dm} \right\} &= \mathbb{E} \left\{ \frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}} \sum_{x=S}^E \left(\frac{1}{\sqrt{2\pi} s^2} \exp \left(-\frac{(m-x)^2}{2s^2} \right) - q_x \right) \cdot \dots \right. \\
&\quad \left. \dots \exp \left(-\frac{(m-x)^2}{2s^2} \right) \cdot (m-x) \right\} \\
&= \frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}} \sum_{x=S}^E \mathbb{E} \{ \mathbb{E}(q_x) - q_x \} \cdot \exp \left(-\frac{(m-x)^2}{2s^2} \right) \cdot (m-x) \\
&= 0 \\
\mathbb{E} \left\{ \frac{d}{ds^2} \right\} &= \mathbb{E} \left\{ -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left[\left(\frac{1}{\sqrt{2\pi} s^2} \exp \left(-\frac{(m-x)^2}{2s^2} \right) - q_x \right) \cdot \dots \right. \right. \\
&\quad \left. \left. \dots \exp \left(-\frac{(m-x)^2}{2s^2} \right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right] \right\} \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{x=S}^E \left\{ \mathbb{E}(\mathbb{E}(q_x) - q_x) \cdot \exp \left(-\frac{(m-x)^2}{2s^2} \right) \cdot \left(\frac{(m-x)^2}{s^4} - \frac{1}{s^2} \right) \right\} \\
&= 0
\end{aligned}$$

5. Die Ableitungen der Log-Likelihoodfunktion sind stetig in θ : Die Log-Likelihoodfunktion ist gegeben durch:

$$l_{(q_S, \dots, q_E)}(m, s) = \sum_{x=S}^E \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right)^2}{2\sigma_\epsilon^2} \right\}$$

Da $l_{(q_S, \dots, q_E)}$ eine Verknüpfung stetig differenzierbarer Funktionen ist, ist die Ableitung der Log-Likelihoodfunktion ebenfalls stetig.

6. Der Punkt θ_0 erfüllt $\frac{d}{d\theta} l(\theta_0 - \epsilon) < 0 < \frac{d}{d\theta} l(\theta_0 + \epsilon)$ für alle $\epsilon > 0$ und $\theta = \theta_1, \theta_2, \dots$: Der Parameterraum wurde genau so gewählt, dass dies zutrifft. Als Begründung betrachte die Abbildungen der Log-Likelihoodfunktion.
7. Der Parameterraum Θ ist offen: Der Parameterraum $\Theta = (30, 85) \times (4, 10)$ ist genau passend gewählt.
8. Das Modell hat ein Maß P_θ mit Dichte p_θ die stetig differenzierbar in θ und in x ist: Die Dichte ist gegeben durch:

$$f_{(m,s)}((q_S, \dots, q_E)) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp \left(-\frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x \right)^2}{2\sigma_\epsilon^2} \right)$$

Da alle Bestandteile von $f_{(m,s)}$ stetig differenzierbar in x und θ sind, ist auch $f_{(m,s)}$ stetig differenzierbar.

9. Die Fisher-Informationsmatrix $i(\theta)$ existiert, ist nicht singulär und ist stetig in θ : Die Fisher-Information ist definiert als:

$$\begin{aligned} i(\theta) &= -\mathbb{E} \begin{pmatrix} \frac{d^2}{dm^2} l_{(q_S, \dots, q_E)}(\theta) & \frac{d}{dm} \frac{d}{ds} l_{(q_S, \dots, q_E)}(\theta) \\ \frac{d}{ds} \frac{d}{dm} l_{(q_S, \dots, q_E)}(\theta) & \frac{d^2}{ds^2} l_{(q_S, \dots, q_E)}(\theta) \end{pmatrix} \\ &= \begin{pmatrix} -\mathbb{E} \frac{d^2}{dm^2} l_{(q_S, \dots, q_E)}(\theta) & -\mathbb{E} \frac{d}{dm} \frac{d}{ds} l_{(q_S, \dots, q_E)}(\theta) \\ -\mathbb{E} \frac{d}{ds} \frac{d}{dm} l_{(q_S, \dots, q_E)}(\theta) & -\mathbb{E} \frac{d^2}{ds^2} l_{(q_S, \dots, q_E)}(\theta) \end{pmatrix} \end{aligned}$$

Alle Bestandteile wurden in Gleichung (51), (52) und (53) bereits ausgerechnet. Man erhält als Determinante:

$$\det(i(\theta)) = \frac{1}{2\pi s^4 \sigma_\epsilon^2} \sum_{x=S}^E a^2 b \cdot \frac{1}{2\pi \sigma_\epsilon^2} \sum_{x=S}^E d_s^2 - \left(\frac{1}{2\pi s^2 \sigma_\epsilon^2} \sum_{x=S}^E d_s a \sqrt{b} \right)^2$$

Die Cauchy-Schwarz-Ungleichung liefert:

$$\sum_{x=S}^E a^2 b \cdot \sum_{x=S}^E d_s^2 \geq \left(\sum_{x=S}^E d_s a \sqrt{b} \right)^2$$

Gleichheit gilt genau dann, wenn die Zeilen-oder Spaltenvektoren der Matrix linear abhängig sind. Dies ist nicht der Fall, wenn man die Definition von linearer Abhängigkeit auf den i -ten Eintrag von $a\sqrt{b}$ und d_s anwendet. Man erhält, dass die folgende Gleichung für alle $x = \{0, \dots, 95\}$ gelten muss.

$$\begin{aligned} \lambda \cdot a\sqrt{b} &= d_s \\ \iff \lambda \cdot \exp\left(-\frac{(m-x)^2}{2s^2}\right) \sqrt{\frac{(m-x)^2}{s^2}} &= \exp\left(-\frac{(m-x)^2}{2s^2}\right) \cdot \left(\frac{(m-x)^2}{2s^2} - \frac{1}{s^2}\right) \\ \iff \lambda \cdot \sqrt{\frac{(m-x)^2}{s^2}} &= \left(\frac{(m-x)^2}{2s^2} - \frac{1}{s^2}\right) \\ \iff \lambda &= (m-x)^2 - \sqrt{m-x} - 1 \end{aligned}$$

Diese Gleichung kann nicht für alle x gelten, da man aus den ersten beiden Einträgen der Vektoren, das heißt $x \in \{0, 1\}$, den Wert $m = 1$ erhält. Setzt man diesen für $x = 2$ ein, erhält man eine strikt imaginäre Zahl aus der Wurzel und diese ist sicher nicht -1 , der Wert für λ , den man aus den ersten beiden Einträgen erhält.

Damit ist die Cauchy-Schwarz-Ungleichung eine strikte Ungleichung, die Determinante der Fisher-Informationsmatrix ist ungleich Null und somit sind alle Eigenwerte größer Null und die Matrix existiert und ist nicht singulär. Weiterhin ist $l_{(q_S, \dots, q_E)}$ gegeben durch:

$$l_{(q_S, \dots, q_E)}(m, s) = \sum_{x=S}^E \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - q_x\right)^2}{2\sigma_\epsilon^2} \right\}$$

Man sieht, dass $l_{(q_S, \dots, q_E)}$ als Verkettung glatter Funktionen in θ wieder glatt ist und insbesondere sind die zweiten Ableitungen stetig in θ .

10. Es existiert eine Funktion g mit $\mathbb{E}(g^2) < \infty$ und

$$|l_{m_1, s_1}(y) - l_{m_2, s_2}(y)| \leq g(y) \|(m_1, s_1) - (m_2, s_2)\|$$

Diese Eigenschaft kann man als Lipschitz-Stetigkeit der Log-Likelihoodfunktion auffassen. Eine Verknüpfung von Lipschitz-stetigen Funktionen ist wieder Lipschitz-stetig mit dem Produkt der Lipschitzkonstanten als Lipschitzkonstante. Siehe dazu Gleichung 54.

$$|f(g(x)) - f(g(y))| \leq L_f |g(x) - g(y)| \leq L_f L_g |x - y| \quad (54)$$

Eine Dichte ist nach Proposition 2.2 in [HR10, S. 118] Lipschitz-stetig, wenn sie beschränkt ist. Die Normalverteilung ist beschränkt, also hat man Lipschitz-Stetigkeit der Funktion g aus Gleichung (54). Weiterhin folgt aus $x, y \in \{S, \dots, E\}$, dass die folgende Gleichung gilt:

$$g(x), g(y) \in \left[\min_{z \in \{S, \dots, E\}} g(z), \max_{z \in \{S, \dots, E\}} g(z) \right] \quad (55)$$

Die Logarithmusfunktion wird also auf einem abgeschlossenen Intervall angewandt und ist dort Lipschitz-stetig.

Also sind alle Voraussetzungen für das Theorem 3 aus Abschnitt 4.1 erfüllt und das Modell ist identifizierbar, konsistent und asymptotisch normal.

4.3 Eigenschaften des Lee-Carter-Modells

Das Ziel dieses Abschnitts ist, die Voraussetzungen des Theorem von Cramér zu zeigen. Hierbei untersuchen wir die Schätzer für β_x und ν getrennt. Man beachte wieder, dass die Aussage der asymptotischen Normalität für $n \rightarrow \infty$ so zu verstehen ist, dass $m_{x,t}$ das logarithmierte Verhältnis von Summen von Zufallsvariablen ist. Die Zufallsvariablen sind Personen und um so mehr Personen man beobachtet, über um so mehr Zufallsvariablen summiert man.

4.3.1 Der Schätzer für β_x

Wir geben zuerst das Modell, in dem β_x geschätzt wird, an, berechnen den Maximum-Likelihood-Schätzer und zeigen dann die fünf Eigenschaften. In Abschnitt 3.4 haben wir gesehen, dass im Lee-Carter-Modell die folgenden Gleichungen gelten:

$$\begin{aligned} \alpha &= \frac{1}{t} \sum_{t=1}^T m_{x,t} = \bar{m}_x \\ \gamma &= \sum_{x=S}^E (m_{x,t} - \bar{m}_x) = \sum_{x=S}^E \tilde{m}_{x,t} \end{aligned}$$

Bezeichnen wir diese Werte nun mit $\hat{\alpha}$ und $\hat{\gamma}$, um zu betonen, dass wir einen Schätzer betrachten. Man erhält dann das folgende Modell:

$$\tilde{m}_{x,t} \sim N(\beta_x \hat{\gamma}_t, \sigma_{\epsilon_{x,t}})$$

Entsprechend erhalten wir die Dichte:

$$\prod_{t=1}^T f(\tilde{m}_{x,t}|\beta_x, \sigma_{\epsilon_{x,t}}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2}{2\sigma_{\epsilon_{x,t}}^2}\right)$$

Für die Log-Likelihoodfunktion erhalten wir:

$$\begin{aligned} & \ln\left(\prod_{t=1}^T f(\tilde{m}_{x,t}|\beta_x, \sigma_{\epsilon_{x,t}})\right) \\ &= \ln\left(\prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2}{2\sigma_{\epsilon_{x,t}}^2}\right)\right) \\ &= \sum_{t=1}^T \left(\frac{\ln(2\pi\sigma_{\epsilon_{x,t}}^2)}{2} - \frac{(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \end{aligned}$$

Wir erhalten als Ableitung in β_x -Richtung:

$$\begin{aligned} & \frac{d}{d\beta_x} l_{\tilde{m}_{x,t}}(\beta_x) \\ &= \frac{d}{d\beta_x} \sum_{t=1}^T \left(\frac{\ln(2\pi\sigma_{\epsilon_{x,t}}^2)}{2} - \frac{(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\ &= \frac{1}{2\sigma_{\epsilon_{x,t}}^2} \sum_{t=1}^T \frac{d}{d\beta_x} \left(-(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2\right) \\ &= \frac{1}{2\sigma_{\epsilon_{x,t}}^2} \sum_{t=1}^T (-2(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})\hat{\gamma}_t) \\ &= \frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t - \sum_{t=1}^T \beta_x \hat{\gamma}_t^2\right) \\ &= \frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t - \beta_x \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2\right)\right) \end{aligned}$$

Man erhält als Nullstelle von $\frac{d}{d\beta_x} l_{\tilde{m}_{x,t}}(\beta_x)$:

$$\begin{aligned}
0 &= \frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t - \beta_x \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \right) \\
\iff 0 &= \sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t - \beta_x \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \\
\iff \sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t &= \beta_x \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \\
\iff \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)^{-1} &= \beta_x
\end{aligned}$$

Das heißt, man erhält als Maximum-Likelihood-Schätzer:

$$\hat{\beta}_x = \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)^{-1}$$

Dieser Schätzer erfüllt immer noch die Bedingung $\sum_x \beta_x = 1$. Diese Bedingung wurde bei der Modellbildung angenommen. Die folgende Rechnung benutzt die Identität $\sum_x \tilde{m}_{x,t} = \hat{\gamma}_t$, die ebenfalls in der Modellbildung gezeigt wurde:

$$\begin{aligned}
\sum_{x=S}^E \hat{\beta}_x &= \sum_{x=S}^E \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)^{-1} \\
&= \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)^{-1} \cdot \sum_{x=S}^E \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t \right) \\
&= \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)^{-1} \cdot \left(\sum_{t=1}^T \sum_{x=S}^E \tilde{m}_{x,t} \hat{\gamma}_t \right) \\
&= \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)^{-1} \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \\
&= 1
\end{aligned}$$

Als nächstes wird die zweite Ableitung berechnet:

$$\begin{aligned}
&\frac{d^2}{d\beta_x^2} l_{\tilde{m}_{x,t}}(\beta_x) \\
&= \frac{d}{d\beta_x} \frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \tilde{m}_{x,t} \hat{\gamma}_t - \beta_x \cdot \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \right) \\
&= -\frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)
\end{aligned}$$

Die zweite Ableitung ist immer negativ. Also haben wir in der Tat ein Maximum gefunden. Jetzt werden die fünf Eigenschaften von Theorem 1 aus Abschnitt 4.1 gezeigt:

- Die Menge Θ ist eine offene Teilmenge des \mathbb{R}^n : Dies ist kein Problem, da man Θ passend wählen kann. Setzen wir also $\beta_x \in (c_1, c_2)$ mit c_1, c_2 passend, zum Beispiel $c_1 = -1, c_2 = 2$. Dann liegen die Schätzer für β mittig im Parameterraum für das Beispiel mit großem Abstand zu den Rändern.
- Das Modell ist identifizierbar: Wir zeigen Identifizierbarkeit für alle Parameter gleichzeitig. Das heißt, wir zeigen Identifizierbarkeit für $\theta = (\alpha_x, \beta_x, \gamma_t, \nu)$. Wir setzen wieder die Definition an:

$$P_{\theta_1} = P_{\theta_2} \iff \alpha_x + \beta_x \gamma_t + \epsilon = \alpha'_x + \beta'_x \gamma'_t + \epsilon \text{ und } \gamma_{t-1} + \nu + \xi = \gamma'_{t-1} + \nu' + \xi$$

Betrachten wir zuerst den ersten Teil und setzen die Nebenbedingungen für $\alpha_x = \bar{m}_x = \alpha'_x$ und für $\gamma_t = \sum_x (m_{x,t} - \bar{m}_x)$ ein:

$$\begin{aligned} \alpha_x + \beta_x \gamma_t + \epsilon = \alpha'_x + \beta'_x \gamma'_t + \epsilon &\iff \alpha_x + \beta_x \gamma_t = \alpha'_x + \beta'_x \gamma'_t \\ &\iff \bar{m}_x + \beta_x \sum_x (m_{x,t} - \bar{m}_x) = \bar{m}_x + \beta'_x \sum_x (m_{x,t} - \bar{m}_x) \\ &\iff \beta_x = \beta' \end{aligned}$$

Betrachten wir nun den zweiten Teil:

$$\begin{aligned} \gamma_{t-1} + \nu + \xi = \gamma'_{t-1} + \nu' + \xi &\iff \gamma_{t-1} + \nu = \gamma'_{t-1} + \nu' \\ &\iff \sum_x (m_{x,t-1} - \bar{m}_x) + \nu = \sum_x (m_{x,t-1} - \bar{m}_x) + \nu' \\ &\iff \nu = \nu' \end{aligned}$$

Das heißt insgesamt haben wir $\theta = (\alpha_x, \beta_x, \gamma_t, \nu) = (\alpha'_x, \beta'_x, \gamma'_t, \nu') = \theta'$ gezeigt und das Modell ist somit identifizierbar.

- Die zweite Ableitung von f_θ existiert und ist stetig für alle $x \in \{S, \dots, E\}$ und kann unter das Integral $\int f_\theta(x) dx$ geschoben werden: Es handelt sich wieder um eine Verkettung glatter Funktionen. Deshalb ist die zweite Ableitung von f_θ stetig. Außerdem handelt es sich wieder um eine Exponentialfamilie:

$$\begin{aligned}
f(\tilde{m}_{x,t}|\beta_x, \sigma_{\epsilon_{x,t}}) &= \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t - \tilde{m}_{x,t})^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t)^2 - 2\beta_x \hat{\gamma}_t \tilde{m}_{x,t} + \tilde{m}_{x,t}^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{x,t}}^2}} \exp\left(-\frac{(\beta_x \hat{\gamma}_t)^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \cdot \exp\left(\frac{\beta_x \hat{\gamma}_t \tilde{m}_{x,t}}{\sigma_{\epsilon_{x,t}}^2}\right) \cdot \exp\left(-\frac{\tilde{m}_{x,t}^2}{2\sigma_{\epsilon_{x,t}}^2}\right) \\
&=: h(\tilde{m}_{x,t}) A(\theta) \exp(\eta(\theta) T(\tilde{m}_{x,t}))
\end{aligned}$$

- Es existiert eine messbare, integrierbare Funktion $K(x)$, sodass $|l''_\theta(x)| \leq K(x)$ für alle $\theta : |\theta - \theta_0| < \epsilon$:

$$\begin{aligned}
|l''_\theta(x)| &= \left| -\frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \right| \\
&= \frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \\
&=: K(x)
\end{aligned}$$

Man kann also in diesem Fall einfach $l''_\theta(x) = K(x)$ setzen, da die Summe über t endlich ist und K nicht von x abhängt. Die Stammfunktion ist dann gleich $K(x) \cdot x + C$.

- Die Fisher-Informationsmatrix $i(\theta_0)$ ist positiv definit:

Die zweite Ableitung wurde als der folgende Ausdruck bestimmt:

$$\frac{d^2}{d\beta_x^2} l_{\tilde{m}_{x,t}}(\beta_x) = -\frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)$$

Benutzt man diesen Ausdruck und den Fakt, dass $\hat{\gamma}$ nicht zufällig ist, erhält man die folgende Fisher-Informationsmatrix:

$$\begin{aligned}
i(\theta_0) &= -\mathbb{E} \left(\frac{d^2}{d\beta_x^2} l_{\tilde{m}_{x,t}}(\beta_x) \right) \\
&= -\mathbb{E} \left(-\frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right) \right) \\
&= \frac{1}{\sigma_{\epsilon_{x,t}}^2} \left(\sum_{t=1}^T \hat{\gamma}_t^2 \right)
\end{aligned}$$

Dieser letzte Ausdruck ist immer positiv. Weiterhin sind in \mathbb{R} alle Matrizen symmetrisch und gleich ihrem Eigenwert. Damit ist die Fisher-Informationsmatrix positiv definit.

4.3.2 Schätzer für ν

Zeigen wir als nächstes das Cramér-Theorem 1 aus Abschnitt 4.1 für den Schätzer von ν . Zuerst geben wir wieder das Modell an und berechnen den Maximum-Likelihood-Schätzer:

$$\hat{\gamma}_t \sim N(\hat{\gamma}_{t-1} + \nu, \sigma_{\xi_t}^2)$$

Dann ist die gemeinsame Dichte:

$$\prod_{t=2}^T f(\nu | \hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(-\frac{(\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t)^2}{2\sigma_{\xi_t}^2}\right)$$

Man beachte, dass die Summe bei $t = 2$ beginnt, da für $t = 1$ der Random-Walk beginnt. Die Log-Likelihoodfunktion ist gegeben durch:

$$\begin{aligned} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) &= \ln\left(\prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(-\frac{(\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t)^2}{2\sigma_{\xi_t}^2}\right)\right) \\ &= \sum_{t=2}^T -\frac{\ln(2\pi\sigma_{\xi_t}^2)}{2} - \frac{(\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t)^2}{2\sigma_{\xi_t}^2} \end{aligned}$$

Die erste Ableitung der Log-Likelihoodfunktion ist gegeben durch:

$$\begin{aligned} \frac{d}{d\nu} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) &= \frac{d}{d\nu} \sum_{t=2}^T -\frac{\ln(2\pi\sigma_{\xi_t}^2)}{2} - \frac{(\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t)^2}{2\sigma_{\xi_t}^2} \\ &= -\frac{1}{2\sigma_{\xi_t}^2} \sum_{t=2}^T 2(\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t) \\ &= -\frac{1}{\sigma_{\xi_t}^2} \sum_{t=2}^T (\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t) \end{aligned}$$

Damit ist der Maximum-Likelihood Schätzer gegeben durch:

$$\begin{aligned}
0 &= \frac{d}{d\nu} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) \\
\iff 0 &= -\frac{1}{\sigma_{\xi_t}^2} \sum_{t=2}^T (\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t) \\
\iff 0 &= \sum_{t=2}^T \hat{\gamma}_{t-1} + (T-1)\nu - \sum_{t=2}^T \hat{\gamma}_t \\
\iff \nu &= \frac{\sum_{t=2}^T \hat{\gamma}_t - \sum_{t=2}^T \hat{\gamma}_{t-1}}{T-1}
\end{aligned}$$

Dieser Ausdruck vereinfacht sich zu:

$$\begin{aligned}
\frac{\sum_{t=2}^T \hat{\gamma}_t - \sum_{t=2}^T \hat{\gamma}_{t-1}}{T-1} &= \frac{\hat{\gamma}_2 + \hat{\gamma}_3 + \dots + \hat{\gamma}_T - \hat{\gamma}_1 - \hat{\gamma}_2 - \hat{\gamma}_3 - \dots - \hat{\gamma}_{T-1}}{T-1} \\
&= \frac{\hat{\gamma}_T - \hat{\gamma}_1}{T-1}
\end{aligned}$$

Man erhält als zweite Ableitung:

$$\begin{aligned}
\frac{d^2}{d\nu^2} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) &= -\frac{d}{d\nu} \frac{1}{\sigma_{\xi_t}^2} \sum_{t=2}^T (\hat{\gamma}_{t-1} + \nu - \hat{\gamma}_t) \\
&= \frac{1}{\sigma_{\xi_t}^2} \sum_{t=2}^T (\hat{\gamma}_t - \hat{\gamma}_{t-1}) \\
&= \frac{\hat{\gamma}_T - \hat{\gamma}_1}{\sigma_{\xi_t}^2}
\end{aligned}$$

Es bleibt, Eigenschaften drei, vier und fünf zu zeigen, da man Eigenschaft eins durch passende Definition von Θ sicherstellen kann und Eigenschaft zwei über die Identifizierbarkeit bereits bei der Identifizierbarkeit von β mitbewiesen wurde. Siehe dazu Abschnitt 4.3.1.

- Die zweite Ableitung von f_θ existiert und ist stetig für alle $t \in \{1, \dots, T\}$ und kann unter das Integral $\int f_\theta(x) dx$ geschoben werden: Die zweite Ableitung von f_θ existiert als glatte Verknüpfung glatter Funktionen. Das Integral kann unter die Ableitung geschoben werden, da wir uns in einer Exponentialfamilie befinden:

$$\begin{aligned}
f(\hat{\gamma}_t, \nu, \sigma_{\xi_t}^2) &= \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(-\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1} - \nu)^2}{2\sigma_{\xi_t}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1})^2 - 2(\hat{\gamma}_t - \hat{\gamma}_{t-1})\nu + \nu^2}{2\sigma_{\xi_t}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{\xi_t}^2}} \exp\left(\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1})^2}{2\sigma_{\xi_t}^2}\right) \cdot \exp\left(\frac{\nu^2}{2\sigma_{\xi_t}^2}\right) \cdot \exp\left(\frac{(\hat{\gamma}_t - \hat{\gamma}_{t-1})\nu}{\sigma_{\xi_t}^2}\right) \\
&= h(\hat{\gamma}_t)A(\theta)\exp(\eta(\theta)T(\hat{\gamma}_t))
\end{aligned}$$

- Es existiert eine messbare, integrierbare Funktion $K(x)$, sodass $|l''_{\theta}(x)| \leq K(x)$ für alle $\theta : |\theta - \theta_0| < \epsilon$: Die zweite Ableitung hängt weder von $\theta = \nu$ noch von $x = \hat{\gamma}_t$ ab. Setzt man $K(x) = l''_{\theta}(x)$ reicht es zu zeigen, dass $K(x) < \infty$:

$$\begin{aligned}
\frac{d^2}{d\nu^2} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) &= \frac{\hat{\gamma}_T - \hat{\gamma}_1}{2\sigma_{\xi_t}^2} \\
&= \frac{1}{\sigma_{\xi_t}^2} \left(\sum_x (m_{x,T} - \bar{m}_x) - \sum_x (m_{x,1} - \bar{m}_x) \right) \\
&= \frac{1}{\sigma_{\xi_t}^2} \left(\sum_x m_{x,T} - \sum_x m_{x,1} \right) \\
&= \frac{1}{\sigma_{\xi_t}^2} \left(\sum_x \beta_x \gamma_T - \sum_x \beta_x \gamma_1 \right) \\
&= \frac{1}{\sigma_{\xi_t}^2} (\gamma_T - \gamma_1) < \infty
\end{aligned}$$

In der letzten Gleichung wurde $\sum_x \beta_x = 1$ und $\gamma_T, \gamma_1 < \infty$ nach Annahme benutzt.

- Die Fisher-Informationsmatrix $i(\theta_0)$ ist positiv definit: Die Fisher-Informationsmatrix ist gegeben durch:

$$\begin{aligned}
i(\theta) &= -\frac{d^2}{d\nu^2} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) \\
&= \frac{(\gamma_1 - \gamma_T)}{\sigma_{\xi_t}^2}
\end{aligned} \tag{56}$$

Da wir davon ausgehen, dass die Sterblichkeit sinkt, folgt $\gamma_1 > \gamma_T$. Damit sind sowohl Zähler, als auch Nenner positiv. Daraus folgt, dass der Bruch positiv ist und somit auch die Fisher-Informationsmatrix.

Die asymptotischen Eigenschaften der Modelle ermöglichen es, Konfidenzbänder zu konstruieren. Betrachten wir hier beispielsweise den Schätzer $\hat{\nu}$ für ν . In diesem Abschnitt

wurde gezeigt, dass das Theorem von Cramér anwendbar ist. Das heißt, es wurde gezeigt, dass die Eigenschaft (57) gilt.

$$\sqrt{n}(\hat{\nu} - \nu) \xrightarrow{D} \mathcal{N}(0_k, i(\theta_0)^{-1}) \quad (57)$$

Kombiniert man Eigenschaft (57) und den Wert der Fisher-Matrix aus Gleichung (56), erhält man das Konfidenzintervall (58). Hier bezeichnet $z_{1-\alpha/2} \approx 1.96$ das $1 - \alpha/2$ Quantil der Standardnormalverteilung.

$$\begin{aligned} 1 - \alpha &= \mathbb{P} \left(z_{\alpha/2} < \sqrt{n} \sqrt{i(\theta_0)} (\hat{\nu} - \nu) < z_{1-\alpha/2} \right) \\ &= \mathbb{P} \left(-\frac{z_{1-\alpha/2}}{\sqrt{n} \sqrt{i(\theta_0)}} < (\hat{\nu} - \nu) < \frac{z_{1-\alpha/2}}{\sqrt{n} \sqrt{i(\theta_0)}} \right) \\ &= \mathbb{P} \left(\nu \in \left[\hat{\nu} - z_{1-\alpha/2} \cdot \frac{(\hat{\gamma}_E - \hat{\gamma}_S)}{\sqrt{n}}, \hat{\nu} + z_{1-\alpha/2} \cdot \frac{(\hat{\gamma}_E - \hat{\gamma}_S)}{\sqrt{n}} \right] \right) \end{aligned} \quad (58)$$

4.4 Eigenschaften des einfachen Trendmodells

Wir haben in Abschnitt 3.5 eine Verbesserung des einfachen Modells eingeführt und gesehen, dass diese Verbesserung leider auf Grund der Zensur zu Ergebnissen führt, die nicht mit den Daten vereinbar sind. Das einfache Trendmodell, welches die Zensur ignoriert, ist allerdings ebenfalls konsistent und asymptotisch normal, da wieder die beiden Teilstücke konsistent und asymptotisch normal sind. Zeigen wir beispielsweise die Identifizierbarkeit des einfachen Trendmodells. Wir haben bereits in dem Kapitel über das einfache Modell gezeigt, dass die folgende Identität gilt:

$$P_\theta = P_{\theta'} \iff (m, s) = (m', s')$$

Daran ändert sich auch durch das Hinzufügen des Random Walk with Drift nichts. Es bleibt zu zeigen, dass $\nu = \nu'$:

$$\begin{aligned} P_\theta = P_{\theta'} &\iff \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp \left(-\frac{(m_t - X)^2}{2 \cdot s^2} \right) + \epsilon = \frac{1}{\sqrt{2 \cdot \pi \cdot s'^2}} \cdot \exp \left(-\frac{(m'_t - X)^2}{2 \cdot s'^2} \right) + \epsilon \\ &\quad \text{und } m_{t-1} + \nu + \xi_t = m'_{t-1} + \nu' + \xi_t \\ &\iff m_{t-1} + \nu = m'_{t-1} + \nu' \\ &\iff \nu = \nu' \end{aligned}$$

Also folgt insgesamt:

$$P_\theta = P_{\theta'} \iff (m, s, \nu) = (m', s', \nu') \iff \theta = \theta'$$

Das heißt, man erhält Identifizierbarkeit. Konsistenz und asymptotische Verteilung folgen ebenfalls einfach, indem man die gleichen Rechnungen wie in den entsprechenden Abschnitten durchführt.

4.5 Qualitätsvergleich durch Simulation

In diesem Unterkapitel werden Simulationen vorgestellt, um die verwendeten Methoden zu vergleichen. Es werden drei Arten von Simulationen durchgeführt: In der ersten Art wird der historische Datensatz 1956 Westdeutschland halbiert und die frühere Hälfte, der sogenannte Schätzzeitraum, benutzt, um die spätere Hälfte, den sogenannten Vorhersagezeitraum, vorherzusagen. Dabei wird für das einfache Modell und das Whittaker-Henderson-Verfahren der Datensatz als Generationstafel benutzt, während für das Lee-Carter-Modell dieselben Daten im Format einer Periodentafel benutzt werden. In der zweiten Art Simulation werden die Parameter des einfachen Modells und des Lee-Carter-Modells in ihrem Modell geschätzt. In der dritten Art Simulation wird mit demselben Lee-Carter-Modell wie in Simulation zwei das Verfahren aus Simulation eins getestet. Das heißt, es wird der simulierte Datensatz halbiert und mit der ersten Hälfte die zweite vorhergesagt. Außerdem wird noch eine ergänzende Simulation durchgeführt, bei der untersucht wird, ob Störungen im Parameter γ auch im Schätzer $\hat{\gamma}$ zu sehen sind.

Die Größe des Fehlers wird jeweils in der L_2 Norm bestimmt. Das heißt, es wird $\sum_i (\hat{q}_i - q_i)^2$ für den Datensatz und $\sum_i (\hat{\theta}_i - \theta_i)^2$ für die Parameter betrachtet.

Es werden das einfache Modell und das Lee-Carter-Modell getrennt betrachtet. Da das Whittaker-Henderson-Verfahren kein passendes Modell hat, wird es nur auf den historischen Datensatz angewendet.

Hier noch einige Informationen darüber, wo die Programme für die Berechnungen in diesem Abschnitt zu finden sind. Die Daten werden in der Funktion *MakeData* in der Datei *GenerateData* erzeugt und als globale Variable abgespeichert. Das Schätzen der Abstände erfolgt in der Datei *EstimateGoodness* in der jeweiligen Funktion. Es gibt die Funktionen *GoodnessEinfachData*, *GoodnessWhittakerData*, *GoodnessLeeData* für den Datensatz und *GoodnessEinfachEinfachSim*, *GoodnessLeeSim* für die Parameter. Außerdem gibt es *GoodnessLeePredict* für die Vorhersagesimulation im Lee-Carter-Modell. Die Parameter werden in den Funktionen *einfachstesModellAlterEstimateParameters*, *Whittaker* beziehungsweise *Lee* bestimmt. Die erste Funktion liegt in der Datei *einfachstesModell* zusammen mit einer Hilfsfunktion *einfachstesModellAlterProjektion*, die eine Generationstafel auf das Alter projiziert. Es wird $\left(\sum \left\{ Y - \frac{1}{\sqrt{2\pi s^2}} \exp \left(-\frac{(x-m)^2}{2s^2} \right) \right\}^2 \right)^{1/2}$ numerisch minimiert. Die Funktionen für das Whittaker-Henderson-Verfahren und das Lee-Carter-Modell befinden sich in der Datei *DAV*. Die Simulation für den gestörten γ -Parameter befindet sich in der Funktion *GoodnessLeeSimSprung* in der Datei *EstimateGoodness* statt.

4.5.1 Simulationen des einfachen Modells

Beginnen wir damit, dass einfache Modell auf den geteilten Datensatz anzuwenden. Die geschätzte Abweichung zwischen q_i und \hat{q}_i im Vorhersagezeitraum in l_2 Norm. Präziser gesagt wird der Fehler wie in (59) berechnet und beträgt 0.023406. Ein Fit ist in Abbildung 18 zu sehen.

$$\sum_{i=S}^E \left(q_i - \frac{1}{\sqrt{2\pi\hat{s}^2}} \exp \left(-\frac{(x - \hat{m})^2}{2\hat{s}^2} \right) \right)^2 \quad (59)$$

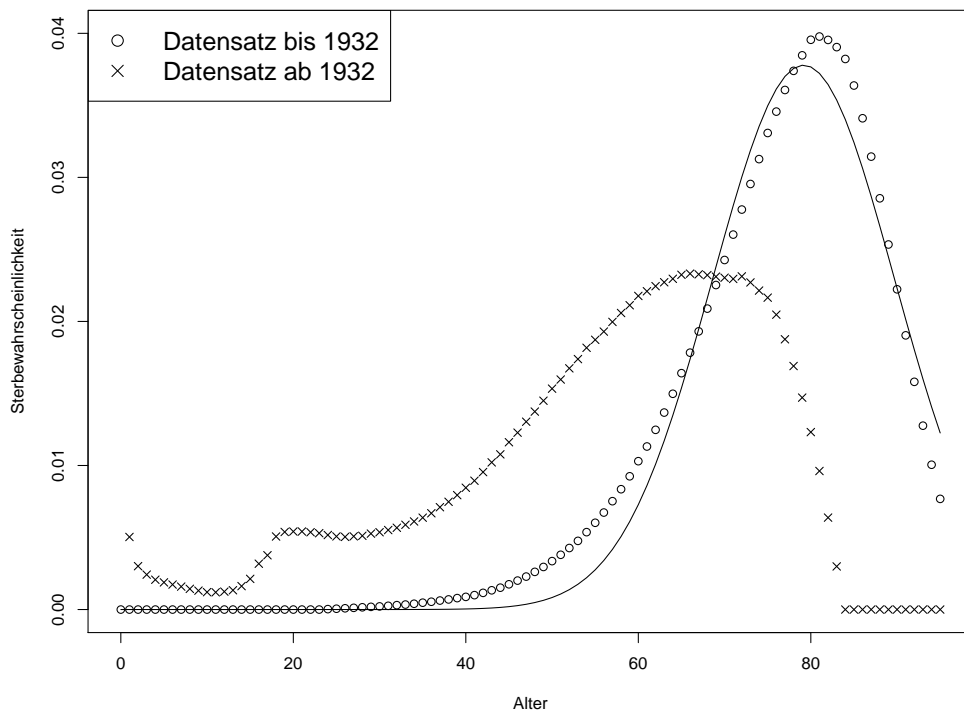


Abbildung 18: Vorhersage des einfachen Modells für den historischen Datensatz. Die Kreise zeigen die erste Hälfte des historischen Datensatzes nach der Transformation, während die Kreuze die zweite Hälfte nach der Transformation zeigen. Die durchgezogene Linie ist der Schätzer im einfachen Modell.

In Abbildung 18 sieht man die Zensur, da bei dem Datensatz, aus dem geschätzt wird, die ersten 24 Werte Null sind, während bei der zweiten Hälfte die letzten 12 Werte Null sind. Außerdem gibt es in der zweiten Hälfte einen Anstieg zwischen den Altern 16 und 20, der bei der ersten Hälfte des Datensatzes nicht zu sehen ist. Ich vermute, dass dies unter anderem mit einer geringeren Anzahl an Beobachtungen zusammenhängt. Die erste Hälfte besteht aus 33.315.629 Todesfällen, während die zweite Hälfte nur aus 7.657.841 Todesfällen

besteht. Da durch die Gesamtzahl an Todesfällen geteilt wird, wirkt die Sterblichkeit für den zweiten Datensatz in frühen Altern höher. Dagegen wird bei dem ersten Datensatz durch eine größere Anzahl an Todesfällen geteilt und somit erscheint die Sterblichkeit in jungen Altern niedriger. Der relativ große Anteil von Personen, die in hohen Altern sterben, wurde im zweiten Datensatz noch nicht beobachtet. Der Anstieg zwischen 16 und 20 wird als Unfallbuckel bezeichnet und mit größerem Todesrisiko durch Arbeit beziehungsweise Transport/Auto fahren erklärt.

Diskutieren wir jetzt die Parameter des einfachen Modells. Es ist $X = (0, \dots, 95)$. Es werden die Parameter $\mu = 77$, $\sigma = 12$ und $\sigma_\epsilon = 0.002427235$ verwendet. Diese sind die gerundeten Werte, die aus dem Datensatz Generationstafel 1879 Westdeutschland, die der Periodentafel 1956 Westdeutschland entspricht, geschätzt wurden. Diese Schätzung erfolgt in der Funktion *MakePlotEinfacheModelle* in der Datei *einfachstesModell*. Die Fehlervarianz wurde mittels $\hat{\sigma}_\epsilon^2 = \frac{1}{96} \sum_{x=0}^{95} (\hat{q}_x - q_x)^2$ geschätzt. Man erhält also das folgende Modell:

$$Y = f(x) + \epsilon = \frac{1}{\sqrt{2\pi}12^2} \exp\left(-\frac{(77-x)^2}{2 \cdot 12^2}\right) + \epsilon, \epsilon_i \sim N(0, 0.000006)$$

Insgesamt werden $m = 10000$ Datensätze erzeugt. Diese Werten in der Matrix *simple_data* nach dem Format Alter, Daten ohne Fehler, Datensatz 1, ..., Datensatz 10000 abgelegt. Dieses Format ist in Tabelle (3) veranschaulicht.

Alter	Daten ohne Fehler	Datensatz 1	...	Datensatz 100
0	$f(0)$	$f(0) + \epsilon_{1,1}$...	$f(0) + \epsilon_{1,10000}$
\vdots	\vdots	\vdots	\vdots	\vdots
95	$f(95)$	$f(95) + \epsilon_{95,1}$...	$f(95) + \epsilon_{95,10000}$

Tabelle 3: Der Datensatz *simple_data*.

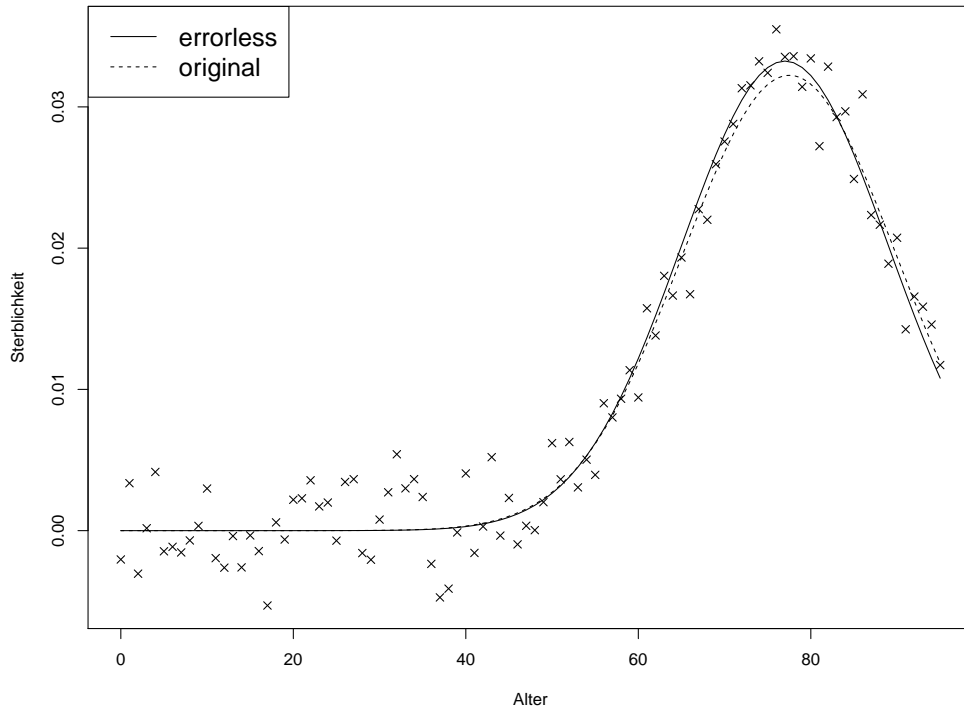


Abbildung 19: Ein Beispiel der Simulationsdaten für das einfache Modell. Man beachte, dass in diesem Modell die Fehler normalverteilt sind.

In Abbildung 19 sieht man den ersten Datensatz in Form von Kreuzen. Die durchgezogene Linie zeigt $f(x)$. Das heißt die Kreuze kennzeichnen $f(x) + \epsilon_{x,1}$. Man sieht, dass die Annahme, dass die Fehler normalverteilt sind, den Nachteil hat, dass es in bestimmten Altern eine negative Wahrscheinlichkeit gibt, dass Personen sterben. Dies ist nicht sinnvoll und eine Schwachstelle des Modells.

Als nächstes sieht man in Abbildung 20 den Abstand zwischen (m, s) und (\hat{m}, \hat{s}) pro Simulation $i \in (1, \dots, 10000)$ als Histogramm. Präziser gesagt, sieht man die empirische Verteilung der folgenden beiden Werte:

$$\sum_{i=1}^{10000} (m - \hat{m}_i)^2 \quad \text{und} \quad \sum_{i=1}^{10000} (s - \hat{s}_i)^2$$

Man sieht gut, dass die Fehler normalverteilt sind. Dies ist genau, was das Modell vorhergesagt hat. Insgesamt ergibt sich eine mittlere Abweichung der Schätzer zu ihren Parametern in L_2 -Norm von 0.1364178.

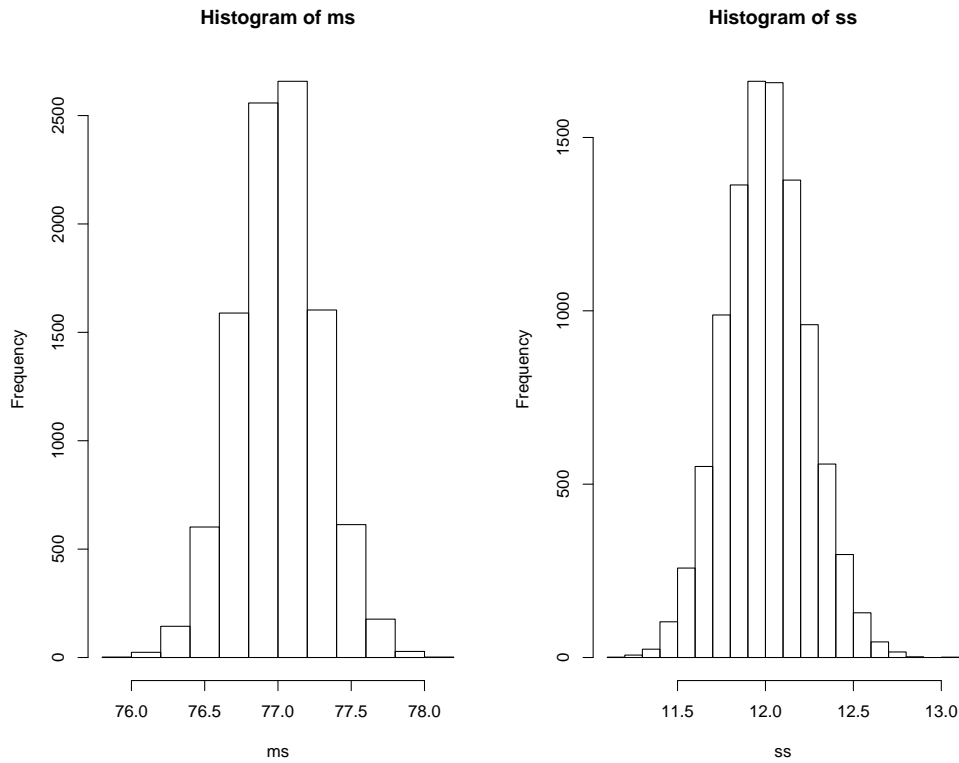


Abbildung 20: Fehler der Simulation im Detail: Links sieht man ein Histogramm der L_2 -Abstände von m und \hat{m} , während im rechten Histogramm die Verteilung der L_2 -Abstände von s und \hat{s} zu sehen ist.

4.5.2 Simulationen zum Lee-Carter-Modell

Diskutieren wir wieder zuerst das Ergebnis der Anwendung des Lee-Carter-Modells auf den historischen Datensatz. Insgesamt erhält man eine Abweichung von 0.0035201, also eine um etwa einen Faktor von 10 kleinere Abweichung als bei dem einfachen Modell. Der Fit ist in Abbildung 21 dargestellt.

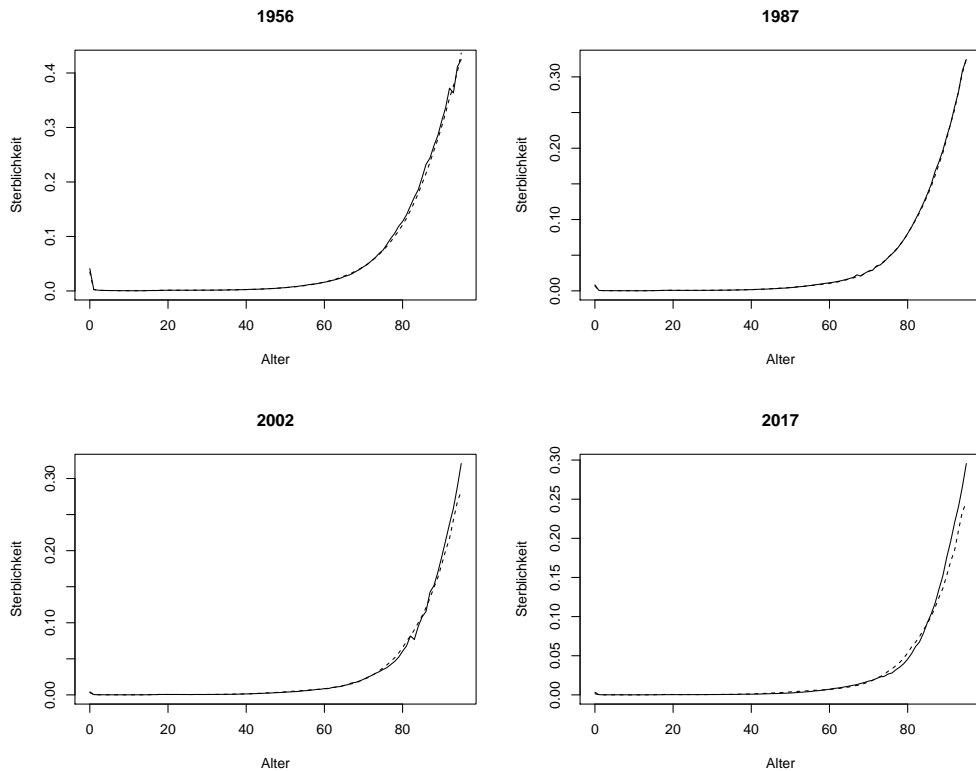


Abbildung 21: Das Lee-Carter-Modell (gestrichelte Linie) angewendet auf den historischen Datensatz (durchgezogene Linie) in vier ausgewählten Beobachtungsjahren: Beginn des Schätzzeitraums (1956), Ende des Schätzzeitraums (1987), Mitte des Vorhersagezeitraums (2002), Ende des Vorhersagezeitraums (2017).

Man sieht, dass der Fit in dem Schätzzeitraum sehr gut ist und in dem Vorhersagezeitraum stetig schlechter wird. Der Fit ist in dem ersten Jahr, in dem vorhergesagt wird, deutlich näher an dem wahren Datensatz, als zum Ende des Schätzzeitraums. In Abbildung 22 sieht man die selben Werte logarithmiert. Man sieht, dass die Abweichungen im letzten Jahr des Vorhersagezeitraums am größten sind. Der wesentliche Unterschied zu Abbildung 21 ist, dass man bei niedrigeren Altern höhere Abweichungen als bei späteren Altern sieht.

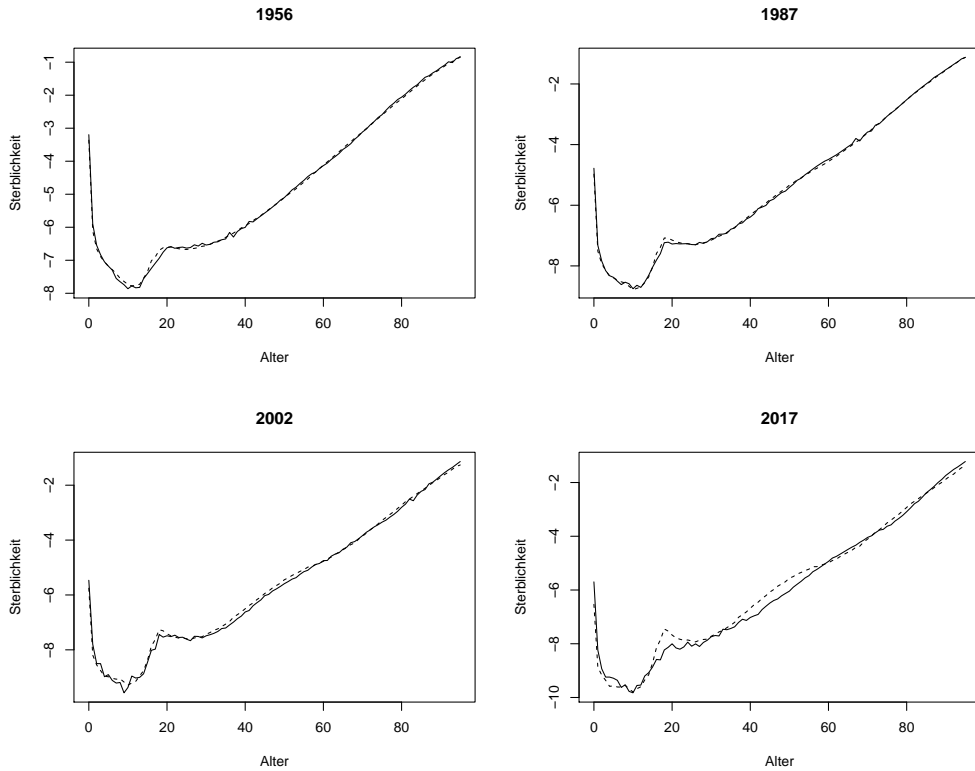


Abbildung 22: Die logarithmierten Werte wie in Abbildung 21.

Als Nächstes werden wieder die Parameter des Lee-Carter-Modells diskutiert, die für die Simulation verwendet werden. Da dieses Modell die Parameter $(\alpha_x, \beta_x, \gamma_t, \nu)$, $x \in (0, 95)$, $t \in (0, 40)$ und damit $2 \cdot x + t + 1$ Parameter hat, sind die Parameter in Abbildung 23 zusammengefasst.

Man sieht oben links die α -Parameter, oben rechts die β -Parameter und unten links die γ -Parameter mit ihren Schätzern. Hier sind α, β und $\nu = -1.679$ aus der ersten Hälfte der Periodentafel 1965 Westdeutschland geschätzt. Die γ -Parameter wurden dann mittels des Modells $\gamma_t = \gamma_{t-1} + \nu + \xi$, $\xi \sim N(0, \sigma_\xi^2)$ generiert. Der Parameter σ_ξ wurde ebenfalls aus der Periodentafel 1965 Westdeutschland geschätzt und beträgt $\sigma_\xi = 2.270456$. Diese Parameter werden in der Funktion *GoodneesLeeData* in der Datei *EstimateGoodness* mittels $\hat{\sigma}_\xi^2 = (\sum (\hat{\gamma}_{t+1} - \hat{\gamma}_{t-1} - \nu)^2) / (\text{Anzahl Beobachtungsjahren})$ bestimmt. Auch für dieses Modell wurde eine Tabelle analog zu Tabelle 3 erzeugt. Allerdings musste vor der Spalte mit der Überschrift Alter noch eine Spalte mit den Beobachtungsjahren eingefügt werden. Beobachtungsjahre sind von 0 bis 40 durchnummeriert. Der Fehler für die Datensätze wurde ebenfalls aus der Periodentafel 1965 Westdeutschland geschätzt und beträgt 0.05933034. Dieser Wert wird in der Funktion *GoodneesLeeData* in der Datei *EstimateGoodness* als die empirische Standardabweichung der Abstände des Datensatzes 1956 Westdeutschland zwischen der Schätzung und dem letzten Beobachtungsjahr für den Schätzzeitraum bestimmt.

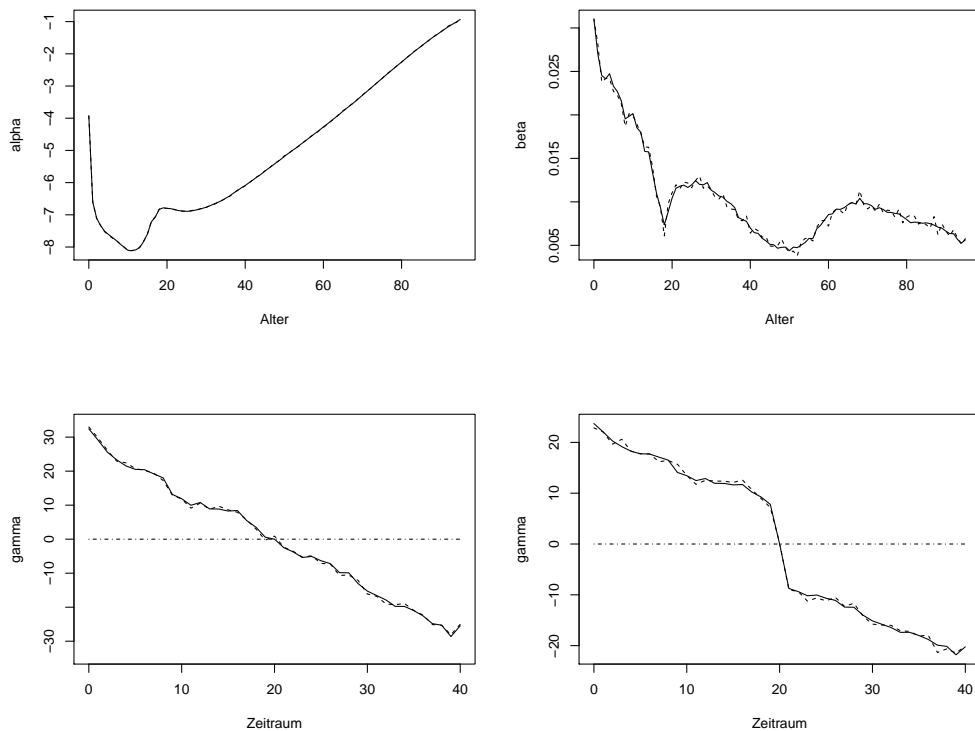


Abbildung 23: Parameter des Lee-Carter-Modells: Oben links alpha (durchgezogen) mit Schätzer (gestrichelt), oben rechts beta (durchgezogen) mit Schätzer (gestrichelt) unten links gamma (durchgezogen) mit Schätzer (gestrichelt) und Nulllinie (dashdot), unten rechts gamma-sprung (durchgezogen) mit Schätzer (gestrichelt) und Nulllinie (dashdot)

Insgesamt sieht man in Abbildung 23, dass der Schätzer für alpha sehr gut ist, während die beta-Werte überschätzt und die gamma-Werte unterschätzt werden. Da die beta- und gamma-Werte im Modell multipliziert werden, um die Sterblichkeit zu schätzen, erhält man im Mittel vermutlich einen guten Wert. Außerdem sieht man rechts unten was passiert, wenn beta nicht wie in dem Modell generiert wird, sondern eine Störung auftritt. Der Schätzer folgt dann dieser Störung und es wird nicht der gesamte Schätzer beeinträchtigt.

In Abbildung 24 sieht man einen Beispieldatensatz, der aus den oben genannten Parametern generiert wurde. Die durchgezogene Linie ist das wahre Modell, die Kreuze sind das wahre Modell Plus einen log-normalverteilten Fehler und die gestrichelte Linie zeigt die Periodentafel 1965 Westdeutschland im entsprechenden Beobachtungsjahr seit dem Start der Periodentafel im Jahr 1965.

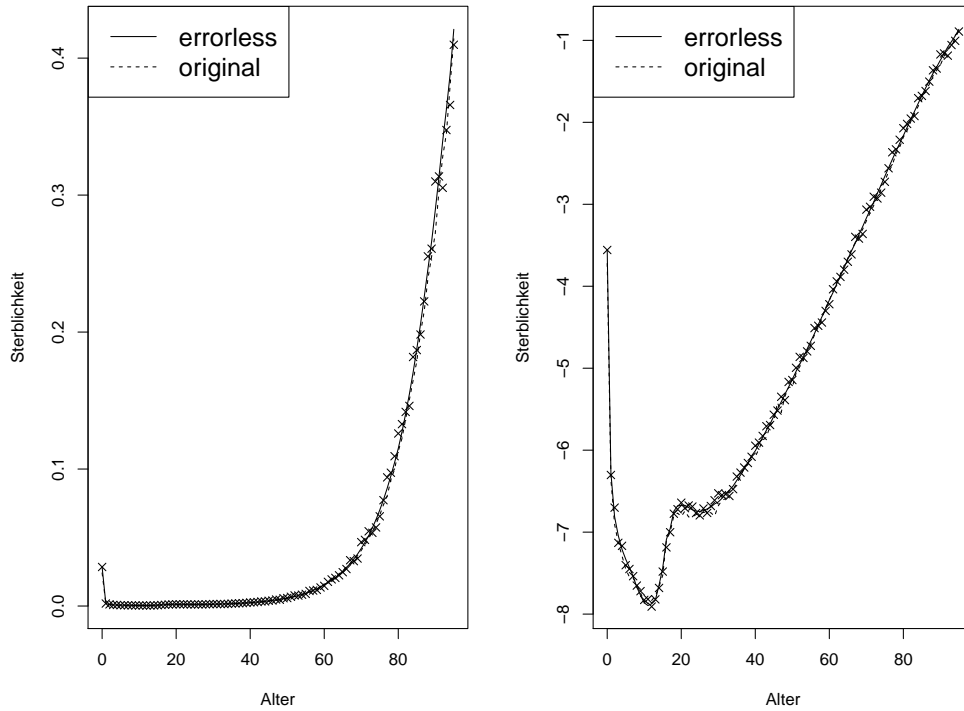


Abbildung 24: Das Lee-Carter-Modell (durchgezogen) mit den Beispielparametern und der historische Datensatz Periodentafel 1965 Westdeutschland (gestrichelt). Außerdem ist ein Beispieldatensatz (gestrichelt) eingetragen. Man beachte, dass in diesem Modell die Fehler log-normalverteilt sind.

In Abbildung 25 sieht man Histogramme der Abstände der Parameter und ihrer Schätzer über die 10000 Simulationen. Man sieht, dass die Fehler insgesamt recht gut einer Normalverteilung folgen und somit die Annahmen des Modells bestätigen. Das heißt, dass das Modell in sich konsistent ist. Weiterhin sieht man, dass die Schätzer für β und ν nicht ganz normalverteilt scheinen. Es wirkt so, als ob mehr Masse in der linken Hälfte ist. Gerade bei ν ist dies nicht sehr verwunderlich, da ν aufgrund der Schätzer von γ bestimmt wurde und somit Folgefehler auftreten. Insgesamt erhält man ein Abweichung der tatsächlichen und der geschätzten Parameter von 0.2746259.

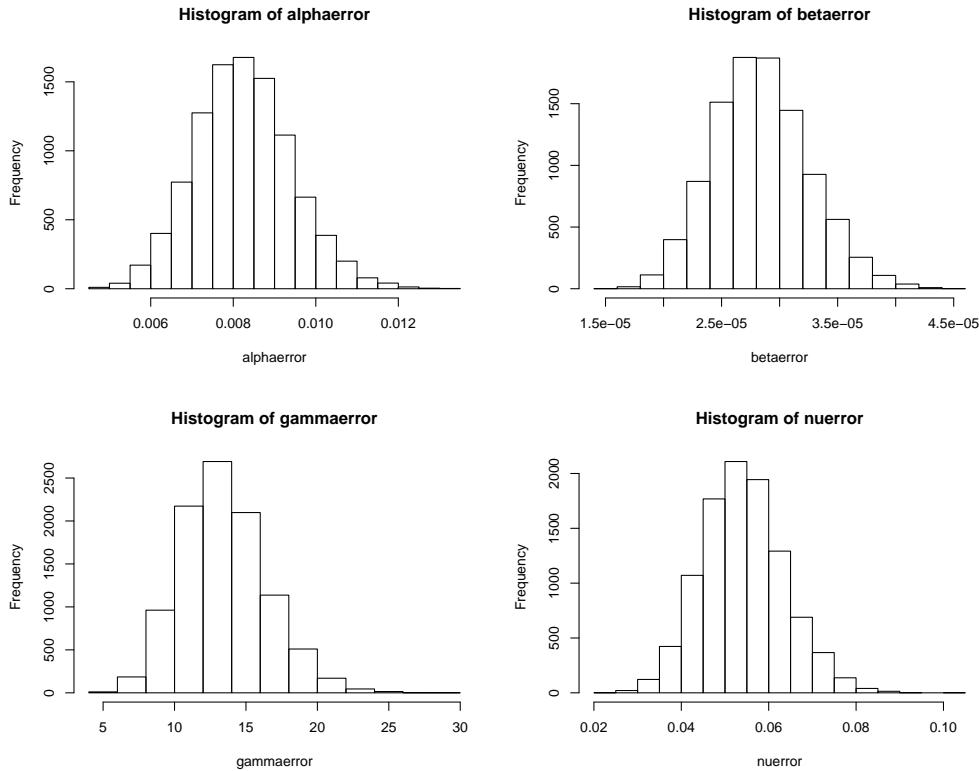


Abbildung 25: Histogramme der Fehler in der Lee-Carter Simulation: oben links Verteilung der Abstände von α und $\hat{\alpha}$, oben rechts β und $\hat{\beta}$ unten links γ und $\hat{\gamma}$, unten rechts ν und $\hat{\nu}$

Abschließend werden die Ergebnisse der Vorhersagesimulation im Lee-Carter-Modell besprochen. Bei dieser Simulation wird der bisher besprochene, simulierte Datensatz zum Lee-Carter-Modell in zwei Hälften aufgeteilt und das Modell nur auf der ersten Hälfte bestimmt und dann mittels des geschätzten ν wie im Modell auf der zweiten Hälfte geschätzt. Die Ergebnisse sind genau wie für den historischen Datensatz in den Abbildungen 26 und 27 aufbereitet. Insgesamt erhält man bei der Vorhersage eine Abweichung von 0.0009655, während man nur eine Abweichung von 0.0000016 erhält, wenn man das Lee-Carter-Modell für den ganzen Datensatz bestimmt. Das heißt, es ist schwieriger Sterblichkeiten vorherzusagen, als sie zu modellieren. Leider sieht man nicht, dass die Vorhersage schlechter wird, wenn man sich weiter von dem Vorhersagezeitraum entfernt. Vermutlich liegt dies daran, dass der γ Wert in dem Beispielmmodell den Modellannahmen sehr genau folgt und γ sehr gut auf einer Linie liegt. Das γ , dass aus dem historischen Datensatz geschätzt wurde, ist weniger gut auf einer Linie. Siehe dazu Abbildung 28. Man sieht, dass γ erst später die Nulllinie erreicht und dann schneller abfällt, als man es von einem RWD erwarten würde. Die Fehler sind außerdem eher positiv als negativ.

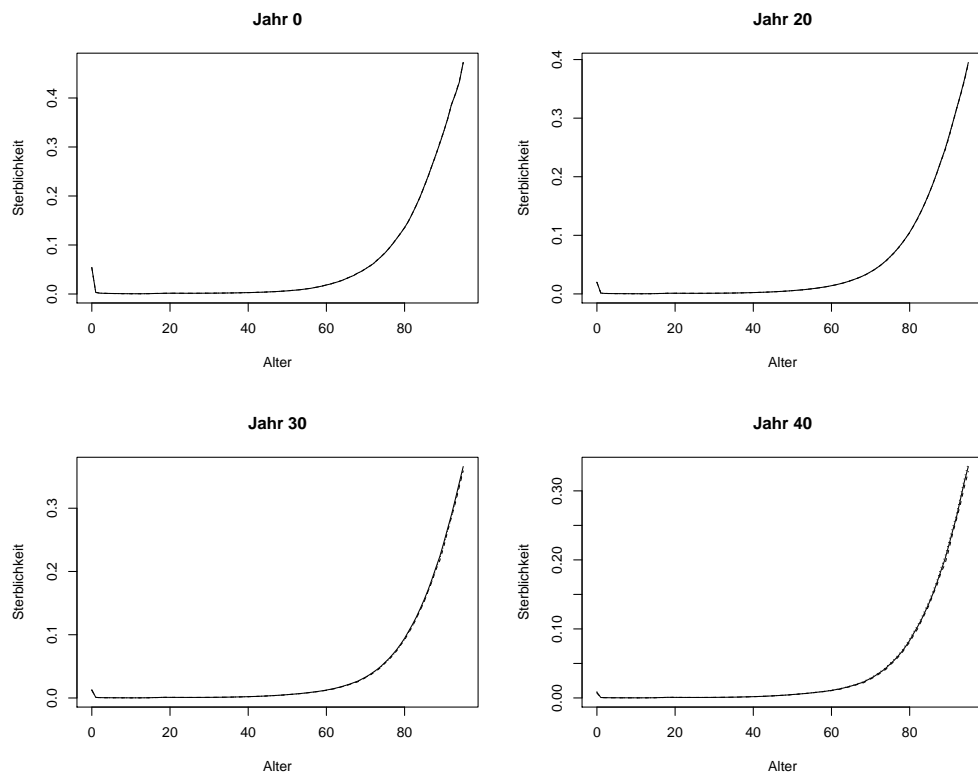


Abbildung 26: Das Lee-Carter-Modell angewendet auf den simulierten Datensatz in vier ausgewählten Beobachtungsjahren: Beginn des Schätzzeitraums (0), Ende des Schätzzeitraums (20), Mitte des Vorhersagezeitraums (30), Ende des Vorhersagezeitraums (40).

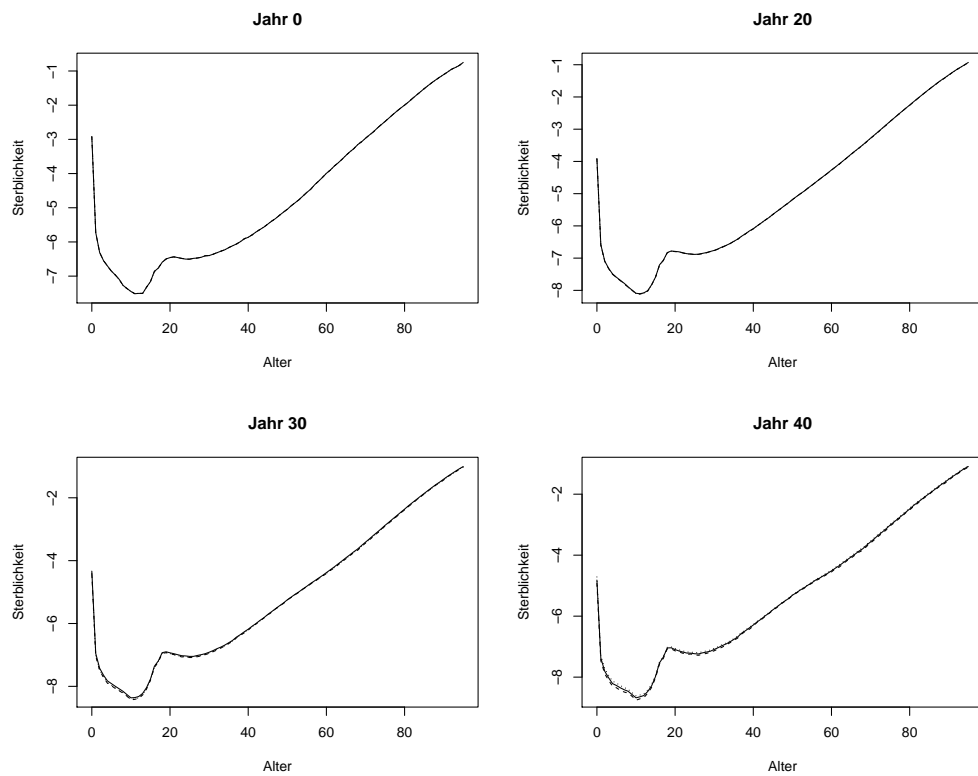


Abbildung 27: Die logarithmierten Werte wie in Abbildung 26.

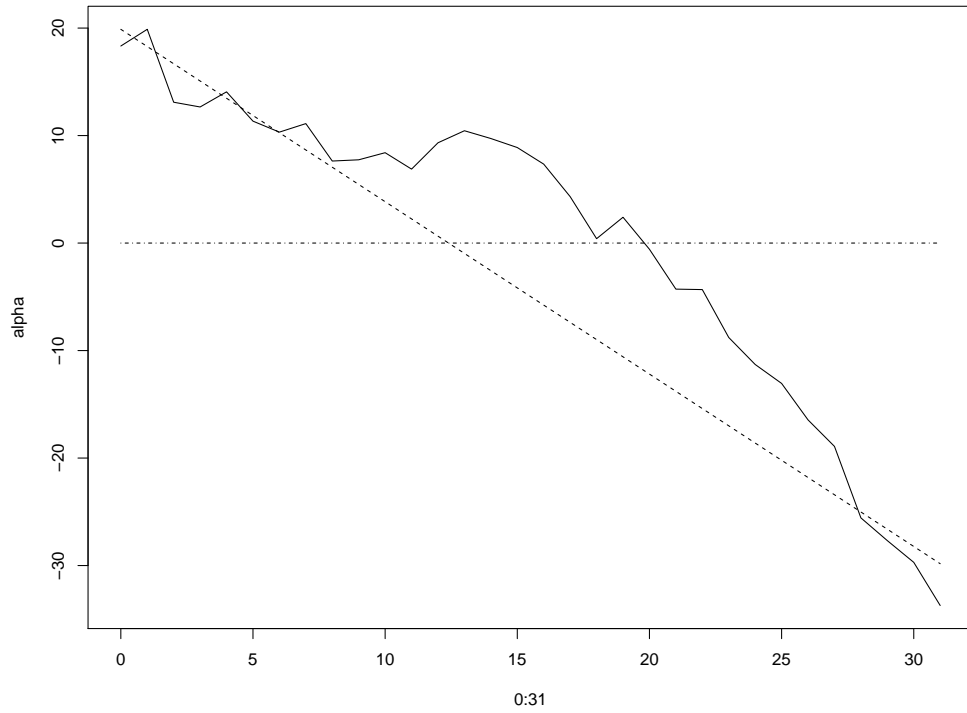


Abbildung 28: Der γ -Wert aus dem historischen Datensatz(durchgezogen), der Wert von $(\max(\hat{\gamma}_S, \dots, \hat{\gamma}_E), \max(\hat{\gamma}_S, \dots, \hat{\gamma}_E) + \nu, \max(\hat{\gamma}_S, \dots, \hat{\gamma}_E) + 2\nu, \dots)$ (dashed) und die Nulllinie(dashdot).

5 Zusammenfassung

Die Aufgabenstellung dieser Arbeit bestand, wie in Kapitel 1 angegeben, darin, einen Überblick über die Erstellung von Sterbetafeln zu geben. Dazu wurden in Kapitel 2 in den ersten drei Abschnitten die zentralen Begriffe der rohen Sterbehäufigkeit und der Sterbetafel definiert und Probleme bei ihrer Bestimmung angegeben. Im darauf folgenden Abschnitt 2.3 wurden verschiedene Glättungsverfahren vorgestellt. Dabei wurde besonders detailliert das Whittaker-Henderson-Verfahren besprochen. Am Schluss des Kapitels wurden historische Daten vorgestellt.

Nachdem in Kapitel 2 die beschreibenden Grundlagen dieser Arbeit gelegt wurden, konnte in Kapitel 3 der schließende Teil dieser Arbeit beginnen. Es wurden verschiedene Regressionsmodelle vorgestellt, mit denen Sterbetafeln erstellt werden können. Insbesondere ein einfaches Modell, das keine Sterblichkeitstrends berücksichtigt, und das Lee-Carter-Modell, das Sterblichkeitstrends berücksichtigt, wurden vorgestellt.

In Abschnitt 4 wurden Eigenschaften der Modelle nachgerechnet und Simulationen durchgeführt. Es wurde insbesondere gezeigt, dass sowohl das einfache Modell, als auch das Lee-Carter-Modell identifizierbar, konsistent und asymptotisch normal sind. Dann wurde für den ν -Parameter ein Konfidenzband (58) bestimmt. Setzt man die Werte der Simulation in Gleichung (58) ein, erhält man das Konfidenzintervall (60). Der Wert $\nu = -1.679$ liegt in der Tat in diesem Konfidenzband.

$$\begin{aligned}
 1 - \alpha &= \mathbb{P} \left(\nu \in \left[\hat{\nu} - z_{1-\alpha/2} \cdot \frac{\frac{(\hat{\gamma}_E - \hat{\gamma}_S)}{2\sigma_\xi^2}}{\sqrt{n}}, \hat{\nu} + z_{1-\alpha/2} \cdot \frac{\frac{(\hat{\gamma}_E - \hat{\gamma}_S)}{2\sigma_\xi^2}}{\sqrt{n}} \right] \right) \\
 &= \mathbb{P}(\nu \in [-1.60371 + 0.4176962, -1.60371 - 0.4176962]) \\
 &= \mathbb{P}(\nu \in [-2.021406, -1.186013])
 \end{aligned} \tag{60}$$

Berechnet man mit dem minimalen und dem maximalen Wert des Konfidenzintervalls für ν ein Konfidenzband für die Sterblichkeit des letzten Vorhersagejahres, erhält man Abbildung 29. Der wahre Wert der Sterblichkeit ist ebenfalls eingezeichnet. Es sieht so aus, als ob der wahre Wert in der Tat in dem Konfidenzintervall liegt und eine Rechnung in dem Programmcode zeigt dies. Es wirkt mit einer maximalen Breite von $\approx 3.0664\%$ auch relativ schmal, allerdings müsste man in einem Beispiel nachrechnen, welchen Einfluss dies auf die Prämienhöhe eines Versicherungsproduktes hat, um die Nützlichkeit in der Praxis beurteilen zu können.

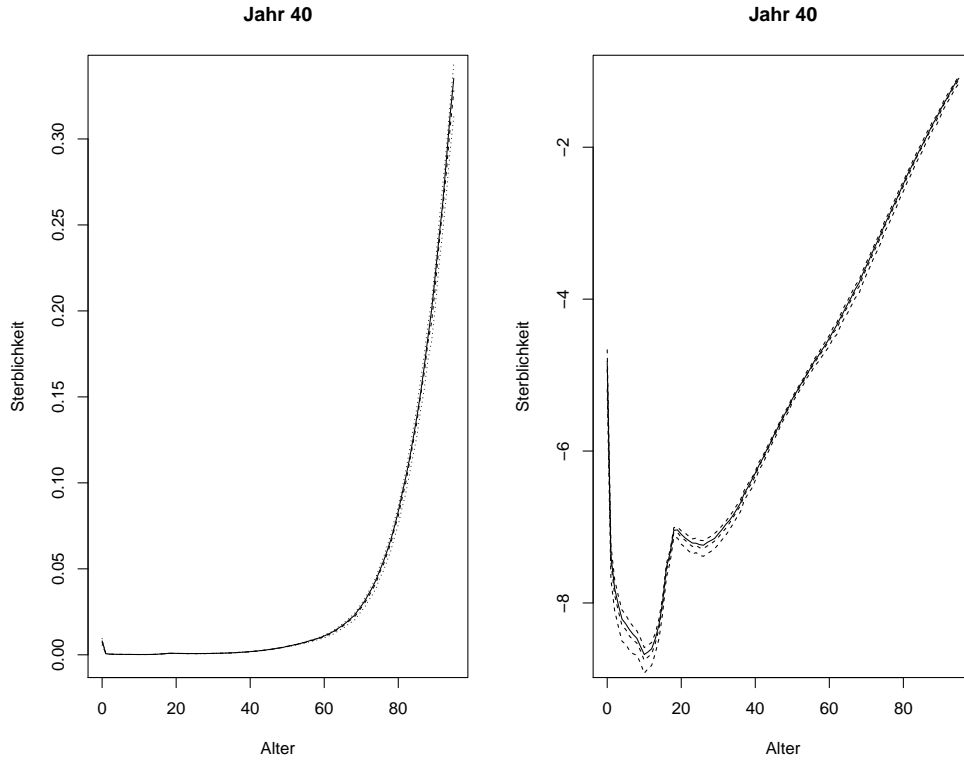


Abbildung 29: Simulation wie in Abbildung 26 mit Konfidenzintervall (gestrichelt) und wahrem Modell (durchgezogen).

Als nächstes sollen die Ergebnisse der Simulation aus Abschnitt 4.5 vergleichend diskutiert werden. Die Ergebnisse sind in Tabelle 4 zusammengefasst. Hier bezeichnet Whittaker das Whittaker-Henderson-Verfahren.

Methode	Einfaches Modell	Whittaker	Lee-Carter	Lee-Carter Vorhersage
$\ q_x - f_\theta(x)\ _2^2$	0.0234060	0.0230494	0.0035201	x
$\ \hat{\theta} - \theta\ _2^2$	0.1364178	x	13.50123	x
$\ \hat{\theta} - \theta\ _2^2 / \dim(\theta)$	0.0000358	x	0.0584469	x
$\ q_x - f_\theta(x)\ _2^2$	x	x	0.0000001	0.0000106

Tabelle 4: Ergebnisse der Simulationen. Alle Werte sind auf 7 Nachkommastellen gerundet. In der obersten Zeile sieht man die Abweichung der Schätzung zum jeweiligen historischen Datensatz, in den nächsten beiden die Abweichung zu dem wahren Parameter und in der letzten die Abweichung zur Vorhersage beziehungsweise zu den wahren Daten im Lee-Carter-Modell.

Betrachten wir zuerst die Abweichung zwischen dem Schätzer für die zweite Hälfte des Datensatzes und den beobachteten Daten. Man sieht, dass diese für das einfache Modell

und das Whittaker-Henderson-Verfahren ähnlich sind. Bei dem Lee-Carter-Modell sind die Fehler dagegen um fast einen Faktor von 10 kleiner. Entsprechend kann man sagen, dass das Lee-Carter-Modell für Sterbetafeln aufgrund dieser Berechnung besser ist. Dies ist genau das Ergebnis, welches man erwarten würde.

Betrachten wir als nächstes, wie gut die Modelle ihre Parameter im eigenen Modell schätzen können. Man sieht, dass die Fehler für das Lee-Carter-Modell um einen Faktor von 2 größer sind, als die Fehler im einfachen Modell. Dies ist auch nicht weiter erstaunlich, da man $2 \cdot x + t + 1$ Parameter schätzt, während man im einfachen Modell nur zwei Parameter schätzt. Deshalb ist in der nächsten Zeile der durchschnittlich Fehler pro Parameter über die 10000 Durchgänge angegeben. Der Fehler im einfachen Modell ist jetzt um einen Faktor von 2 größer, als der Fehler im Lee-Carter-Modell. Bei dem Whittaker-Henderson-Verfahren gibt es keine Parameter und deshalb auch keine Ergebnisse.

Betrachten wir jetzt die letzte Simulation, bei der wieder die Abweichung von einem Datensatz und der Vorhersage angegeben ist. Das Modell, das zur Erzeugung verwendet wurde, ist wieder das Lee-Carter-Modell. Man sieht, dass das Lee-Carter-Modell einen deutlich größeren Fehler macht, wenn es eine Hälfte des Datensatzes vorhersagen muss, als wenn es einfach nur modelliert. Außerdem sieht man, dass der Fehler der Vorhersage im eigenen Modell um einen Faktor von 3 besser ist als auf dem historischen Datensatz.

Insgesamt kann man den Vergleich der Simulation damit zusammenfassen, dass es besser scheint, das Lee-Carter-Modell als das einfache Modell zu verwenden. Dies bedeutet ebenfalls, dass es besser ist mit Periodentafeln zu arbeiten als mit Generationstafeln, da die Transformation von Beobachtungsjahren zu Geburtsjahren zu einer unerwünschten Zensur führt, die das Schätzen von Sterbetafeln deutlich erschwert.

Das Ergebnis der Simulation ist auch das Ergebnis der gesamten Arbeit: Sowohl das einfache Modell als auch das Lee-Carter-Modell erfüllen die drei Eigenschaften Identifizierbarkeit, Konsistenz und asymptotische Normalität. In anderen Worten haben sowohl das einfache als auch das Lee-Carter-Modell die gleichen theoretischen Eigenschaften. Allerdings ist das Lee-Carter-Modell in der Lage, Sterblichkeitstrends zu berücksichtigen und ist somit besser darin, sich ändernde Sterbehäufigkeiten zu modellieren. Das Lee-Carter-Modell, ist also nicht grundlos zum “leading statistical model of mortality [forecasting] in the demographic literature” [GK07] geworden.

Literatur

- [Bel96] E. Beleites, »Grenzen der Medizin: Weil die Medizin in Europa so gut ist ...«, *Dtsch Arztebl International*, Bd. 93, Nr. 6, A–305–, 1996. eprint: <https://www.aerzteblatt.de/pdf.asp?id=357>. Adresse: <https://www.aerzteblatt.de/int/article.asp?id=357>.
- [Bom93] E. Bomsdorf, »Generationensterbetafel Geburtsjahrgang 1987 und Periodensterbetafel 1986/88«, *Blätter der DGVFM*, Bd. 21, Nr. 2, S. 265–276, Okt. 1993, ISSN: 1864-0303. DOI: 10.1007/BF02809409. Adresse: <https://doi.org/10.1007/BF02809409>.
- [DAV04] DAV, *Dav 2004r*, <https://aktuar.de/unsere-themen/lebensversicherung/Seiten/sterbetafeln.aspx>, Accessed: 22-05-2019, 2004.
- [DP01] A. Deaton und C. Paxson, »Mortality, income, and income inequality over time in britain and the united states«, National Bureau of Economic Research, Working Paper 8534, Okt. 2001. DOI: 10.3386/w8534. Adresse: <http://www.nber.org/papers/w8534>.
- [EM86] R. Engle und D. McFadden, Hrsg., *Handbook of Econometrics*, 1. Aufl. Elsevier, 1986, Bd. 4. Adresse: <https://EconPapers.repec.org/RePEc:eee:econhb:4>.
- [Far13] U. Farooq, *Importance of demography*, <http://www.studylecturenotes.com/basics-of-sociology/importance-of-demography>, Accessed: 22-08-2019, 2013.
- [Fer02] T. S. (S. Ferguson, *A course in large sample theory*, English, First edition. London : Chapman und Hall, 2002, Bibliography: p236-237. - Includes index, ISBN: 0412043718.
- [For06] D. O. Forfar, »Mortality laws«, in *Encyclopedia of Actuarial Science*. American Cancer Society, 2006, ISBN: 9780470012505. DOI: 10.1002/9780470012505.tam029. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470012505.tam029>. Adresse: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470012505.tam029>.
- [GK07] F. Girosi und G. King, »Understanding the lee-carter mortality forecasting method«, 2007.
- [Gre81] T. Greville, »Moving-weighted-average smoothing extended to the extremities of the data iii. stability and optimal properties«, *Journal of Approximation Theory*, Bd. 33, Nr. 1, S. 43–58, 1981, ISSN: 0021-9045. DOI: [https://doi.org/10.1016/0021-9045\(81\)90088-5](https://doi.org/10.1016/0021-9045(81)90088-5). Adresse: <http://www.sciencedirect.com/science/article/pii/0021904581900885>.

- [Ham00] M. Hammond, »The forces of mortality at ages 80 to 120.«, *International Journal of Epidemiology*, Bd. 29, Nr. 2, S. 384–384, Apr. 2000, ISSN: 0300-5771. DOI: 10.1093/ije/29.2.384-a. eprint: <http://oup.prod.sis.lan/ije/article-pdf/29/2/384/18478226/290384.pdf>. Adresse: <https://doi.org/10.1093/ije/29.2.384-a>.
- [HR10] R. Henrion und W. Roemisch, »Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions«, *Annals of Operations Research*, Bd. 177, S. 115–125, Juni 2010. DOI: 10.1007/s10479-009-0598-0.
- [KLO+85] P. Kakies, H. Loebus, B. Oehlers-Vogel, B. Zschoyan und H. Behrens, *Methodik von Sterblichkeitsuntersuchungen*, Ser. Schriftenreihe angewandte Versicherungsmathematik. Verlag Versicherungswirtschaft, 1985. Adresse: <https://books.google.de/books?id=GZvRSgAACAAJ>.
- [LC92] R. D. Lee und L. R. Carter, »Modeling and forecasting u. s. mortality«, *Journal of the American Statistical Association*, Bd. 87, Nr. 419, S. 659–671, 1992, ISSN: 01621459. Adresse: <http://www.jstor.org/stable/2290201>.
- [Loe94] H. Loebus, »Bestimmung einer angemessenen Sterbetafel für Lebensversicherungen mit Todesfallcharakter«, *Blätter der DGVFM*, Bd. 21, Nr. 4, S. 497–524, Okt. 1994, ISSN: 1864-0303. DOI: 10.1007/BF02809488. Adresse: <https://doi.org/10.1007/BF02809488>.
- [LR05] E. L. Lehmann und J. P. Romano, *Testing statistical hypotheses*, Third, Ser. Springer Texts in Statistics. New York: Springer, 2005, S. xiv+784, ISBN: 0-387-98864-5.
- [MD19] H. R. Max Roser und B. Dadonaite, *Child and infant mortality*, 2019. Adresse: <https://ourworldindata.org/child-mortality>.
- [Mor18] Mortality, *Mortality*, <https://www.mortality.org/>, Accessed: 22-05-2019, 2018.
- [NS12] A. S. Nocon und W. F. Scott, »An extension of the whittaker–henderson method of graduation«, *Scandinavian Actuarial Journal*, Bd. 2012, Nr. 1, S. 70–79, 2012. DOI: 10.1080/03461238.2010.534257. eprint: <https://doi.org/10.1080/03461238.2010.534257>. Adresse: <https://doi.org/10.1080/03461238.2010.534257>.
- [Vaa98] A. W. v. d. Vaart, *Asymptotic Statistics*, Ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. DOI: 10.1017/CB09780511802256.
- [Wil99] R. Willets, »Mortality in the next millennium«, 1999. Adresse: <http://www.sias.org.uk/papers/mortality.pdf>.
- [Wol] P. D. K. Wolfsdorf, *Versicherungsmagazin*, <https://www.versicherungsmagazin.de/lexikon/selektionstafel-1946621.html>, Accessed: 05-12-2019.

Selbstständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig, nur mit Hilfe meiner Betreuer und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Göttingen,

Henning Hause