

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:

xxx

Erstgutachter:

Frank Werner

Zweitgutachter:

xxx

Studiengang:

M. Sc. Mathematik

Inhaltsverzeichnis

1	Motivation und Aufgabenstellung	2
1.1	Verwendung von Sterbetafeln	2
1.2	Ziele und Aufgabenstellung	3
2	Grundlegende Definitionen und historische Daten	5
2.1	Definitionen relativer Sterbehäufigkeiten und Beschreibung von Personen- gesamtheiten	5
2.2	Ein Problem bei der Schätzung roher Sterbewahrscheinlichkeiten: Welche Personen sind zu berücksichtigen	7
2.3	Weitere Probleme beim Schätzen roher Sterbewahrscheinlichkeiten	11
2.4	Arten von Sterbetafeln	13
2.5	Öffentlich zugängliche Sterblichkeitsdaten	14
3	Modellbildung	18
3.1	Ein einfaches Modell: Geburtsjahr als erklärende Variable	18
3.2	Ein einfaches Modell: Alter als erklärende Variable	20
3.3	Das Vorgehen der DAV	24
3.4	Glättung von Sterbewahrscheinlichkeiten	26
3.4.1	Überblick über Glättungsverfahren	26
3.4.2	Das Verfahren von Whittaker-Henderson	29
3.5	Das Lee-Carter Modell	32
3.6	Verbesserung des Einfachen Modells	41
4	Eigenschaften der Modelle und Simulationen	44
4.1	Das Theorem von Cramér	44
4.2	Eigenschaften des einfachen Modells	49
4.3	Eigenschaften des Whittaker-Henderson Verfahrens	64
4.4	Eigenschaften des Lee-Carter Modells	65
4.4.1	Der Schätzer für β	65
4.4.2	Schätzer für ν	69
4.5	Eigenschaften des einfachen Trendmodells	72
4.6	Qualitätsvergleich durch Simulation	72

1 Motivation und Aufgabenstellung

Die Idee für diese Arbeit ist entstanden, als ich im Winter 2018/2019 ein Praktikum bei einer Lebensversicherung gemacht habe. Dort habe ich eine Programmbibliothek geschrieben, die aus einer Sterbetafel versicherungsmathematische Kennzahlen bestimmt. Da ich einige Vorlesung zu Statistik gehört habe, hat mich das Schätzen von Sterbetafeln interessiert und ich habe in dieser Arbeit verschiedenen Methoden Sterbetafeln zu bestimmen zusammengefasst. Dabei beginnen wir mit einfachen Modellen und entwickeln dann komplizierte Modelle, die auch Sterblichkeitstrends berücksichtigen. Außerdem werden die Eigenschaften der Modelle bestimmt und es werden mittels Simulation durchgeführt. In diesem einleitenden Kapitel wird in Abschnitt 1.1 der Begriff der Sterbetafel durch die Anwendungsgebiete motiviert. Dann werden in Abschnitt 1.2 die Ziele dieser Arbeit genauer diskutiert.

1.1 Verwendung von Sterbetafeln

Sterbetafeln werden vor allem in der Lebensversicherungswirtschaft und in der Demographieforschung 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 selbsterklärend, während man unter den biometrischen Risiken eine Reihe von Risiken zusammenfasst, die alle die körperliche Unversehrtheit betreffen. Beispiele hierfür sind vor allem Tod und Invalidität, aber man kann auch das Risiko zu heiraten betrachten, wie bei der Aussteuerversicherung. Ein einfaches Beispiel für ein Netto-Versicherungsprodukt, das heißt ein Versicherungsprodukt ohne Kosten, ist die Einmalprämie einer n -jährigen Erlebnisfallversicherung mit Versicherungssumme 1 Geldeinheit für eine Person mit Alter x . Der Wert dieses Produktes kann über die folgende Gleichung beschrieben werden:

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

Dabei bezeichnet l_x die Lebenden im Alter x und i ist der Rechenzins. 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 Sterbewahrscheinlichkeiten und wir werden in dieser Arbeit Methoden sehen q_x zu schätzen. Eine Sterbetafel ist dann eine Tabelle, in der 1-jährige Sterbewahrscheinlichkeiten q_x tabelliert sind.

Betrachten wir jetzt noch kurz das andere Anwendungsgebiet für Sterbetafeln, die Demographieforschung. Laut der Website [Far13] beschäftigt sich die Demographieforschung 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 anderen Bereich, der auf der Website [Far13] nicht erwähnt wird, aber trotzdem relevant ist ist die staatliche Rentenpolitik.

1.2 Ziele und Aufgabenstellung

In dem obigen Abschnitt 1.1 wurden einige Anwendungsgebiete von Sterbetafeln zusammengefasst. In diesem Abschnitt werden die Ziele dieser Arbeit vorgestellt.

Ein Problem bei der Verwendung von Sterbetafeln ist, dass sich die Sterblichkeit im Laufe der Zeit verändert. Genauer gesagt, hat sich die Lebenserwartung in den letzten Jahrzehnten erhöht. Diese Veränderung in der Lebenserwartung muss auch in Sterbetafeln berücksichtigt werden. Betrachten wir dazu zwei Säulendiagramme, die die Entwicklung der Sterblichkeit der Jahre 1990 und 2017 in der Bundesrepublik Deutschland zeigen:

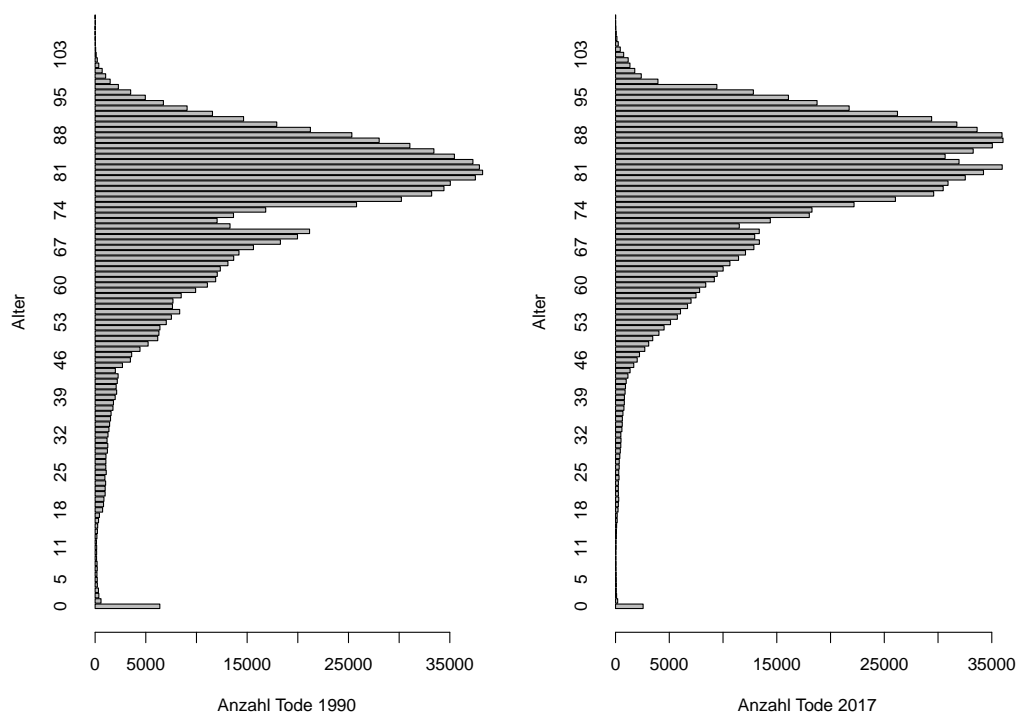


Abbildung 1: Anzahl an Toden in der Bundesrepublik Deutschland. Auf der linken Seite im Jahr 1990 und auf der rechten Seite im Jahr 2017

Wenn man die beiden Diagramme vergleicht, sieht man, dass im Jahr 2017 im Vergleich zum Jahr 1990 deutlich mehr Personen später sterben. Das Diagramm für das Jahr 2017 ist nach oben, zu den höheren Altern, verschoben. Diese Erhöhung der Lebenserwartung nennt man Sterbetrends. Weil es auch in der Praxis sehr relevant ist, wie man

diese Sterblichkeitstrends schätzt, beschäftigt sich auch die Deutschen Aktuarvereinigung (DAV) mit diesem Problem. Mit der Sterbetafel 2004R hat die DAV eine Richtlinie für deutsche Personenversicherer herausgegeben. Die Dokumentation für die Sterbetafel 2004R [DAV04], sowie weiterer von der DAV veröffentlichte Sterbetafeln, können unter <https://aktuar.de/unsere-themen/lebensversicherung/Seiten/sterbetafeln.aspx> eingesehen werden und waren der Ansatzpunkt für diese Arbeit.

Diese Arbeit beginnt im nächsten Kapitel 2 damit grundlegende Definitionen zu geben und historische Daten und ihre Qualität zu diskutieren. Für die Definitionen wurde das Buch "Methodik von Sterblichkeitsuntersuchungen" [Beh85] von Peter Kakiès verwendet. Die historischen Daten stammen von der Website <https://www.mortality.org>, [Mor18], wo man auch weitere Informationen zu den Daten finden kann.

Im nächsten Kapitel, 3, werden verschiedene Modelle betrachtet, mit denen Sterbetafeln erstellt werden. Zuerst wird ein einfaches Modell betrachtet, das zum Vergleich mit den üblicherweise verwendeten Modellen dienen sollen. Daraufhin wird diskutiert, wie die DAV Sterbetafeln erstellt. Insbesondere wird das Whittaker-Henderson Verfahren zur Glättung von Sterbetafeln betrachtet. Dazu wird wieder das Buch "Methodik von Sterblichkeitsuntersuchungen" [Beh85] von Peter Kakiès verwendet. Außerdem wurden "An extension of the Whittaker-Henderson method" von Nocon und Scott, [NS12], sowie "Bestimmung einer angemessenen Sterbetafel für Lebensversicherungen mit Todesfallcharakter" von Loebus, [Loe94] verwendet. Ein weiteres häufig verwendetes Modell ist das Lee-Carter Modell, das auch diskutiert wird. Für dieses Modell wird die Originalquelle "Modeling and Forecasting U. S. Mortality" von Lee und Carter, [LC92], sowie "Understanding the Lee-Carter Mortality Forecasting Method" von Girosi und King [GK07] verwendet. Am Ende des Kapitels 3 wird dann eine Verbesserung für das einfache Modell aus dem Anfang von Kapitel 3 vorgeschlagen.

In Kapitel 4 wird für die Modelle aus Kapitel 3 Identifizierbarkeit, starke Konsistenz und asymptotische Normalität mit dem Satz von Cramér gezeigt. Danach werden die Modelle mittels Simulationen verglichen. Diese Simulationen, sowie fast alle Graphiken in dieser Arbeit wurden in **R** erstellt. Der Code für dieses Diagramm, sowie alle anderen Codes, die in dieser Arbeit verwendet werden, 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** für die Programmiersprache **R**.

2 Grundlegende Definitionen und historische Daten

In diesem Kapitel werden im ersten Abschnitt grundlegende Begriffe wie Kollektiv und relative Sterbehäufigkeiten geklärt. In den nächsten zwei Abschnitten werden Probleme bei der Bestimmung rohen Sterbewahrscheinlichkeiten angegeben. Diese ersten drei Abschnitte orientieren sich an dem Buch "Methodik von Sterblichkeitsuntersuchungen", siehe [Beh85]. Im nächste Abschnitt werden verschiedene Arten von Sterbetafeln angegeben. Im letzten Abschnitt dieses Kapitels werden Daten vorgestellt, auf deren Grundlage Sterbetafeln erstellt wurden und werden. Insbesondere wird auf die öffentlich zugänglichen Sterbetafeln der Website mortality.org eingegangen, siehe [Mor18].

2.1 Definitionen relativer Sterbehäufigkeiten und Beschreibung von Personengesamtheiten

Da das Ziel dieser Arbeit das Schätzen von Sterbewahrscheinlichkeiten ist, müssen wir Notation einführen, um Sterblichkeit zu definieren. Dazu definieren wir zuerst verschieden Arten von Jahren und was Personen und Personengemeinschaften sind:

- Jahre: Seien $\iota, \tau, \gamma \in \mathbb{N}$ Jahre. Dabei bezeichne ι immer ein Kalenderjahr, τ ein Geburtsjahr und γ immer ein Todesjahr. Daraus werden gleich Personen definiert. Ich verwende γ für das Todesjahr, da wir später eine Regression von einem passend gewählten X auf das zufälligen Todesalter Y machen möchten. Die analoge Größe zum zufälligen Todesalter ist das zufällige Todesjahr, dass mit Γ bezeichnet wird.
- Personen: Seien P_1, \dots, P_n Personen. Eine Person P_i ist definiert durch ein Geburtsjahr τ_i und durch ein Todesjahr γ_i . Ist $\iota = \gamma_i$ ist die Person P_i im Jahr ι verstorben.

Aus diesen Definitionen kann man direkt die folgenden Begriffe ableiten:

$$\begin{aligned}\text{Alter} &= X = \text{Kalenderjahr} - \text{Geburtsjahr} = \iota - \tau, \text{ für } \iota \leq \gamma \\ \text{Todesalter} &= Y = \text{Todesjahr} - \text{Geburtsjahr} = \Gamma - \tau \\ \text{Lebenserwartung} &= LE = \mathbb{E}(Y)\end{aligned}$$

Seine weiterhin E und S jeweils End- und Startalter einer Person. Wir setzen in dieser Arbeit $S = 0$ und $E = 95$. Alle Alter größer 95 die beobachtet werden, werden auf 95 gesetzt. Außerdem unterscheiden wir Personen nicht nach Geschlecht.

- Kollektive: Sei weiterhin K_j eine Menge von lebenden Personen P_i . Das heißt, aus $P_i \in K_j$ folgt, dass $\gamma_i < \iota_i$. Alle Personen unter Beobachtung werden mit K bezeichnet. Der Index j dient dazu, Personen in Risikoklassen zu unterteilen. Zum Beispiel $J = \{1900, \dots, 2000\}$, also eine Unterscheidung nach Geburtsjahren. Sei J immer derart, dass $K = \cup_{j \in J} K_j$. Die Bezeichnung K kommt daher, dass man von einem Kollektiv spricht. Ein Synonym zu Kollektiv ist eine Personengesamtheit. Wir

bezeichnen die Kardinalität einer Personengemeinschaft K_j mit k_j also $|K_j| = k_j$ beliebige Personengemeinschaften.

- Austrittsgründe: Es gibt verschiedene Gründe für den Austritt aus einem Kollektiv. Von besonderem Interesse ist der Austritt durch Tod. Andere Austrittsgründe können Invalidität oder Auswanderung sein. Invalidität wird in dieser Arbeit keine Rolle spielen. Wir werden Auswanderung später als Störgröße diskutieren. Man unterscheidet weiterhin offene und geschlossene Personengesamtheit. Falls aus einer Personengesamtheit Personen nur durch Tod ausscheiden, nennen wir diese Personengesamtheit geschlossen.

Führen wir nun Zufallsvariablen(Z.V.) ein: Sei Y als das zufällige Todesalter und y als eine Realisierung von Y . Analog ist das Todesjahr zufällig und wird mit Γ bezeichnet. Eine Realisierung von Γ wird mit γ bezeichnet. Die Notation soll unterstreichen, dass wir Γ_i beziehungsweise Y_i voraussagen wollen, gegeben dass wir für eine Person P_i die Größe τ_i kennen. Außerdem sind wir in einem bestimmten Kalenderjahr ι . Das heißt, wir wollen eine Methode M finden um $Y_i \approx M(\iota_i, x_i)$ sinnvoll zu schätzen. Aus den Größen Γ beziehungsweise Y wird in der Literatur dann q_x abgeleitet:

$$\text{1-jährige-Sterbewahrscheinlichkeit} = q_x = \mathbb{P}(Y \leq x+1 | Y > x), x \in [S, E] \subset \mathbb{N}_0 \quad (1)$$

Das heißt, q_x ist die Wahrscheinlichkeit, dass eine Person, die bereits x Jahre alt geworden ist, vor ihrem $x+1$ -ten Geburtstag verstirbt. Diese Wahrscheinlichkeiten nennt man dann die 1-jährige-Sterbewahrscheinlichkeit

- Teilgesamtheit der Ausgeschiedenen: Um die q_x für $x \in \{S, S+1, \dots, E\}$ zu schätzen, betrachtet man Kollektive und zählt die Personen, die zu einem bestimmten Kalenderjahr ι , das Risiko haben zu sterben und im darauf folgenden Jahr $\iota+1$, die Anzahl an Todesfällen. Das konkrete vorgehen wird im nächsten Abschnitt definiert. An dieser Stelle sollen nur die beiden Teilmengen T und K' definiert werden: Wir bezeichnen mit T die Teilmenge der Toten aus K und mit K' die Teilmenge der Überlebenden. Seien weiterhin die Kardinalitäten mit t beziehungsweise k' bezeichnet. Es gilt:

$$\text{Kollektiv} = K = T \cup K' = \text{Tote} \cup \text{Überlebende}$$

- Rohe Sterbehäufigkeiten: Mit den Definitionen der Teilgesamtheiten kann man rohe (1-jährige) Sterbewahrscheinlichkeiten schätzen:

$$\hat{q}^{roh} = \frac{t}{k} = \frac{t}{t+k'} \quad (2)$$

Man betrachtet also ein Kollektiv und schätzt daraus die Sterbewahrscheinlichkeit. Das in dieser Arbeit meistens betrachtete Kollektiv ist die west-deutsche Gesamtbevölkerung. Warum genau dieses Kollektiv betrachtet wird, wird in dem Abschnitt 2.5 diskutiert.

- Basistafel und Sterbetafel: Es fehlt noch die Definition einer Sterbetafel beziehungsweise einer Basistafel. Leider werden diese beiden Größen in der Literatur nicht definiert, sondern intuitiv verwendet. Eine Sterbetafel ST ist intuitiv $ST = (q_S, q_1, \dots, q_E)$. Allerdings hat man mit dieser Definition das Problem, dass man gerne Sterbewahrscheinlichkeiten pro Geburtsjahrgang angeben möchte. Das heißt es interessiert ST_1, \dots, ST_m für Indizes $1, \dots, m \in \{\tau_1, \dots, \tau_m\}$. Deshalb sei $B = (q_S, q_1, \dots, q_E)$ eine Basistafel und $ST = (B_1, \dots, B_m)$ eine Sterbetafel.

2.2 Ein Problem bei der Schätzung roher Sterbewahrscheinlichkeiten: Welche Personen sind zu berücksichtigen

Im letzten Kapitel wurde bereits definiert, was man unter einer rohen Sterbewahrscheinlichkeit versteht. In diesem Kapitel werden die verschiedenen Methoden rohe Sterbewahrscheinlichkeiten zu gewinnen diskutiert. Es gibt die Geburtsjahrmethode, die Sterbejahrmethode, das Sterbeziffernverfahren und die Verweildauermethode. Diese vier Methoden existieren, weil nicht ganz klar ist welche Personen zu berücksichtigen sind. Dieser Sachverhalt soll an Graphik 2 veranschaulicht werden:

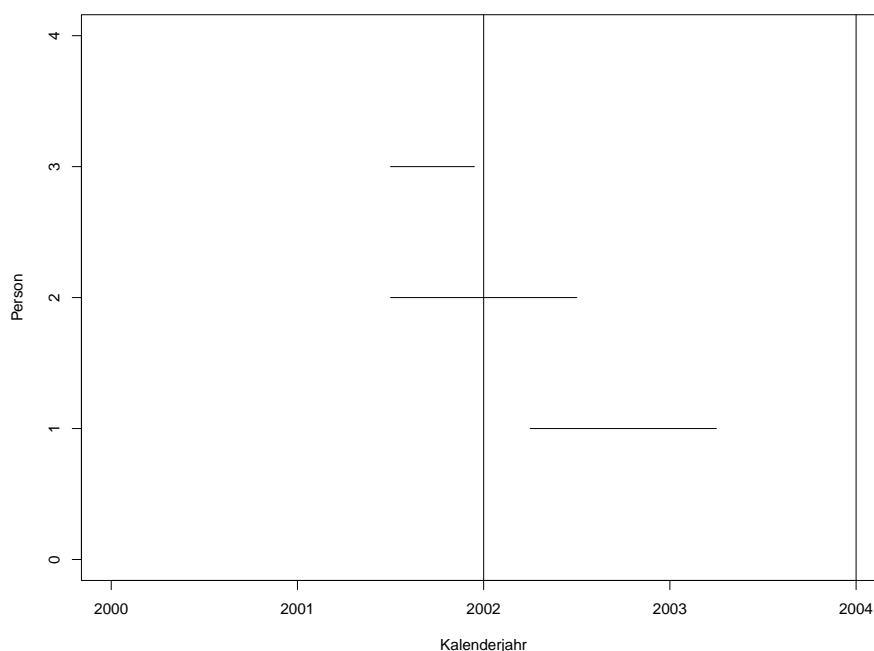


Abbildung 2: Lebensdauer von verschiedenen x -jährigen Personen. Die Frage ist, welche Personen zu berücksichtigen sind.

Zu sehen sind die Lebensdauer von drei x -jährigen Personen. Die Lebensdauer der Personen sind durch einen horizontalen Strich dargestellt. Der Beginn zeigt den Eintritt ins Kollektiv durch Erreichen des Alters x und das Ende den Austritt aus dem Kollektiv der x -jährigen. In diesem Fall gehen wir von Tod aus und die Frage ist, welche die mittlere zu erwartende Lebensdauer einer x -jährigen Person ist. Wir beobachten den Zeitraum 2002 bis 2004, der durch senkrechte Linien eingegrenzt wird. Man sieht, dass Person 1 innerhalb des Beobachtungszeitraums in das Kollektiv eintritt und innerhalb des Beobachtungszeitraums ausscheidet. Das heißt diese Person kann ohne Probleme erfasst werden. Person 2 dagegen ist bereits vor Beginn der Beobachtungsperiode in das Kollektiv eingetreten. Hier stellt sich nun die Frage, ob diese Person erfasst werden sollte. Das Problem ist, dass es auch Personen wie Person 3 gibt, die gleichzeitig mit Person 2 in das Kollektiv eingetreten sind, allerdings bereits vorher das Kollektiv verlassen haben. Man würde also, wenn man Person 2 mit in die Schätzung aufnimmt und Person 3 nicht, die Sterblichkeit von x -jährigen Personen mit dem selben Geburtsjahr unterschätzen.

Bevor wir die verschiedenen Lösungsmethoden betrachten, sei auf Tabelle (1) hingewiesen, in der die bisherige und für diesen Abschnitt relevante Notation zusammengefasst ist.

Betrachten wir jetzt die verschiedenen Methoden, um rohe Sterbewahrscheinlichkeiten zu bestimmen. Beginnen wir mit der Geburtsjahrmethode. Diese berücksichtigt nur die Personen, die innerhalb des Beobachtungszeitraums in das Kollektiv eintreten. Das heißt,

Variable	Bedeutung
ι	Kalenderjahr
τ	Geburtsjahr
γ	Todesjahr
P_i	Person i
X	Alter
Y	Todesalter
E	Endalter einer Sterbetafel
S	Startalter einer Sterbetafel
K, K_j	Kollektiv, Kollektiv j
k_i	Kardinalität von K_i
T	Teilmenge von K , die aus K ausscheidet
t	Kardinalität von T
$t_{x,n}(b, \tau)$	Teilmenge von K die mit Alter x und Geburtsjahr τ im Beobachtungszeitraum b ausscheiden, nachdem sie mindestens Zeit n in Jahren zum Kollektiv gehört haben
$K' = L$	Teilmenge von K die nicht ausscheidet
$k' = l'$	Kardinalität von $K' = L$
$l_x(b, \tau)$	Teilmenge der Überlebenden im Alter x im Beobachtungszeitraum b und Geburtsjahr τ
q_x	Sterbewahrscheinlichkeit
b	Beobachtungszeitraum
b_i	i -te Jahr des Beobachtungszeitraums b
ST	Sterbetafel
B	Basistafel

Tabelle 1: Bisher verwendete Notation

alle Personen mit dem selben Geburtsjahr werden gleich behandelt und man erhält die folgende Gleichung (3):

$$q_x^G = t_{x,1}(b, \tau) / l_{x,1}(b, \tau) \quad (3)$$

Bei Gleichung (3) bezeichnet der Index G das es sich um die Geburtsjahrmethode handelt und die 1 im Index der Anzahl an Toten und Lebenden, dass die Person ein Jahr zum Kollektiv gehört haben muss. Wir gehen von einer Beobachtungsperiode von mindestens einem Jahr aus. Der Parameter b steht für den Beobachtungszeitraum und τ für das Geburtsjahr der Personen, die berücksichtigt werden.

Alternativ gibt es die Sterbejahrmethode (4), bei der man alle in einem Beobachtungszeitraum gestorbenen berücksichtigt. In Formeln erhält man

$$q_x^S = \frac{t_x(b)}{l_x(b_1/b_2, \tau_1/\tau_2) - 1/2 \cdot l_x(b_2, \tau_2) + 1/2 l_x(b_1, \tau_0)} \quad (4)$$

wobei eine Reihe an Korrekturen vorgenommen wurden, um bessere Schätzer zu erhalten. Auf diese Korrekturen wird hier nicht weiter eingegangen. Mehr Informationen kann man in [Beh85] finden. In [Beh85] wird als nächstes das Sterbezeiffernverfahren beschrieben. Dieses Verfahren ist technisch einfacher als die Sterbejahrmethode. Es handelt sich um ein Anfangs-Plus-Endbestand-Halbe-Methode. Zuerst muss man die Sterbezeiffer wie in (5) definieren. Hier sind ι_1 und ι_2 Anfangs-beziehungsweise Endjahr von b_1 .

$$z_x = \frac{t_x(b_1, \tau_0/\tau_1)}{1/2 \cdot (l_x(\iota_1) + l_x(\iota_2))} \quad (5)$$

Damit erhält man als rohe Sterbezeiffer die folgende Gleichung (6):

$$q_x^Z = \frac{2 \cdot z_x}{2 + z_x} \quad (6)$$

Abschließend gibt es noch das Verweildauerverfahren (7). Dieses Verfahren erfasst die Teilzeiten, die Personen zu einem Kollektiv gehören und löst damit das Problem welche Personen zu erfassen sind am besten. Allerdings ist es auch das aufwendigste. Personenversicherer sind eventuell in der Lage Sterblichkeiten in ihren Beständen durch diese Methode zu schätzen, aber um Bevölkerungssterblichkeiten zu schätzen ist es zu kompliziert. Um die Sterblichkeit in einer Bevölkerung mit dieser Methode zu schätzen, müsste man tagesgenau wissen wann eine Person Teil des zu beobachtenden Kollektivs geworden ist und wann es dieses verlässt. Diese Informationen wie Geburten, Tode, Emigrationen und Immigrationen werden sicherlich nicht immer erfasst und sind auch nicht immer akkurat. Nehmen wir als Gedankenexperiment eine Person, die in Deutschland geboren ist, in die USA emigriert und dann in einem Urlaub in Zentralamerika verstirbt. Ich vermute, dass solch ein Todesfall nicht immer in Deutschland gemeldet wird.

$$q_x^D = \frac{t_x(b)}{t_x(b) + \sum_d d \cdot l'_{x,d}(b)}, 0 < d \leq 1 \quad (7)$$

Insgesamt lässt sich feststellen, dass die Methoden aufwändiger werden. Dies liegt daran, dass auch Personen erfasst werden, die auch nur einen Teil der Beobachtungsperiode zu dem betrachteten Kollektiv gehören. Das heißt, dass insbesondere die Geburtsjahrmethode geeignet ist, wenn man ein geschlossenes Kollektiv betrachtet, während die Verweildauer-methode insbesondere bei offenen Kollektiven geeignet ist.

Als letzten Punkt möchte ich noch erwähnen, dass in Deutschland 1910/1911 und 1932/34 sowie 1949/1951 die Geburtsjahrmethode angewandt wurde. Für die Sterbetafel von 1960/1962 wurde die Sterbejahrmethode angewandt. Das heißt, wenn man Daten aus diesen Sterbetafeln zusammenfasst um zum Beispiel Sterblichkeitstrends zu Schätzen, sollte man diese Änderung in der Bestimmung der rohen Sterbewahrscheinlichkeit berücksichtigen.

2.3 Weitere Probleme beim Schätzen roher Sterbewahrscheinlichkeiten

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

- Jahreszeitliche Schwankungen: Wie man in der folgenden Graphik (3) sieht, schwankt die Sterblichkeit, je nach Jahreszeit. Das heißt, je nachdem zu welchem Monat der Beobachtungszeitraum beginnt und endet hat eventuell einen Einfluss auf die Sterbetafel.

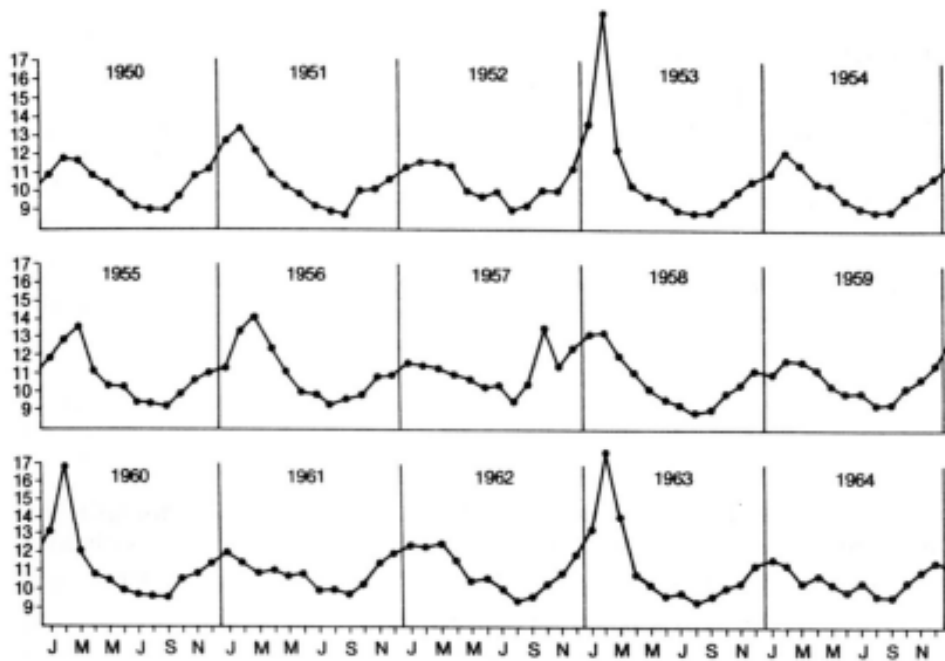


Abb. 10. Monatlische, auf ein Jahr normierte Sterbeziffern 1950–1964 in Prozent.

Abbildung 3: Diese Graphik zeigt jahreszeitliche Schwankung der Sterblichkeit. Diese Graphik ist Abbildung 10 in [Beh85] Seite 4.2

- Extremereignisse: Ein weiteres Problem ist, wie man mit Extremereignissen umgeht. Sowohl die DAV in der Dokumentation zur Sterbetafel 2004R, als auch Lee und Carter 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].

2.4 Arten von Sterbetafeln

In diesem Abschnitt werden Periodentafeln, Kohortentafeln, Generationstafeln und Selektionstafeln definiert. Die Definitionen hier sind recht locker gehalten. Zum Schluss werden noch als Beispiele für Sterbetafeln die 2004R und 2008 diskutiert. In dieser Arbeit werden Periodentafeln und Generationstafeln benutzt

- **Periodensterbetafeln/Aggregattafeln:** Um eine Periodensterbetafel zu erstellen wird meist über einen relativ kurzen Zeitraum ein Kollektiv bestehend aus Personen P_1, \dots, P_n betrachtet. Aus diesem Kollektiv werden dann, im einfachsten Fall, die beiden Kennziffern I_x und t_x bestimmt. Dabei bezeichnet 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 die Anzahl an Todesfällen von Personen im Alter x . Man bestimmt dann mittels der folgenden Gleichung (8) eine rohe Sterbewahrscheinlichkeit.

$$q_x^{\text{roh}} = \frac{t_x}{I_x} \quad (8)$$

Welche Personengruppen im Detail zum Schätzen der q_x betrachtet werden, wurde in Kapitel 2.2 diskutiert. Hier sei nur noch einmal darauf hingewiesen, dass Sterblichkeiten von gleichzeitig lebenden Personen ermittelt werden und daraus q_x geschätzt wird. Das Geburtsjahr geht nicht als Faktor in die Betrachtung mit ein. Deshalb sind solche Tafeln langfristiger Fragestellungen ungeeignet.

- **Kohortensterbetafeln:** Im Gegensatz zu Periodentafeln sind Kohortensterbetafeln Tafeln, die das Absterben eines Kollektivs mit Geburtsjahr τ beschreiben. Das heißt, man beginnt mit einem Kollektiv von n Personen $P_1^\tau, \dots, P_n^\tau$ und erfasst jedes Jahr die Anzahl an Todesfällen. Bei einer solchen Tafel ist das Geburtsjahr als Faktor relevant. Leider sind solche Daten nur schwer zu beschaffen. Dies liegt daran, dass es bei der Gesamtbevölkerung eines Landes 2 Eintritts- und Austrittsgründe gibt. Zum einen kommen Personen in das Kollektiv durch Geburt, zum anderen durch Einwanderung. Bei den Austrittsgründen gibt es analog dazu Tod und Auswanderung. Dieses Problem sieht man auch in den Daten: Betrachtet man die Anzahl an Personen in Deutschland, die in einem gewissen Jahr geboren sind und das Risiko haben, zu sterben, so ist diese Anzahl nicht monoton fallend. Das heißt, es kommen in jedem Alter neue Personen in das Kollektiv, das betrachtet wird. Da diese Daten nicht zu beschaffen sind, betrachtet man die nachfolgende Art von Tafeln.
- **Generationssterbetafel:** Unter einer Generationensterbetafel versteht man eine Sterbetafel, bei der die Sterblichkeit nicht nur vom Alter (und eventuell vom Geschlecht),

sondern zusätzlich vom Geburtsjahrgang abhängt. Es handelt sich also im Prinzip um eine Kohortentafel. Der Unterschied ist, dass bei einer Generationstafel nicht am Anfang ein Kollektiv steht, sondern eine Gesamtbevölkerung, bei der auch das Geburtsjahr τ erfasst wird. Somit ist eine Generationstafel eine Mischung aus einer Periodentafel und einer Kohortentafel. Die Daten haben mehr Information, sind allerdings trotzdem erfassbar. Es gibt allerdings nach wie vor das Problem, dass man mit zensierten Daten arbeitet. Das heißt, man keine Beobachtungen darüber, wie Personen mit einem bestimmten Geburtsjahr τ in Zukunft versterben. Dies muss mit statistischen Verfahren ausgeglichen werden.

- Selektionstafel: Da der Bestand von Lebensversicherern oft nicht repräsentativ für die Gesamtbevölkerung ist, gibt es hier eigene Tafeln. Konkret ist die Mortalität im Bestand von Lebensversicherern niedriger als die Mortalität der Gesamtbevölkerung. Da man um solche Tafeln zu erstellen Bestandsdaten von Lebensversicherern braucht, werden solche Tafeln hier nicht erstellt.

Ein Beispiel für eine Sterbetafel ist die Sterbetafel 2004R. Dabei steht das R für Rente. Das heißt, diese Tafel wird für Versicherungsprodukte mit Erlebnisfallcharakter, wie die Rente, konzipiert. Erlebnisfallprodukte zeichnen sich dadurch aus, dass die versicherte Person nach eine Leistung erhält, wenn sie einen bestimmten Zeitraum n überlebt. In diesen Produkten, wird die Sterblichkeit als Rechnungsgrundlage neben dem Rechnungszins und den Kosten mit einbezogen, da nur bei Erleben geleistet wird. 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 sicher zu gehen, 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 2. Ordnung interessiert.

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 über keine Schätzung von Sterbetrends. Es ist sicherer für Versicherungen mit Todesfallcharakter kein Verbesserung der Sterblichkeit an zu nehmen.

2.5 Öffentlich zugängliche Sterblichkeitsdaten

Die Sterbetafel 2004R basiert laut Kapitel 2 der Herleitung der Sterbetafel 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 1998/200 (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 Jahren

Die Daten der Münchner Rück und der Gen Re sind Privateigentum der jeweiligen Unternehmen und deshalb nicht öffentlich zugänglich. Auch die Daten des Statistischen Bundesamtes 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

Man sieht anhand Tabelle, 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 Kalenderjahr 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 Graphiken und Simulationen dieser Arbeit ausgeführt wurden, sind deshalb die Daten im Format "Kalenderjahr 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 zuordenbar sind. In diesen Fällen werden die Beobachtungen proportional auf die verschiedenen Kategorien verteilt. Dies kann dann dazu führen, dass man bei Todeszahlen kein ganze 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 Prozentszahl entsteht durch zwei Tode, die sich auf 0.5 Personen unter Todesrisiko aufteilen. Hier die entsprechenden Zeilen von der Website. Die Tode sind zuerst aufgeführt:

Year	Age	Female	Male	Total
1987	109	2.00	0.00	2.00
Year	Age	Female	Male	Total
1987	109	0.50	0.00	0.50

Betrachte man die zugehörigen rohen Daten 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

Wir haben uns im letzten Abschnitt grundlegende Definitionen und Probleme bei der Datenerfassung gesehen. In diesem Kapitel werden Modelle eingeführt, um die Sterblichkeitsraten zu schätzen. Wir beginnen mit einfachen Modellen. 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 Abschnitt 3.3 sehen, wie die DAV ihre Sterbetafeln schätzt. In der Motivation habe ich bereits gesagt, dass die Herleitung der Sterbetafel 2004R der Ansatzpunkt dieser Arbeit ist und in diesem Kapitel wird die Sterbetafel 2004R diskutiert. Zwei Methoden, die in der Sterbetafel 2004R erwähnt werden, werden in den Abschnitten 3.4.2 und 3.5 Abschnitt diskutiert. Diese Methoden sind das Whittaker-Henderson Verfahren zum Glätten von Sterbewahrscheinlichkeiten und das Lee-Carter Modell zum Trendschätzen.

Alle Schätzmethoden zu den Modellen sind in dem beiliegenden Paket *moRtRe*, der Name steht für "Mortality Trends", für die Daten aus Kapitel 2.5 implementiert.

3.1 Ein einfaches Modell: Geburtsjahr als erklärende Variable

Beginnen wir mit dem Modell, dass den Todeszeitpunkt in Abhängigkeit vom Geburtszeitpunkt angibt. Wir benutzen die folgenden Definitionen:

- Bezeichne $\gamma = (\gamma_1, \dots, \gamma_n)$ einen Vektor von Todesjahren.
- Bezeichne $\tau = (\tau_1, \dots, \tau_n)$ einen Vektor von Geburtsjahren.
- Bezeichne $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ einen Vektor von Fehlern. Zu diesem Zeitpunkt machen wir keine Annahmen über die Fehler.

Und gehen von einem funktionalen Zusammenhang

$$\gamma = f(\tau) + \epsilon$$

aus. Dann ist das Todesalter $\gamma - \tau$ und die Lebenserwartung $\mathbb{E}(\gamma - \tau)$. Sei weiterhin die Zufallsvariable X definiert durch $X = f(\tau) + \epsilon$ mit Dichte f_X . Dann kann man eine Sterbetafel aufstellen mittels:

$$\begin{aligned} q_x &= \mathbb{P}(\gamma - \tau < x + 1 | \gamma - \tau > x) \\ &= \frac{\mathbb{P}(\gamma - \tau < x + 1 \cap \gamma - \tau > x)}{\mathbb{P}(\gamma - \tau > x)} \\ &= \frac{\mathbb{P}(f(\tau) + \epsilon - \tau < x + 1 \cap f(\tau) + \epsilon - \tau > x)}{\mathbb{P}(f(\tau) + \epsilon - \tau > x)} \\ &= \frac{\int_x^{x+1} f_X(y) dy}{\int_x^\infty f_X(y) dy} \end{aligned}$$

Offensichtlich hat dieses Modell das Problem, dass das Alter einer zu versichernden Person bekannt ist und diese Info nicht berücksichtigt wird. Ein anderes Problem ist, dass bei Anwendung auf die Daten um so später eine Person geboren ist, um so früher stirbt sie. Siehe dazu Graphik (4)

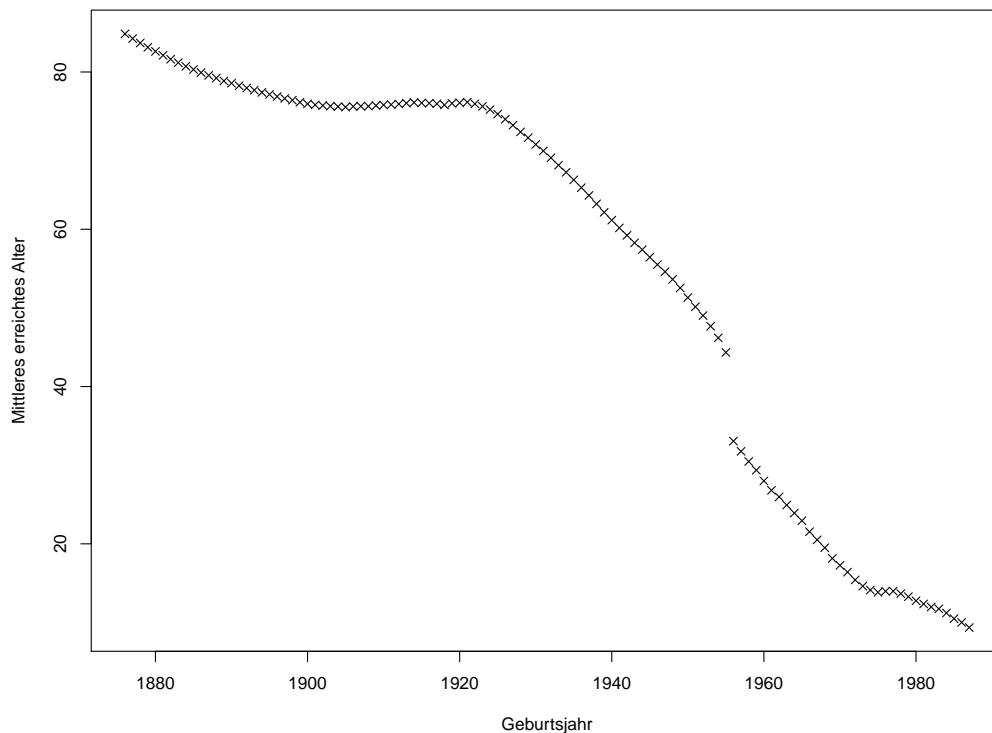


Abbildung 4: Ein Plot der Geburtsjahr versus mittleres erreichtes Alter bei der Generationsstafel 1876 West zeigt.

Dies macht auch Sinn, wenn man einen Moment darüber nachdenkt: Das mittlere Alter, in dem wir Personen sterben sehen, die 1950 geboren sind, ist natürlich höher als das mittlere Alter von Personen, die 2017 geboren sind. Hier sieht man deutlich das Problem mit der Zensur der Daten. Zum Abschluss dieses Kapitels betrachten wir noch eine Heatmap, das die Anzahl an Todesfällen in Abhängigkeit zu Geburtsjahr und Alter beschreibt:

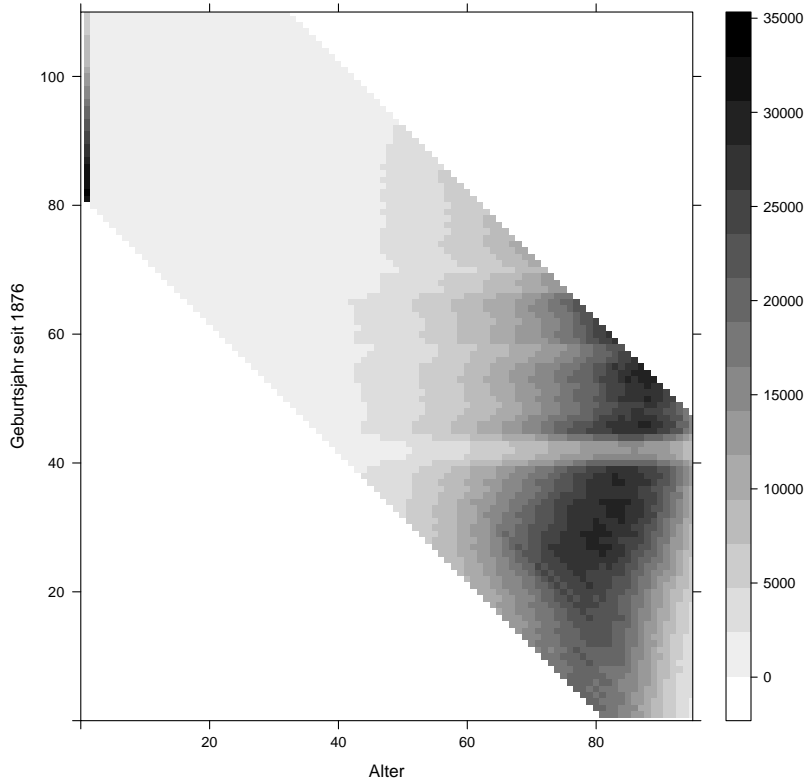


Abbildung 5: Ein Heatmap der Todesfälle der Generationstafel 1876 West. Auf der x-Achse ist das Geburtsjahr und auf der y-Achse ist das Alter der Personen.

Man sieht, 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 nur das Geburtsjahr nicht als einzige erklärende Variable.

3.2 Ein einfaches Modell: Alter als erklärende Variable

Betrachten wir als nächstes ein weiteres einfaches Modell, doch diesmal mit dem Alter als erklärende Variabel. Beginnen wir wieder mit den Definitionen:

- Bezeichne $Y = (Y_1, \dots, Y_n)$ einen Vektor mit der Anzahl an Todesfällen im Alter i geteilt durch die Gesamtzahl an Todesfällen. Das heißt, sei $b = (\iota_1, \iota_2)$ ein Beobachtungszeitraum und γ, τ das Geburtsjahr beziehungsweise Todesjahr der Person P , dann ist Y_i definiert durch:

$$Y_i = \frac{\#\{P \text{ mit } \gamma \in b \text{ und } \gamma - \tau = i\}}{\#\{P \text{ mit } \gamma \in b\}}$$

- Bezeichne $X = (X_1, \dots, X_n)$ einen Vektor mit Altern. Dabei ist $X_1 = S$ das Startalter und X_n das Endalter
- Bezeichne $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ einen Vektor mit Fehlern.

Wir unterstellen einen funktionalen Zusammenhang zwischen X und Y

$$Y = f(X) + \epsilon \quad (9)$$

Betrachten wir einen Plot der Daten (6):

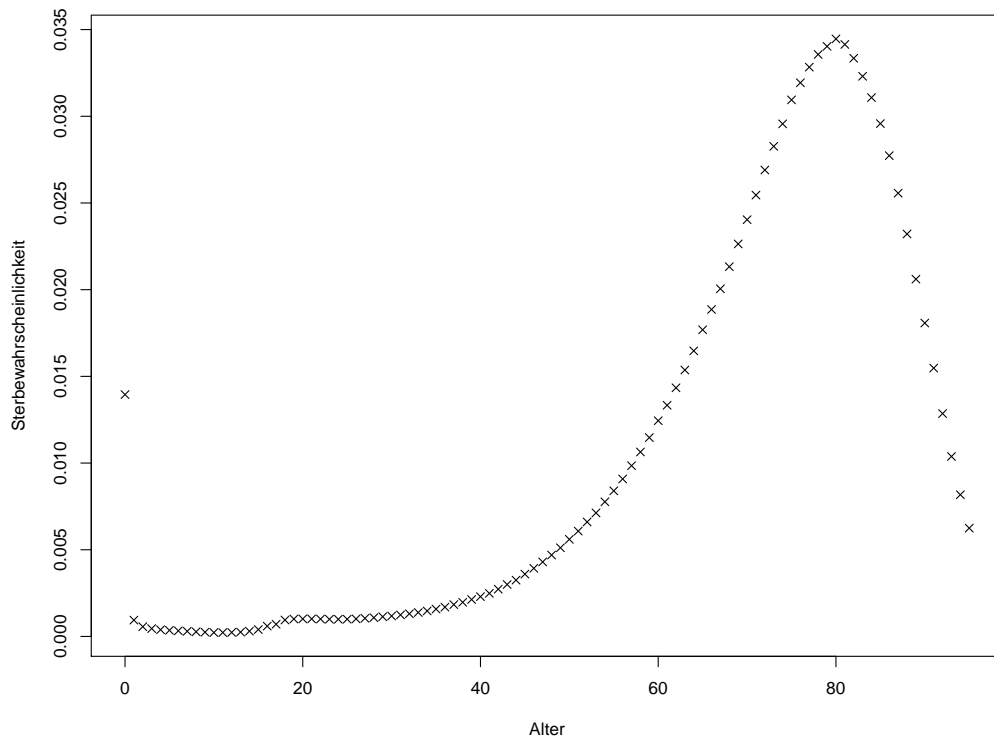


Abbildung 6: Alter versus relative Anzahl an Sterbefällen pro Alter bei der Generationstafel 1876 West

Man sieht, dass der funktionale Zusammenhang f nicht linear ist. Zum Beispiel in [GK07] wird das folgende Modell vorgeschlagen, dass als Lee-Carter Modell bezeichnet wird: Bezeichne $m_{a,t}$ den Logarithmus der Sterblichkeit der Altersgruppen $a(a = 1, \dots, A)$ und $t(t = 1, \dots, T)$ die Zeit. Dann geht man von

$$m_{a,t} = \alpha_a + \beta_a \gamma_t + \epsilon_{a,t} \quad (10)$$

aus. Hier bezeichnen α_α , β_α und γ_t Parameter, die zu schätzen sind und $\epsilon_{a,t}$ zufälligen Schwankungen. Logarithmiert man nun die Sterbewahrscheinlichkeiten und passt dann ein lineares Modell an, so erhält man folgende Graphik (7):

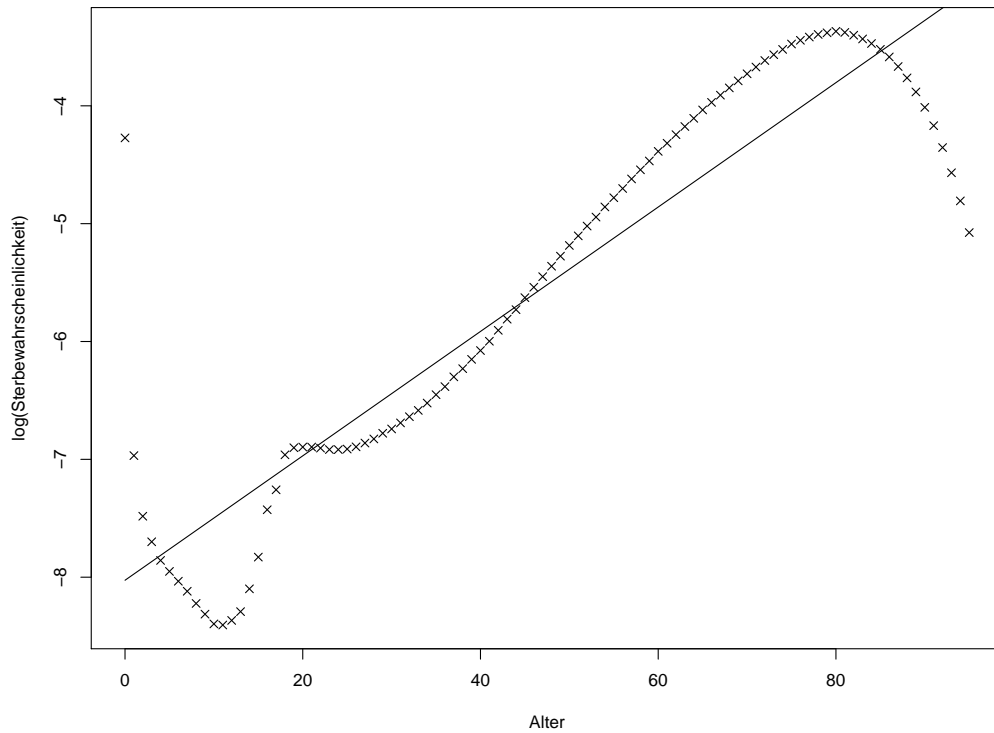


Abbildung 7: Alter versus logarithmierte relative Anzahl an Sterbefällen pro Alter bei der Generationstafel 1876 West. Außerdem ist eine Regressionsgerade eingezeichnet bei der mit der einfachen kleinsten Quadraten Methode (OLS) $\log(Y)$ auf X regressiert wurde.

Durch die Logarithmierung werden die Daten schon deutlich linearisiert, allerdings noch nicht zu einem Grad, der mir ausreichend scheint.

Betrachten wir noch einmal Graphik (6). Die relativen Sterbehäufigkeiten ergeben in Summe 1, sprich $\sum_{x=S}^E Y_i = 1$. Jede Person stirbt in unserem Modell, spätestens bei Erreichen des Alters E . Diese Eigenschaft haben auch Dichten. Es liegt also nahe, Y_i für alle i als Ausschnitt aus einer Dichte zu interpretieren. Das heißt, es scheint sinnvoll den funktionalen Zusammenhang f als Dichte zu betrachten. Die bekannteste Dichte ist die Normalverteilung, und die Daten scheinen optisch auch recht gut zu einer Normalverteilung zu passen. Wir legen also das folgende Modell an:

$$Y = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon \quad (11)$$

Bei Modell (11) sind alle Operationen vektorwertig zu verstehen. Das heißt, $Y, X \in \mathbb{R}^{E-S+1}$. In folgender Graphik sieht man noch einmal die Daten aus Graphik (6) und zusätzlich der geschätzte, funktionale Zusammenhang f :

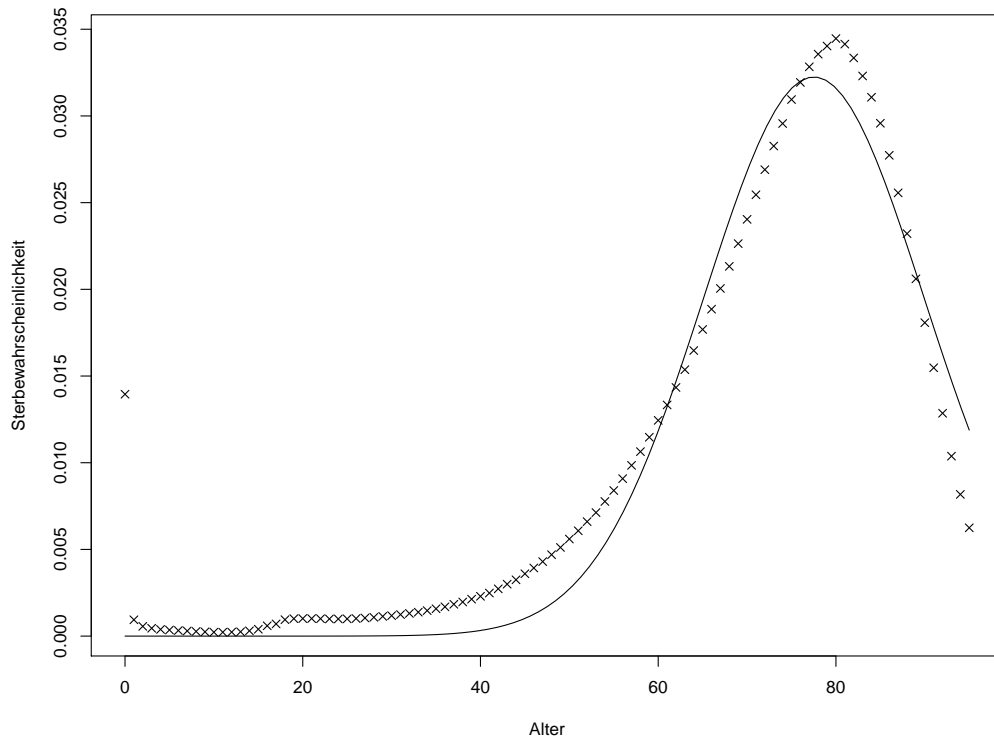


Abbildung 8: Noch einmal das Alter versus die Sterblichkeit bei der Generationstafel 1876 West. In der durchgezogenen Linie ein Schätzer für das Modell.

Die Interpretation der Parameter ist in diesem Modell recht einfach: Der Parameter m ist das mittlere erreichte Lebensalter und der Parameter s ist die Standardabweichung von dem mittleren erreichten Lebensalter. Hier heißt m der Lageparameter Mitte im Sinne des Mittelwertes zu verstehen.

Dieses Modell ist das Basismodell. Das heißt, die anderen beiden Modelle müssen zeigen, ob sie besser sind. Es hat sicherlich viele Schwächen. So hat man beispielsweise Masse im negativen Bereich. Das heißt, man hat eine von Null verschiedene Wahrscheinlichkeit mit einem negativen Alter zu sterben. Außerdem wird das Geburtsjahr als erklärende Variable vernachlässigt. Allerdings ist es als Vergleichsfall sehr praktisch, da man davon ausgehen kann, dass dieses Modell schlechter ist, als die anderen, die in dieser Arbeit betrachtet werden.

3.3 Das Vorgehen der DAV

Die Sterbetafel 2004R geht in zwei Schritten vor: 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 die 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 Kapitel 3.3.1 der Herleitung der Tafel 2004R. Man beachte, dass die 2004R zum einen eine Tafel für die Rentenbezugszeit und zum anderen für die Aufschubszeit hat. Da es sich bei der Tafel zur Rentenbezugszeit um eine Selektionstafel handelt, wird deren Herleitung hier nicht betrachtet.

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

$$\frac{\sum_{t=1995}^{2002} t \cdot VWD(t)}{\sum_{t=1995}^{2002} VWD(t)} \quad (12)$$

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 Randaltern am geringsten sei dürfte.

Als nächstes muss man für das Basisjahr rohe Sterbewahrscheinlichkeiten bestimmen. Wie bereits in Abschnitt 2.5 beschrieben werden zur Bestimmung der rohen Sterbewahrscheinlichkeiten die beiden Kennzahlen I_x und t_x für ein Kollektiv bestimmt. Dabei bezeichnet 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 die Anzahl an Todesfällen von Personen im Alter x . Man bestimmt dann

$$q_x^{\text{roh}} = \frac{t_x}{I_x} \quad (13)$$

Dieses Verfahren nennt man Verweildauerverfahren und wir haben es bereits in Kapitel 2 gesehen. Es findet auch in der Sterbetafel 2008T Anwendung.

Als nächstes wird die Basistafel geglättet. Diesen Vorgang haben wir bisher nicht gesehen, werden ihn aber in Abschnitt 3.4 im Detail betrachten. Dieses Vorgehen ist bereits recht alt, weshalb eine Vielzahl von Methoden zur Verfügung stehen.

Nachdem man eine Basistafel geschätzt und diese geglättet hat, schlägt die DAV vor hohe Alter zu extrapolieren. Die DAV Tafel 2004R hat in der Basistafel nur Alter bis 99. Die Alter 100 bis 120 werden mittels der Methodik aus [Ham00] The force of mortality at Ages 80 to 120 geschätzt. Thatcher gibt eine Reihe an Modellen an und vergleicht ihre Güte. Er kommt zu dem Schluss, dass das logistische Modell das beste für die ihm vorliegenden Daten ist. Hier eine Auswahl aus Modellen, die Thatcher vorschlägt und dann vergleicht:

- Gompertz-Modell: Das Gompertz-Modell (1825) ist das erste Modell, das entdeckt wurde und ist auch das einfachste:

$$q_x = \alpha \exp(\beta x)$$

Im Gompertz-Modell geht man davon aus, dass sich die Sterblichkeit auf eine gleichmäßige exponentielle Art ansteigt. Gompertz entdeckte dieses Sterblichkeitsgesetz, indem er die Sterbewahrscheinlichkeiten der, im zur Verfügung stehenden Sterbetafeln analysierte. Trotz seiner Einfachheit hat es in den vergangenen 170 gute Vorhersagen geliefert.

- Makehams Gesetz: Makehams Gesetz ist eine einfache Modifikation von Gompertz Modell. Es wird ein konstanter, additiver Term c auf das Gompertz Modell addiert, um Risiken, die nicht vom Alter abhängen, zu berücksichtigen:

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

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

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

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

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

Die DAV schätzt dann alle Parameter für die Basistafel und dann wird auf Grund von Gütekriterien das beste Modell ausgewählt. Konkret werden drei herangezogen:

- Vergleich von gemäß extrapolierten Sterbewahrscheinlichkeiten erwarteten Toten mit tatsächlichen Toten $\sum_{x=96}^{99} (\hat{q}_x L_x) / \sum_{x=96}^{99} \hat{q}_x L_x$
- Wert der Loglikelihood (ohne konstanten Summanden): $\sum_{x=96}^{99} T_x \ln(\hat{q}_x) + \sum_{x=96}^{99} (L_x - T_x) \ln(1 - \hat{q}_x)$

- Wert der χ^2 -Statistik: $\sum_{x=96}^{99} (T_x - L_x \hat{q}_x)^2 / (L_x \hat{q}_x (1 - \hat{q}_x))$

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 DAV 2004R Anhang 1 Teil B gefunden werden. Als Sekundärliteratur zu diesen Modellen wird Willets Mortality in the next millennium [Wil99] empfohlen. Da diese Quelle 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, dass wir in Abschnitt 3.5 genauer betrachten werden.

3.4 Glättung von Sterbewahrscheinlichkeiten

Dieses Kapitel basiert auf [Beh85] Kapitel 3. Die DAV hat für die Whittaker-Henderson Verfahren die Quelle [Loe94] empfohlen. Diese fand ich nicht sehr tief gehend und habe deshalb die Quelle [NS12] verwendet.

3.4.1 Überblick über Glättungsverfahren

Im Kapitel 2 haben wir gesehen, wie man rohe Sterbewahrscheinlichkeiten $Q = \{q_S, \dots, q_E\}$ schätzt. Das Ziel von Ausgleichsmethoden ist dann, diese Werte sinnvoll abzuändern. Früher wurde der Ausgleich mittels Millimeterpapier und Bleistift vorgenommen, doch da diese Art Daten auszugleichen sehr subjektiv ist, wurden objektivere Verfahren verwendet. Der theoretische Hintergrund ist die Ausgleichsverfahren als eine Abbildung $q_x \mapsto q'_x$ zu betrachten. Das heißt, man führt eine Funktion F ein mit:

$$q'_x = F(Q, x, a_k(Q, x)), k = 1, \dots, r$$

Bei der Definition der Funktion F , die auch als Ausgleichsfunktion bezeichnet wird bezeichnet x 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.

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

- Mechanische Ausgleichsrechnung: F ist eine lineare Funktion der Form:

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

Die a_k sind unabhängig von x und von q_x . Man beachte, dass k 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 verschieden Vorzeichen haben, hofft man, dass sich einige Fehler gegenseitig aufheben. Die mechanische Ausgleichsrechnung hat eine glättende Wirkung.

- Analytische Ausgleichsrechnung: F ist eine nichtlineare Funktion in x und a_k . Die Parameter a_k können von x und q_x abhängig sein. Die allgemeine Darstellung ist:

$$q'_x = F(x, a_k(x, Q))$$

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 SterbeGesetz, 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.

- Zusammenhang zwischen den Methoden: Der Zusammenhang zwischen den mechanischen und den analytischen Ausgleichsverfahren sieht man, wenn man von dem folgenden Modell ausgeht:

$$q_x = f(x) + \epsilon \tag{14}$$

Dabei bezeichnet ϵ zufällige Fehler und f ist eine unbekannte Funktion. Man geht also davon aus, dass es ein SterbeGesetz gibt, dass 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:

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

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

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

- Beispiele mechanischer Verfahren: Ein bekanntes Verfahren ist die 9-Punkte-Formel von Schärtlin, die q'_x wie folgt bestimmt:

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

Ein anderes ist die 15-Punkte-Formel von Spencer, bei der q'_x durch den folgenden Term bestimmt wird:

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

Ein anderes bekanntes Verfahren ist das Whittaker-Henderson-Verfahren, dass gesondert vorgestellt wird.

- Beispiele analytischer Ausgleichsverfahren: Eine Klasse analytischer Ausgleichsverfahren basiert darauf, dass man die Anzahl an Lebenden als Folge betrachtet und ein Vorschrift angibt, wie man weiter Folgenglieder berechnen kann. Zwei Möglichkeiten sind die SterbeGesetze von de Moivre (1724) und Lambert (1765):

$$l'(x) = l(0) \left(1 - \frac{x}{86} \right) \text{ de Moivre}$$

$$l'(x) = 10.000 \left(\frac{96-x}{96} \right)^2 - 6176 \left(e^{\frac{x}{13,68}} - e^{\frac{x}{2,43}} \right) \text{ Lambert}$$

Die nächsten SterbeGesetze, die entwickelt wurden, sind die von Gompertz (1824) und Makeham (1860). Diese sind von der folgenden Form, wobei k, s, g, c geschätzt werden müssen:

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

Nach der Wahl eines SterbeGesetzes müssen als nächstes die Parameter bestimmt werden. Dazu bieten sich verschiedene Möglichkeiten an. Allen ist gemeinsam, dass

die Beobachtungen in k Gruppen zusammengefasst werden. Dies liegt daran, dass die Anzahl der beobachteten Werte im allgemeinen wesentlich größer als die Anzahl der zu bestimmenden Parameter a_k ist. Dies würde zu überbestimmten linearen Gleichungssystemen führen, die fast immer unlösbar sind. Ein bekanntes Verfahren zum bestimmen der Parameter ist das Verfahren der kleinsten Quadrate, bei dem folgende Bedingung erfüllt werden muss:

$$\sum_{x=x_0}^{x_0+n} g(x, a_k) (F(x, a_k) - q_x)^2 = \text{Minimum}$$

Eine andere ist das Verfahren von King-Hardy, bei dem die Werte in k Gruppen multiplikativ zusammengefasst werden. Es ist durch die folgende Gleichung gegeben:

$$\Pi_{i=0}^{m-1} F(x_0 + (v-1)m + i, a_k) = \Pi_{i=0}^{m-1} q_{x_0+(v-1)m+i} \text{ für } v = 1, \dots, r$$

Eine dritte Möglichkeit ist das Verfahren der Spline-Funktionen nach Reinsch:

$$\int_{x_0}^{x_0+n} \left(\frac{d^2}{dx^2} F(x, a_k) \right)^2 dx = \text{Minimum}$$

mit der Nebenbedingung

$$\sum_{i=0}^n \frac{(F(x_0 + i, a_k) - q_{x_0+i})^2}{g_i} \leq A$$

Schlussendlich sein noch das Verfahren der Ordinatenmomente erwähnt:

$$\sum_{x=x_0}^{x_0+n} x^v F(x, a_k) = \sum_{x=x_0}^{x_0+n} x^v q_x$$

3.4.2 Das Verfahren von Whittaker-Henderson

Wie bereits oben erwähnt, werden wir im Folgenden das Verfahren von Whittaker-Henderson betrachtet. Die Idee dieser Konstruktionsmethode ist, die Bestimmungsgleichungen für die a_k aus einer Optimierungsaufgabe herzuleiten. Sei das Bestimmungsmaß definiert durch:

$$\sum_{k=0}^n (q'_{x_0+k} - q_{x_0+k})^2 \quad (15)$$

Sei weiterhin das Glättemaß definiert durch:

$$\sum_{k=0}^{n-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{x_0+k+v} \right)^2 \quad (16)$$

Das Glättemaß wird in der Literatur auch mit $\Delta^s q'_{x_0+k}$ bezeichnet. Insgesamt erhält man, wenn man Gleichung (15) und (16) kombiniert die Optimierungsaufgabe

$$\sum_{k=0}^n w_k (q'_{x_0+k} - q_{x_0+k})^2 + g \cdot \sum_{k=0}^{n-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{x_0+k+v} \right)^2 = \text{Minimum}_{q'} \quad (17)$$

dabei 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 und für $g \rightarrow \infty$ erhält man ein Polynom zweiter Ordnung.

Zur Lösung des Optimierungsproblems (17) bietet es sich an, das Problem in Matrixschreibweise zu formulieren. Diese Reformulierung orientiert sich an [NS12]. Sei K die $(n-p) \times n$ Matrix:

$$K_{i,j} = \begin{cases} (-1)^{p+j-i} \binom{p}{j-i} \\ 0 \end{cases}$$

Damit lässt sich das Optimierungsproblem (17) schreiben als:

$$\begin{aligned} & \sum_{k=0}^n w_k (q'_{x_0+k} - q_{x_0+k})^2 + g \cdot \sum_{k=0}^{n-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{x_0+k+v} \right)^2 \\ &= (Q' - Q)^t W (Q' - Q) + g \cdot \sum_{k=0}^{n-s} \left(\sum_{v=0}^s K_{k,v} q'_{x_0+k+v} \right)^2 \\ &= (Q' - Q)^t W (Q' - Q) + g (K q')^t (K q') \end{aligned}$$

Dieses Problem ist von der Form:

$$(u - y)^t (u - y) + (Ku)^t H (Ku) = \text{Minimum}_u \quad (18)$$

Für Probleme der Form (18) zeigt Greville ([Gre81], Seite 54), dass eine Lösung durch die folgende Gleichung gegeben ist:

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

Setzt man beim Optimierungsproblem von Whittaker-Henderson, dass ist Gleichung (17), die Parameter $H = g \cdot I$ und $w_k = 1$ für alle k erhält ein Problem der Form (18). Da die Matrix $(I + g K^t K)$ positiv definit und damit invertierbar ist erhält man durch Invertieren der Matrix $(I + g K^t K)$ eine Lösung:

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

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

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

Betrachten wir jetzt ein Beispiel. Es wird ein Datensatz wie in der folgenden Tabelle generiert:

Jahr	Alter	Anzahl an Toden
1900	0	$f(0, \mu, \sigma, 1900)$
\vdots	\vdots	\vdots
1900	95	$f(95, \mu, \sigma, 1900)$

Dieser Datensatz wird wie folgt generiert:

1. Starte mit einem Vektor X , der die möglichen Alter enthält $X = (0, \dots, 95)$
2. Berechne dann die fehlerfreien Y Werte $f(X)$ gemäß:

$$f(X) = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) \quad (19)$$

Dabei wird $m = 70$ und $s = 15$ benutzt.

3. Erzeuge zufällige Fehler $\epsilon \sim N(0, \sigma_\epsilon^2)$ mit $\sigma_\epsilon^2 = 4$
4. Berechne $Y = f(X) + \epsilon$

Das heißt es wird das Regressionsmodell aus Kapitel 3.2 verwendet:

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

Die Parameter für m und s sind so gewählt, dass sie nah an den geschätzten Parametern aus den Daten von [Mor18] sind. Der Glättungsparameter λ wird aus $\{2^r : r = 0, \dots, 10\}$ ausgewählt. Das heißt für $\lambda \in \{2^r : r = 0, \dots, 10\}$ wird der GCV Wert bestimmt und das λ mit dem kleinsten Wert wird verwendet.

Das Ergebnis der Parameterschätzung ist in der folgenden Graphik (9) 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 $f(X)$.

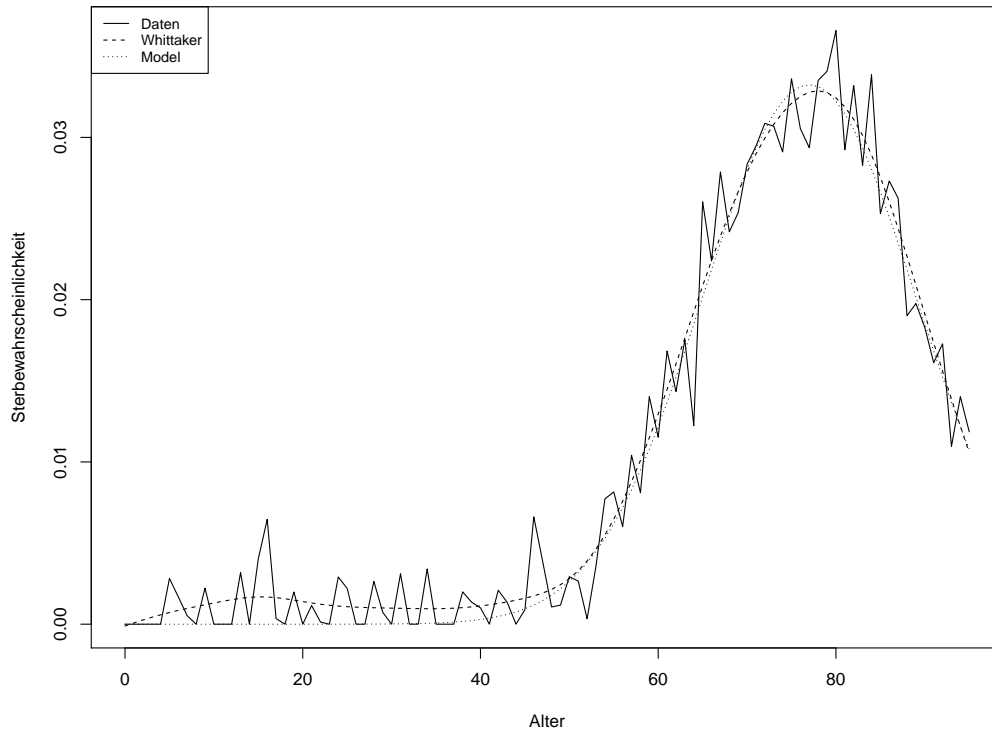


Abbildung 9: Glättung von Daten mittels Whittaker

Man sieht, dass das geglättete Modell dem wahren Modell stark ähnelt.

3.5 Das Lee-Carter Modell

Wir haben bereits in der Motivation in Abschnitt 1 gesehen, dass die Sterblichkeit in der Gesamtbevölkerung in Deutschland in den letzten Jahren zurückgegangen ist. Deshalb werden wir in diesem Abschnitt statistische Modelle angeben, 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 U.S. Sterblichkeiten entwickelt. Laut [GK07] ist es mittlerweile zum leading statistical model of mortality [forecasting] in the demographic literature"geworden. Dieses Zitat hat [GK07] aus Deaton and Paxsons Mortality, Income, and Income Inequality over Time in Britain and the United States"(2004) [DP01] entnommen. Für weitere Informationen für welche Zwecke genau das Lee-Carter Modell verwendet wurde siehe die Einleitung zu [GK07].

Ein dem Lee-Carter Modell sehr ähnliches Modell benutzt Bomsdorf [Bom93], um deutsche Sterblichkeiten zu schätzen. Sein Ansatz ist

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

dabei bezeichnet x ein Alter, t ein Jahr 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 $m_{a,t}$ die logarithmierten Sterblichkeitsraten in den Altersgruppen $a \in \{S, S+1, \dots, E\}$ und zu den Zeitpunkten $t = 1, \dots, T$. Die Sterblichkeitsraten $m_{a,t}$ werden modelliert durch

$$m_{a,t} = \alpha_a + \beta_a \gamma_t + \epsilon_{a,t} \quad (20)$$

hier sind α_a, β_a und γ_t Parameter, die geschätzt werden müssen. Bei $\epsilon_{a,t}$ handelt es sich um zufällige Störungen. Die Parametrisierung (20) ist nicht eindeutig, da sie unter den folgenden beiden Transformationen invariant ist:

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

Dies führt dazu, dass die Likelihood-Funktion, die zu Modell (20) gehört mehr als ein Maximum hat. Deshalb werden in der Literatur zwei Nebenbedingungen eingeführt. Zum einen fordert man

$$\sum_t \gamma_t = 0 \quad (21)$$

und zum anderen

$$\sum_a \beta_a = 1 \quad (22)$$

Stellt man nun das Modell (20) nach γ_t um erhält man

$$\gamma_t = \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\beta_a}$$

Setzt man diese Gleichung für γ_t nun in die Nebenbedingung (21) ein, erhält man

$$\begin{aligned} 0 &= \sum_t \gamma_t = \sum_t \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\beta_a} \\ \iff 0 &= \sum_t (m_{a,t} - \alpha_a - \epsilon_{a,t}) \\ \iff \sum_t \alpha_a &= \sum_t m_{a,t} - \sum_t \epsilon_{a,t} \\ \iff \bar{m}_a &:= \frac{1}{t} \sum_t m_{a,t} = \alpha_a \end{aligned}$$

Dabei benutzt man, dass die Fehler in den Altersgruppen a über die Zeit t im Erwartungswert Null sind und deshalb in der Rechnung als Null angenommen werden. Dadurch lässt sich das Modell (20) neu schreiben, indem man $\alpha_a = \bar{m}_a$ und $\tilde{m}_{a,t} = m_{a,t} - \bar{m}_a$ setzt:

$$\tilde{m}_{a,t} = \beta_a \gamma_t + \epsilon_{a,t}$$

Weiterhin kann man Nebenbedingung (22) nach γ_t umstellen. Dazu stellt man zuerst das Modell (20) nach β_a um und erhält

$$\beta_a = \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\gamma_t}$$

setzt man diese Charakterisierung für β_a nun in Nebenbedingung (22) ein, erhält man

$$\begin{aligned} 1 &= \sum_a \beta_a = \sum_a \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\gamma_t} \\ \iff \gamma_t &= \sum_a (m_{a,t} - \alpha_a) = \sum_a (m_{a,t} - \bar{m}_a) \end{aligned}$$

- Schätzung der Parameter: Wir haben bereits Charakterisierungen für α und für γ gesehen. Es bleibt sich Gedanken über β zu machen. Dazu wird der Ansatz aus [GK07] verwendet. Dort wird behauptet, dass β mittels der Maximum-Likelihood-Methode zu schätzen äquivalent zu folgendem Ansatz ist: Zuerst bestimmt man die Matrix

$$\tilde{M} = \begin{pmatrix} \tilde{m}_{0,1956} & \dots & \tilde{m}_{0,2017} \\ \vdots & \vdots & \vdots \\ \tilde{m}_{95,1956} & \dots & \tilde{m}_{95,2017} \end{pmatrix} \quad (23)$$

falls die Alter $\{0, \dots, 95\}$ und die Kalenderjahre $\{1956, \dots, 2017\}$ sind. 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 *beta* normieren und die Spalte aus B nehmen, die dem größten Singulärwert entspricht.

Nachdem man β geschätzt hat, kann man die Charakterisierungen für α und γ , die wir im vorhergehenden Abschnitt zum Modell gesehen haben, benutzen, um Schätzer zu generieren.

- Vorhersagen: Im Lee-Carter Modell geht man davon aus, dass α und β über die Zeit konstant bleiben. Das heißt, es genügt γ_t für t außerhalb der Beobachtungsperiode zu schätzen.

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

$$\hat{\gamma}_t = \hat{\gamma}_{t-1} + \nu + \xi_t \quad (24)$$

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 ist

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

hier sind E, S das Endalter und das Startalter der Sterbetafel.

- Bedeutung der Parameter: In diesem Punkt geht es darum, die Bedeutung der Parameter zu diskutieren. Gehen wir zuerst die Parameter einzeln durch:
 - Alpha: Der Alpha-Parameter ist auf Grund der Nebenbedingungen der Mittelwert der Sterblichkeiten über die Kalenderjahre. Der folgende Plot 10 zeigt die Werte von Alpha für das oben genannte Beispiel. Man kann diesen Parameter als die Sterblichkeit ohne Trends verstehen.

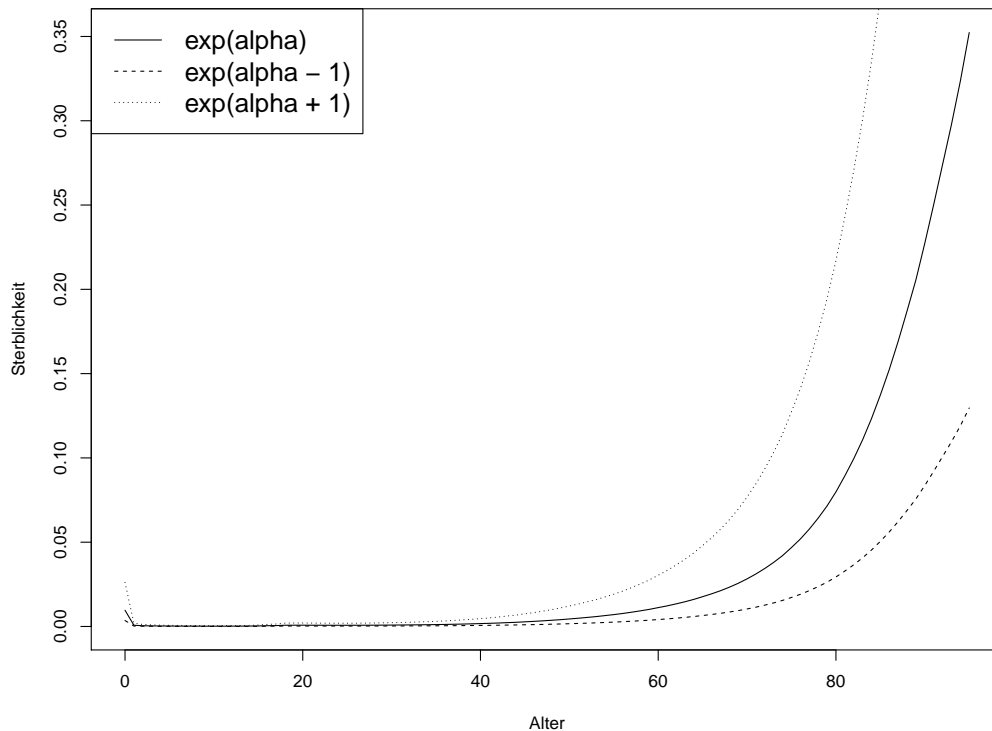


Abbildung 10: Lee-Carter Schätzung für $\exp(\alpha)$ für Westdeutsche Periodendaten 1956 bis 2017

Betrachtet man die Graphik 10, 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: Wir betrachten Periodendaten, das heißt, wir betrachten nur ein Kalenderjahr und in diesem Kalenderjahr die Verteilung von Toden und wir sehen, dass von den älteren Personen relativ mehr sterben als von den jungen Personen.

Außerdem veranschaulicht die Graphik 10, dass ein höherer Wert von Alpha zu höheren Sterbewahrscheinlichkeiten führt und ein niedrigerer Wert zu niedrigeren. Dies wird bei der Diskussion der Sterblichkeitsverbesserung wichtig.

- Beta: Der Beta-Parameter gibt an wie stark der Einfluss der Sterblichkeitsverbesserung ist, indem er den Gamma-Parameter Gewichtet. Die Gewichtungen, die man bei den Westdeutschen Periodendaten bekommt ist wie in Graphik 11 gezeigt.

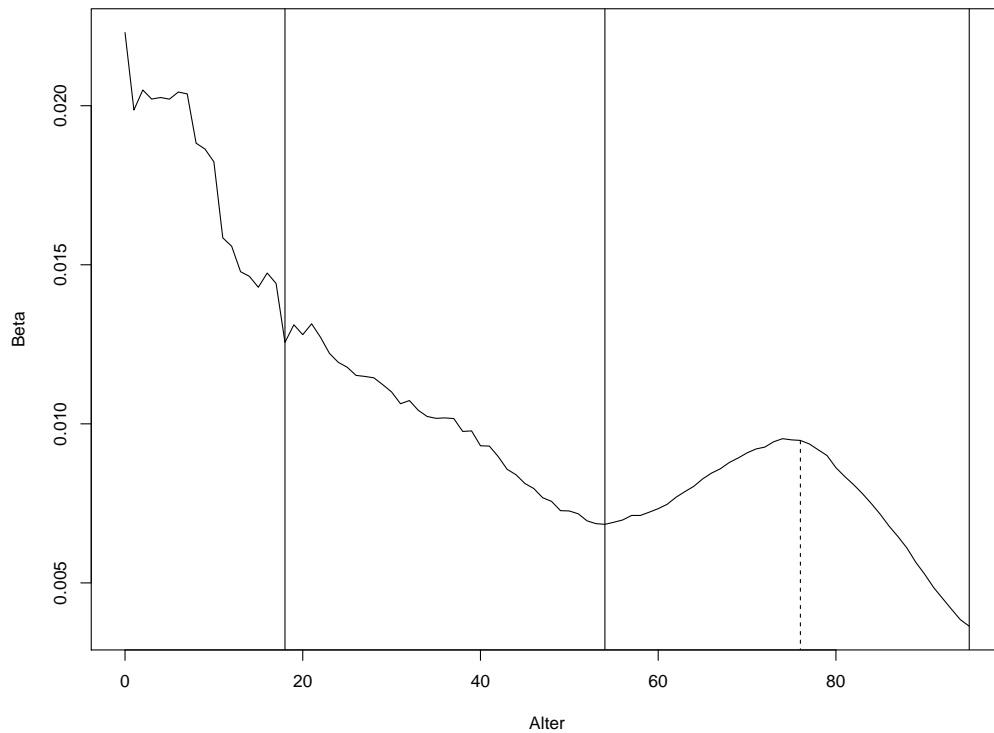


Abbildung 11: Lee-Carter Schätzung für Beta für Westdeutsche Periodendaten 1956 bis 2017. Als optische Hilfe sind 4 Bereiche verschiedener Sterblichkeitsverbesserung eingetragen.

Man sieht, dass der Einfluss der Sterblichkeitsverbesserung zuerst stark ist sich dann abschwächt, dann wieder steigt und dann wieder absinkt. Um den Einfluss auf die Sterblichkeitsverbesserung zu verstehen, muss man Beta und Gamma gemeinsam betrachten. Dies wird später gemacht.

- Gamma: Der Gamma-Parameter gibt die Verbesserung beziehungsweise die Verschlechterung der Sterblichkeit an. Beides ist grundsätzlich möglich. In der folgenden Graphik 12 sieht man wie sich γ_t im Laufe der Jahre entwickelt hat.

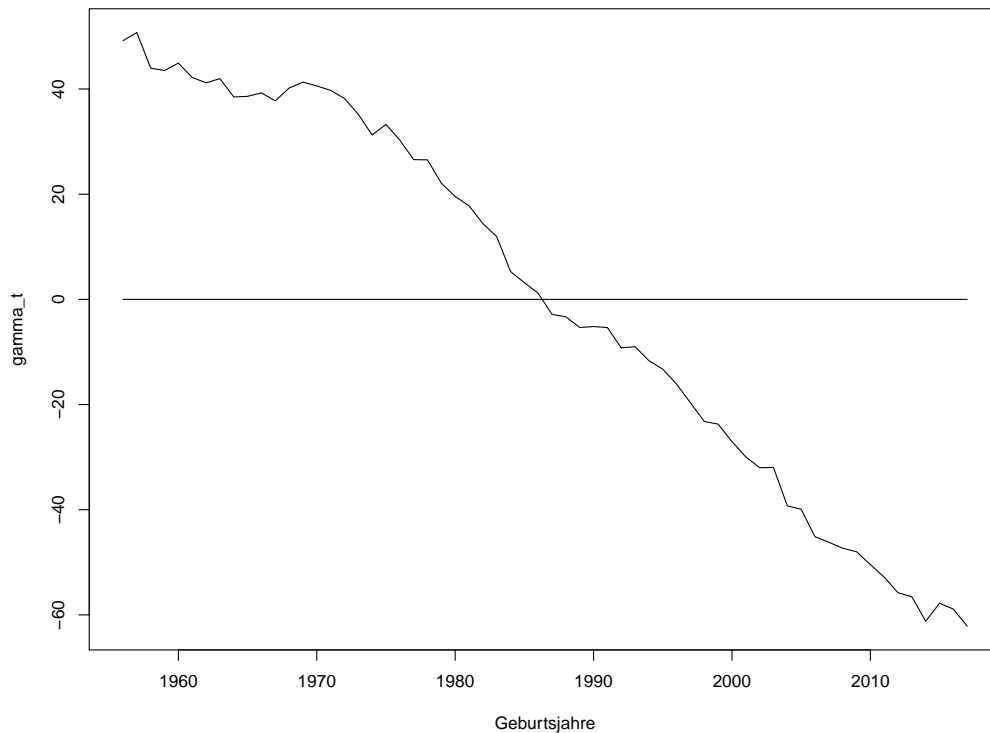


Abbildung 12: Lee-Carter Schätzung für Gamma für Westdeutsche Periodendaten 1956 bis 2017. Zur optischen Hilfe ist eine Linie auf Höhe der Null geplottet.

Man sieht, dass die Sterblichkeitsverbesserung ungefähr linear verläuft und von einem Wert von plus 60 zu einem Wert von -60 läuft. Da der Beta-Parameter immer positiv ist, gibt der Gamma-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 Neugeborenen und alten Personen beeinflusst.

Hier haben wir allerdings eine sehr starke Verbesserung bei unter 19 jährigen dann eine etwas langsermere Verbesserung bis zum Alter 54, dann eine Buckel und ab 98 keine Bestimmte Struktur mehr. Die Sterblichkeit der unter 19 jährigen lässt sich gut durch einen Rückgang der Kindersterblichkeit erklären. So sank laut Wikipedia "die Kindersterblichkeit weltweit [...] von 18,2 % im Jahr 1960 auf 4,3 % im Jahr 2015". Auch der letzte Abschnitt ab 98 lässt sich gut erklären: Für diese Personengruppe sind nicht genug Daten verfügbar, da nur wenige Personen solch ein Alter erreichen. Der Buckel zwischen 54 und 98 lässt sich damit erklären, dass an 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 sind 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 (Arbeitsunfällen) 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 ein Sterblichkeitsverschlechterung ($\gamma_t < 0$ beziehungsweise $\gamma_t > 0$).

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

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

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

Im folgenden Plot 13 sieht man die Werte der Produkte der Beta-Parameter und der Gamma-Parameter. Dass heißt, man sieht die folgende Matrix (25):

$$\begin{pmatrix} \beta_0 \cdot \gamma_S & \dots & \beta_{95} \cdot \gamma_S \\ \vdots & \vdots & \vdots \\ \beta_0 \cdot \gamma_E & \dots & \beta_{95} \cdot \gamma_E \end{pmatrix} \quad (25)$$

Hier der Plot:

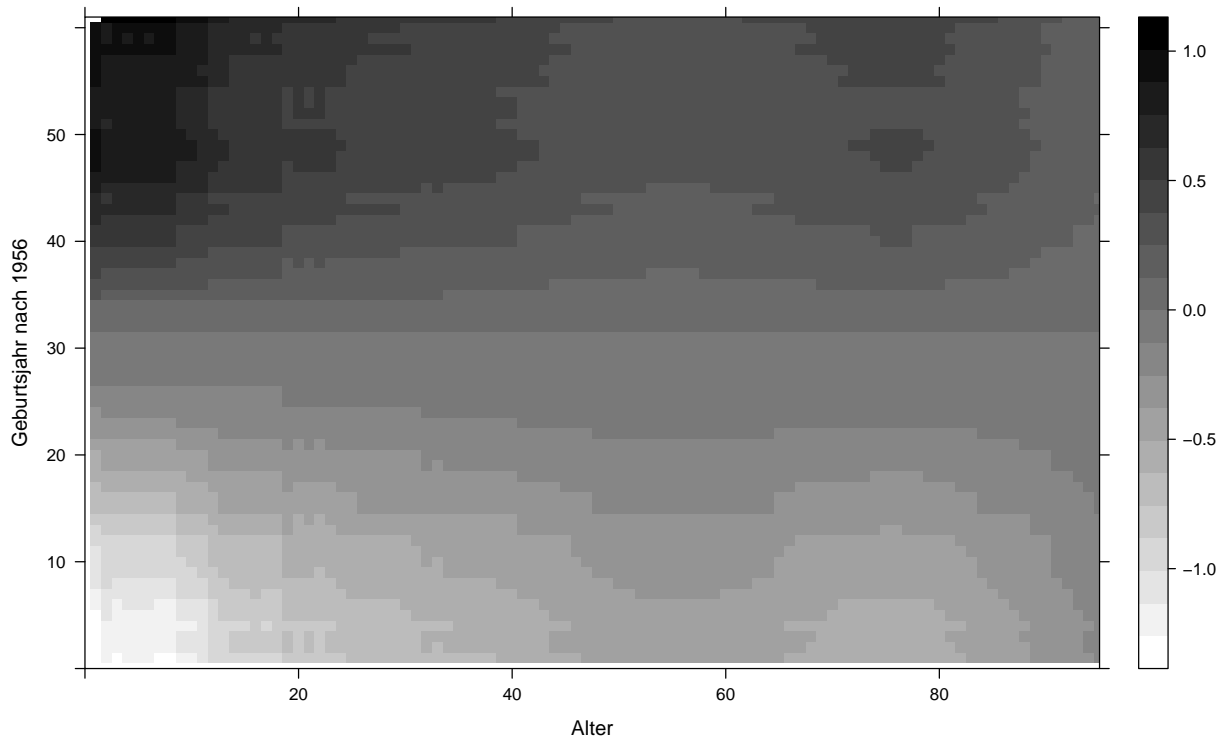


Abbildung 13: Heatmap für (γ_t, β_a)

Auf den ersten Blick ist der Plot 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 Geburtsjahre einen positiven Buckel und bei den selben Altern in der oberen Hälfte einen negativen Buckel. Die rechte Hälfte ab einem Alter von 90 ist wenig strukturiert und der Bereich in der Mitte enthält auch wenig Struktur. Ist dies nun richtig? Oder ist diese Matrix voll Schätzern falsch? Die Antwort liefert Plot (14). Hier sieht man die erste und den letzte Zeile aus der Matrix (25):

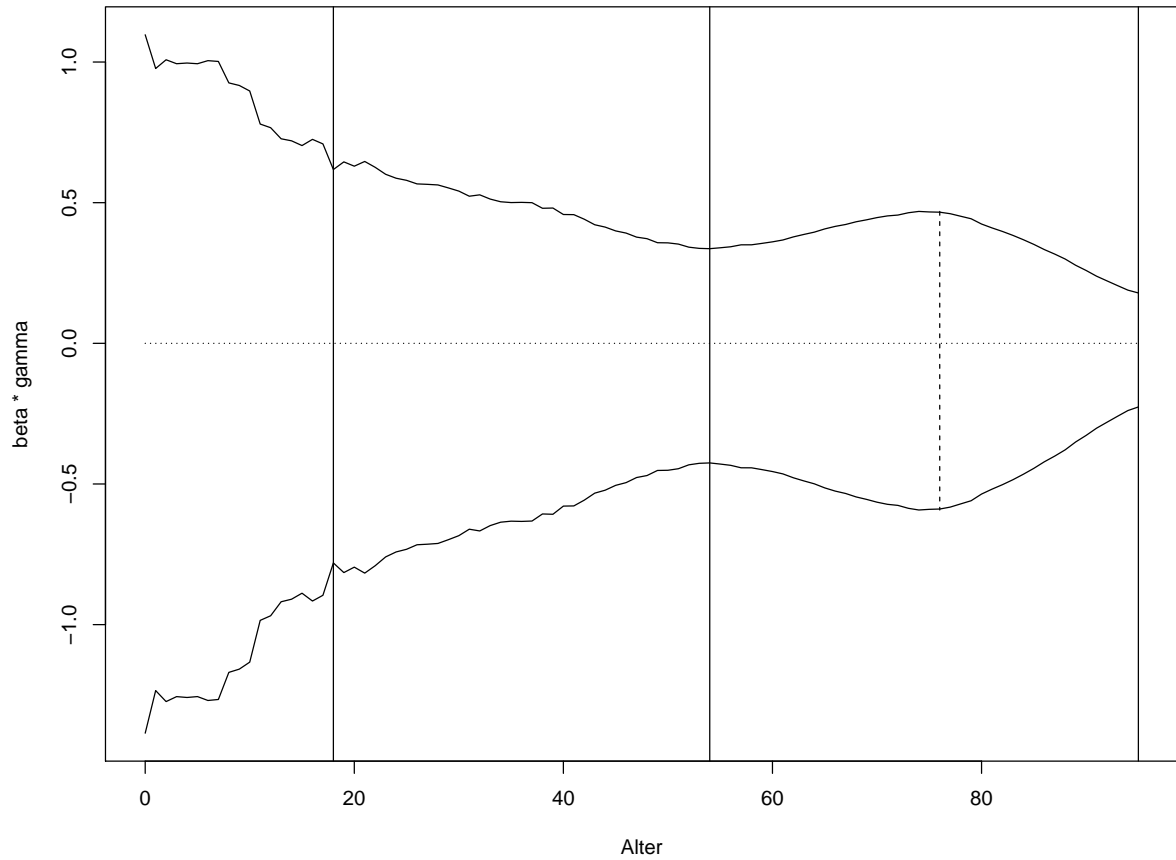


Abbildung 14: Obere Linie ist $\gamma_t \cdot \beta_a$ mit $t = 1956$ und die untere Linie ist $\gamma_t \cdot \beta_a$ mit $t = 1956$. Außerdem sind die Hilfslinie wie in Graphik (11) mit eingetragen

Dieser Plot zeigt also $\gamma_t \cdot \beta_a$ mit $t \in \{1956, 2017\}$ und außerdem die Hilfslinie wie in Graphik (11). Man sieht, dass dieser Plot in der Tat genau die Struktur von Graphik (13) wiedergibt. Die generierten Schätzer scheine also in der Tat sinnvoll.

3.6 Verbesserung des Einfachen Modells

Die folgende Graphik zeigt die Sterblichkeit pro Alter und pro Geburtsjahrgang:

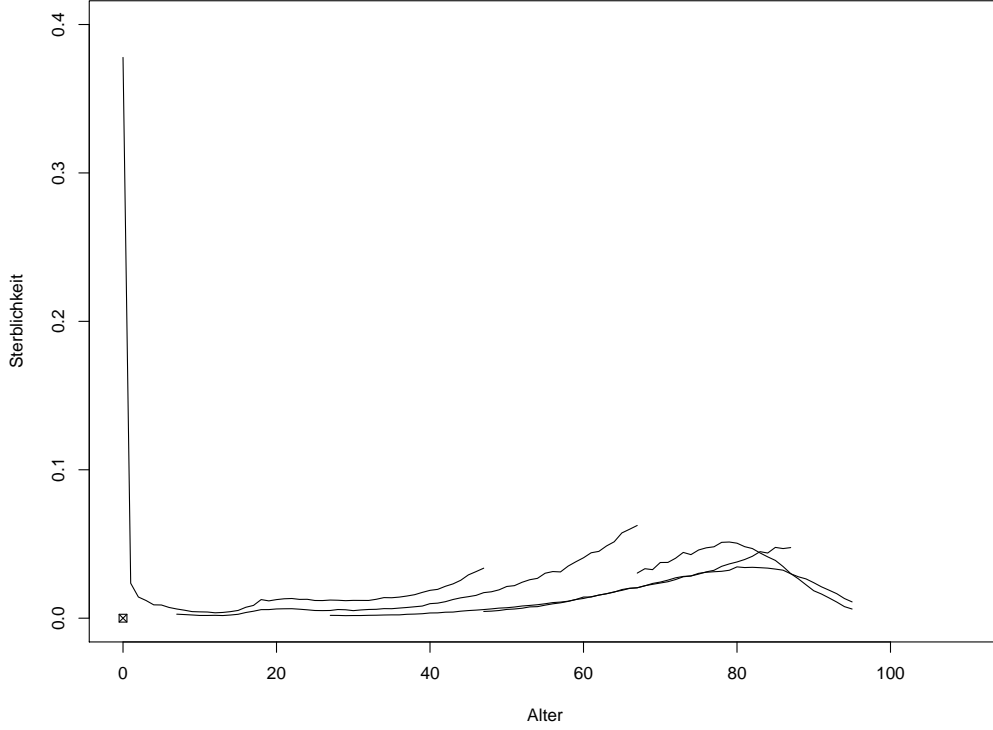


Abbildung 15: abschnittsweises-Dichteschätzen

In diesem Abschnitt werden Ansätze vorgestellt, das einfache Modell zu verbessern. Bisher sind wir von dem Modell

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

ausgegangen. Dieses Modell beachtet in keiner Weise das Geburtsjahr. Das heißt, Änderungen in der Sterblichkeit werden nicht berücksichtigt. Wir haben bereits den Random Walk with Drift (RWD) bei dem Lee-Carter Modell gesehen. Fügen wir also auch diesem Modell eine RWD Komponente hinzu, indem wir von dem Modell

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

mit

$$m_t = m_{t-1} + \nu + \xi_t \tag{26}$$

ausgehen. Dabei ist ν der unbekannte Driftparameter und ξ_t ein Fehler. Schätzt man die μ_t für den Datensatz der Generationssterbetafel erhält man den folgenden Plot:



Abbildung 16: Ein Plot der $\hat{\mu}_t$ gegen die Geburtsjahre zeigt.

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, dass 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.

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 Modell wenig Eigenschaften zeigen, da es sich nicht um ein statistisches Modell, sondern um ein numerisches Verfahren handelt. Dies wird in Abschnitt 4.3 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 4.4 ausgeführt. Die Verbesserung des einfachen Modells hat keine guten Ergebnisse in der Modellbildung geliefert und wird deshalb nur kurz in Abschnitt 4.5 besprochen. Am Ende des Kapitels werden in Abschnitt 4.6 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 1 (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 fast überall besitzt:

$$P_{\theta_1} = P_{\theta_2} \implies \theta_1 = \theta_2 \quad (27)$$

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 $\mathbf{X} = (X_1, \dots, X_n)$ 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 in θ Richtung existieren, sind stetig für alle $x \in \chi$ und können unter das Integral $\int f_\theta(x)dx$ geschoben werden.
4. Es existiert eine messbare, 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 $\theta : |\theta - \theta_0| < \epsilon$.
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 Wurzeln von $n^{-1} \sum_{i=1}^n l'_\theta(X_i) = 0$ mit $\sqrt{n}(\hat{\theta} - \theta) \rightarrow^D \mathcal{N}(0_k, i(\theta_0)^{-1})$.

Beweis. Für den vollständigen Beweis sehen Fergusons Buch *A course in large sample theory*, [Fer02], Theorem 18. \square

Man beachte, dass das Theorem von Cramér nur aussagt, dass ein kritischer Wert der Likelihoodfunktion stark konsistent und asymptotisch normalverteilt ist. Allerdings muss man sicherstellen, dass man auch das richtige Maximum benutzen, beziehungsweise dass es nur ein Maximum gibt.

Das größere Problem ist das Theorem überhaupt erst einmal zu zeigen. Die Eigenschaften 1 kann man durch eine passende Definition von Θ gewährleisten. Eigenschaft 2 folgt für die betrachteten Modell 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. Sei die Funktion $\theta \mapsto \int h(x)e^{\theta^T t(x)} d\mu(x)$ 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]. \square

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 ein 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:

Defintion 2 (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 bezüglich μ der folgenden Form besitzen:

$$f(y, \theta) = h(y)A(\theta)\exp(\eta(\theta)^t T(y)) \quad (28)$$

Dabei sind h und T messbare Funktionen und $A, \eta : \mathbb{R}^k \supset \Theta \mapsto \mathbb{R}$ 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:

$$\begin{aligned} f_{m,s}(y) &= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - y\right)^2}{2\sigma_\epsilon^2}\right) \\ &=: h_1(y) \exp\left(-\frac{\left(\frac{1}{s} h\left(\frac{m-x}{s}\right)\right)^2 - 2\frac{1}{s} h\left(\frac{m-x}{s}\right) y + y^2}{2\sigma_\epsilon^2}\right) \\ &= h_1(y) \exp\left(-\frac{y^2}{2\sigma_\epsilon^2}\right) \cdot \exp\left(\frac{\frac{1}{s} h\left(\frac{m-x}{s}\right) y}{\sigma_\epsilon^2}\right) \cdot \exp\left(-\frac{\left(\frac{1}{s} h\left(\frac{m-x}{s}\right)\right)^2}{2\sigma_\epsilon^2}\right) \\ &=: h(y) \cdot \exp(\eta(\theta)T(y)) \cdot A(\theta) \end{aligned}$$

Man sieht, dass es sich in der Tat um eine Exponentialfamilie handelt. Das Problem ist, dass $\eta(\theta) \in \mathbb{R}$, während $\theta \in \mathbb{R}^2$. Das impliziert, dass η auch keine bijektive Abbildung sein kann, da η von \mathbb{R}^2 nach \mathbb{R} geht. Untersuchen wir als nächstes die beiden Teile des Lee-Carter Modells. Wir beginnen mit dem Schätzer für β :

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

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 :

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

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. Ausgehen 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 auch nicht einfach funktioniert. 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. Präziser nennt man einen Schätzer $\hat{\theta}$ Zero-Schätzer, wenn er die folgende Gleichung löst:

$$\Psi_n(\theta) = \frac{1}{n} \sum_i^n \psi_\theta(y_i) = 0 \quad (29)$$

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 [LR05] in *Testing statistical hypotheses* in Theorem 12.2.2 (S.498). 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 ein Variante von Lemma 5.10 in Verbindung mit Lemma 2.2 (S.14) aus *Handbook of Econometrics* [EM86] 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. Es gelten die folgenden Eigenschaften:

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-Likelihood konvergieren in Wahrscheinlichkeit gegen Ihren Erwartungswert.
5. Die Ableitungen der Log-Likelihood 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 diffbar in θ und in x ist.
9. Die Fisher-Informationsmatrix $i(\theta)$ existiert, ist nicht singulär und ist 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:

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

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 $\epsilon \sim N(0, \sigma_\epsilon)$ unabhängig und identisch (u.i.v.) verteilt ist. Zuerst wird die Likelihood-Funktion aufgestellt und die Maximum-Likelihood Schätzer bestimmt. Danach werden die Voraussetzungen für das Theorem 3 gezeigt. Das Modell ist durch die folgende Gleichung gegeben:

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

Dieses Modell hat die folgende Dichte:

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

und die folgende log-Likelihoodfunktion:

$$\begin{aligned} l_y(m, s) &= \sum_{i=1}^n \ln(f_{y_i}(m, s)) = \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2}\right)\right) \\ &= \sum_{i=1}^n \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2} \right\} \end{aligned}$$

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:

$$\begin{aligned}
& \frac{d}{dm} l_y(m, s) \\
&= \frac{d}{dm} \sum_{i=1}^n \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right)^2}{2\sigma_\epsilon^2} \right\} \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n \frac{d}{dm} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right)^2 \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n 2 \cdot \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \frac{d}{dm} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \\
&= -\frac{1}{2\sigma_\epsilon^2} \cdot \frac{2}{\sqrt{2\pi s^2}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \frac{d}{dm} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi s^2}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{d}{dm} \left(-\frac{(m-x_i)^2}{2s^2} \right) \\
&= \frac{1}{\sigma_\epsilon^2 2s^2 \sqrt{2\pi s^2}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot 2(m-x_i) \\
&= \frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i)
\end{aligned}$$

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. Dieser Plot ist (17).

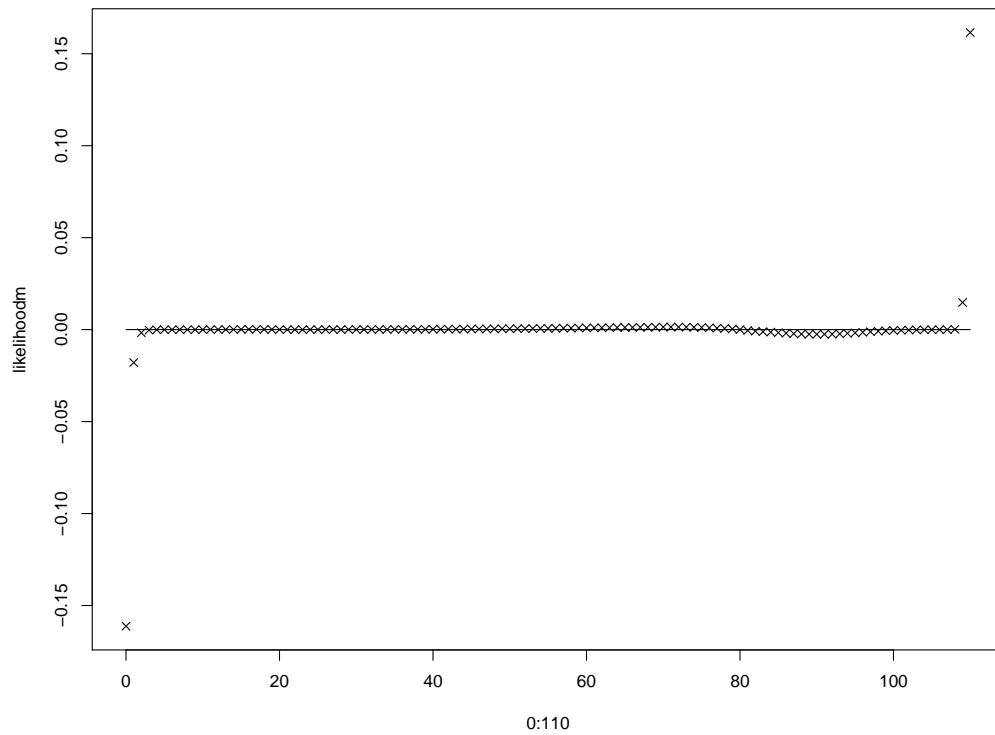


Abbildung 17: Die erste Ableitung der Log-Likelihoodfunktion in m Richtung. Auf der x -Achse mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht, dass die Werte an den Rändern sehr groß sind. Deshalb ist in Graphik (18) dieselbe Funktion noch einmal geplottet, diesmal allerdings ohne die Ränder:

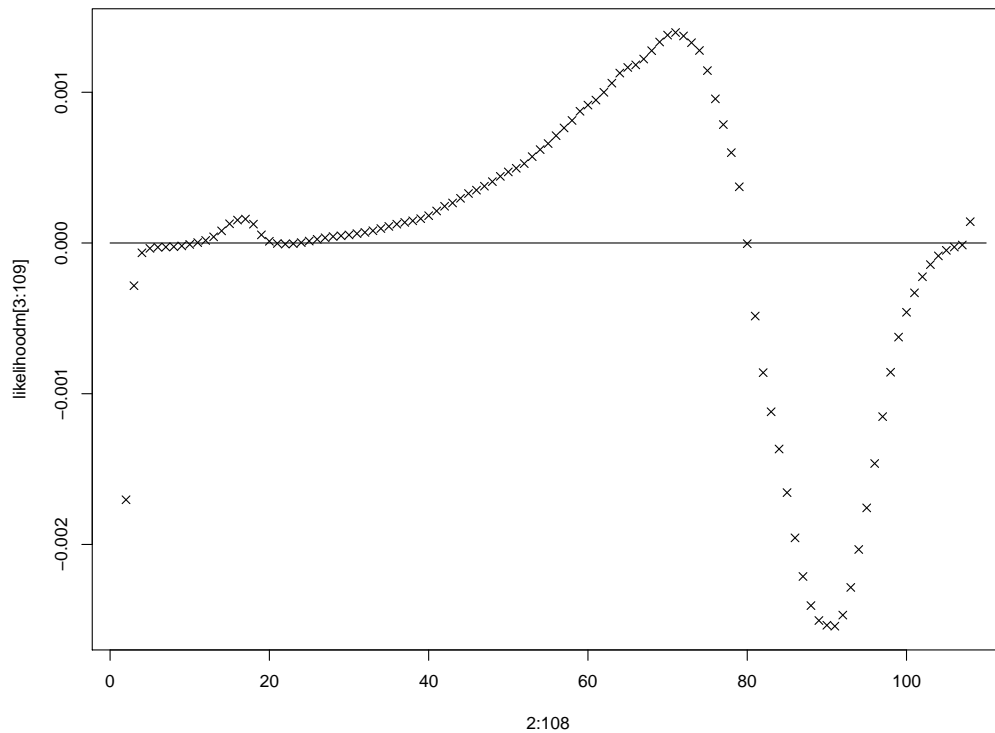


Abbildung 18: Die erste Ableitung der Log-Likelihoodfunktion in m Richtung ohne die Randwerte. Auf der x -Achse sind wieder mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht vier Mögliche Nullstellen. Die Nullstelle bei 80 scheint am sinnvollsten, wenn man sich die Daten (8) anschaut. Berechnen wir jetzt die erste Ableitung in s^2 Richtung:

$$\begin{aligned}
& \frac{d}{ds^2} l_y(m, s) \\
&= \frac{d}{ds^2} \sum_{i=1}^n -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i)^2}{2\sigma_\epsilon^2} \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n \frac{d}{ds^2} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right)^2 \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n 2 \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \frac{d}{ds^2} \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \right\} \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ \left(\frac{d}{ds^2} s^{-1} \right) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + s^{-1} \frac{d}{ds^2} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ (-s^{-2}) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + s^{-1} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \frac{d}{ds^2} \left(-\frac{(m-x_i)^2}{2s^2}\right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ (-s^{-2}) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + s^{-1} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(-\frac{(m-x_i)^2}{2}\right) \cdot (-2s^{-3}) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ (-s^{-2}) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i)^2 \cdot (s^{-4}) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^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 dem Plot (18). Man erhält den Plot (20):

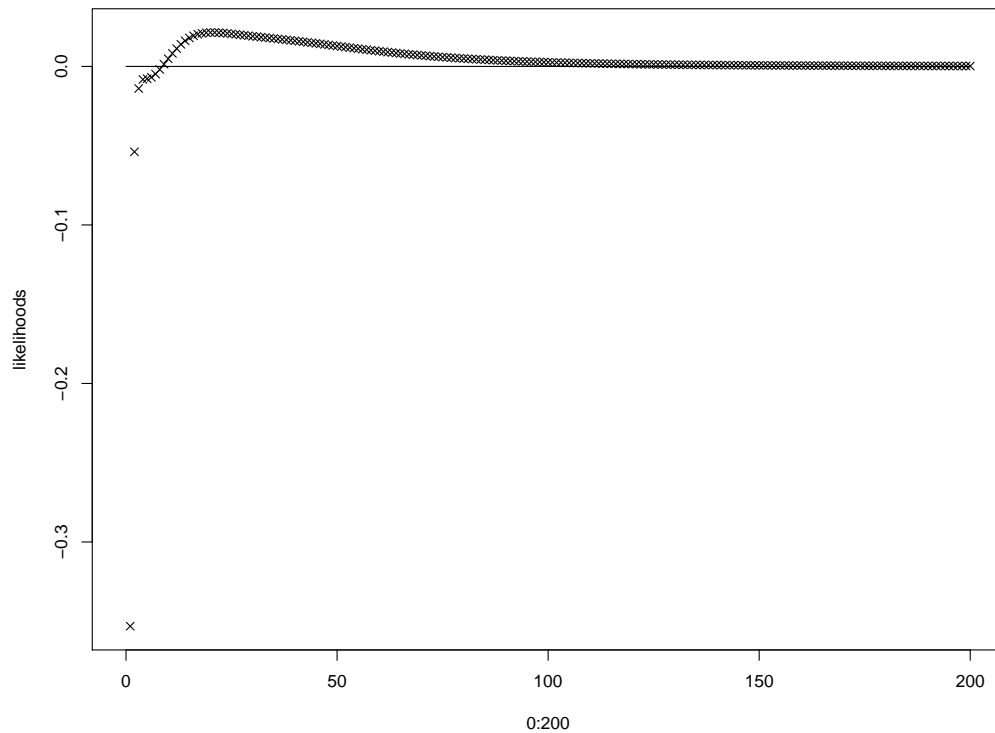


Abbildung 19: Die erste Ableitung der Log-Likelihoodfunktion in s^2 Richtung. Auf der x -Achse mögliche Werte für \hat{s}^2 und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht, dass die ersten paar s^2 Werte stark von den späteren in Größe abweichen. Plotten wir also als nächstes die selben s^2 Werte ohne die ersten drei Werte. Dabei erhält man plot (20):

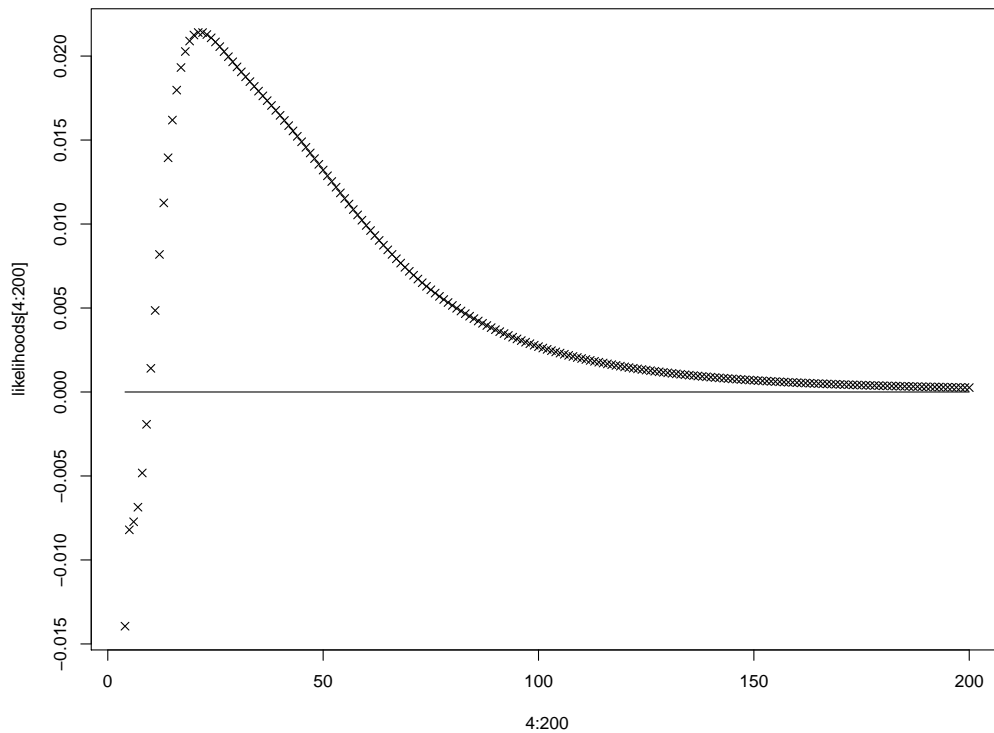


Abbildung 20: Die erste Ableitung der Log-Likelihoodfunktion in s^2 Richtung ohne die ersten drei Werte. Auf der x -Achse sind wieder mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht schon etwas mehr, aber der Maßstab ist noch zu groß. Man beachte also, 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 Graphik (21) in der nur der Bereich von 4 bis 10 in 0.1 Inkrementen zu sehen ist:

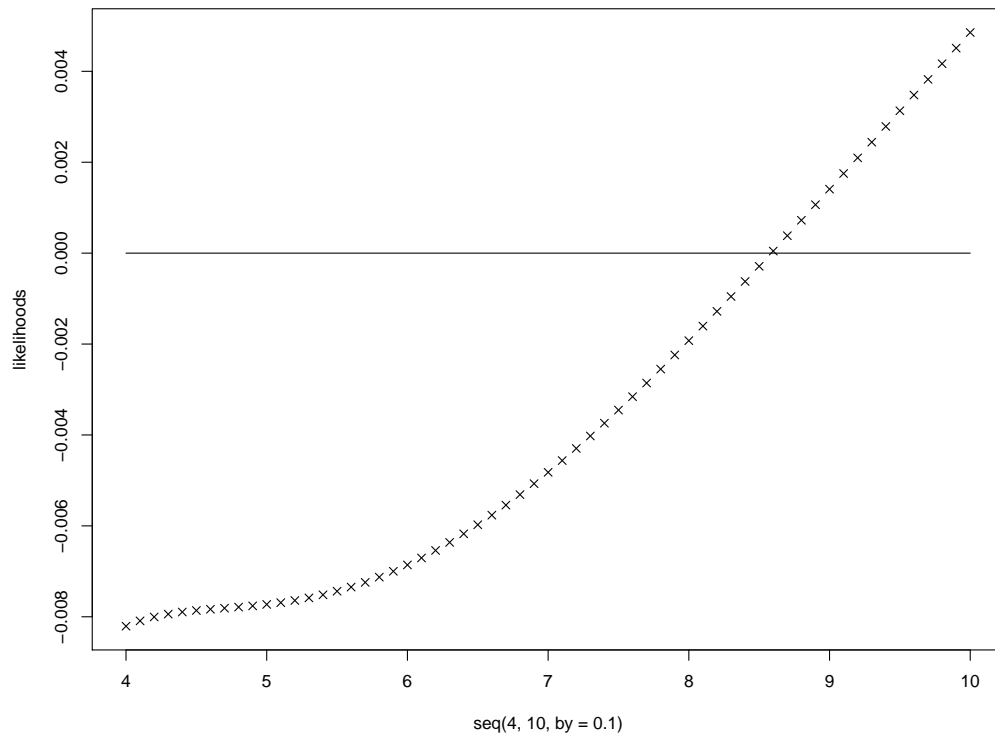


Abbildung 21: Die erste Ableitung der Log-Likelihoodfunktion in s Richtung im Bereich 4 bis 10 in 0.1 Inkrementen. Auf der x -Achse sind wieder mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht in Graphik (21) genau eine Nullstelle bei ungefähr 8,7. Berechnen wir als nächstes die zweiten Ableitungen der Log-Likelihood. Beginnen wir mit der zweiten Ableitung in m Richtung:

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

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

$$\mathbb{E} \left(\frac{d^2}{dm^2} l_{x,y,s,\sigma_\epsilon} \right) = -\frac{1}{2\pi s^4 \sigma_\epsilon^2} \sum_{i=1}^n a^2 b \quad (30)$$

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

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

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

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

Man kann den Erwartungswert direkt ablesen:

$$\mathbb{E} \left(\frac{d}{ds^2} \frac{d}{dm} l_{x,y,\sigma_\epsilon}(m, s) \right) = -\frac{1}{2\pi s^2 \sigma_\epsilon^2} \sum_{i=1}^n d_s a \sqrt{b} \quad (31)$$

Berechnen wir als nächste die zweifache Ableitung in s Richtung:

$$\begin{aligned}
& \frac{d^2}{ds^4} l_{x,y,m,\sigma_\epsilon}(s) \\
= & -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \frac{d}{ds^2} \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right] \\
= & -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \dots \\
& \dots \left[\frac{d}{ds^2} \left(\frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right\} + \dots \right. \\
& \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \frac{d}{ds^2} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) + \dots \\
& \dots \left. \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{d}{ds^2} \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right] \\
= & -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \dots \\
& \left[\left\{ -\frac{1}{s^2 \sqrt{2\pi}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + \frac{1}{\sqrt{2\pi s^2}} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^3} \right) \right\} \dots \right. \\
& \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right\} + \dots \\
& \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{(m-x_i)^2}{s^3} \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) + \dots \\
& \dots \left. \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{2}{s^3} - \frac{4(m-x_i)^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:

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

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:

$$\mathbb{E} \left(\frac{d^2}{ds^4} l_{x,y,m,\sigma_\epsilon}(s) \right) = -\frac{1}{\sigma_\epsilon^2 2\pi} \sum_{i=1}^n d_s^2 \quad (32)$$

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 (11) als eine Menge normalverteilter Maße auf die aus dem folgenden Modell generiert werden:

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

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:

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

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:

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

3. Der Parameterraum $\Theta \subset \mathbb{R}^n$ ist ein möglicherweise offenes Rechteck: Sei $\Theta = (30, 85) \times (4, 10)$. Diese Menge ist so gewählt, dass man genau ein offenes Rechteck in \mathbb{R}^2 erhält.
4. Die Ableitungen der Log-Likelihood 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:

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

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

$$l_y(m, s) = \sum_{i=1}^n \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right)^2}{2\sigma_\epsilon^2} \right\}$$

Da l_y eine Verknüpfung stetig diffbarer Funktionen ist, ist die Ableitung der Log-Likelihood 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 Graphiken der Log-Likelihood. Dieses Argument, Begründung durch zeichnen, verwendet auch Van der Vaart.
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 diffbar in θ und in x ist: Die Dichte ist gegeben durch:

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

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

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

$$\begin{aligned} i(\theta) &= -\mathbb{E} \begin{pmatrix} \frac{d^2}{dm^2} l_y(\theta) & \frac{d}{dm} \frac{d}{ds} l_y(\theta) \\ \frac{d}{ds} \frac{d}{dm} l_y(\theta) & \frac{d^2}{ds^2} l_y(\theta) \end{pmatrix} \\ &= \begin{pmatrix} -\mathbb{E} \frac{d^2}{dm^2} l_y(\theta) & -\mathbb{E} \frac{d}{dm} \frac{d}{ds} l_y(\theta) \\ -\mathbb{E} \frac{d}{ds} \frac{d}{dm} l_y(\theta) & -\mathbb{E} \frac{d^2}{ds^2} l_y(\theta) \end{pmatrix} \end{aligned}$$

Alle Bestandteile wurden in Gleichung (30), (31) und (32) bereits ausgerechnet. Man erhält als Determinante:

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

Die Cauchy-Schwarz Ungleichung liefert:

$$\sum_{i=1}^n a^2 b \cdot \sum_{i=1}^n d_s^2 \geq \left(\sum_{i=1}^n d_s a \sqrt{b} \right)^2$$

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 x, y anwendet. Man erhält, dass die folgende Gleichung für alle $x_i = \{0, \dots, 95\}$ gelten muss.

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

Diese Gleichung kann nicht für alle x_i gelten, da man aus den ersten beiden Einträgen der Vektoren, dass heißt $x_i = 0, 1$ den Wert $m = 1$ erhält. Setzt man diesen für $x_i = 2$ 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_y gegeben durch:

$$l_y(m, s) = \sum_{i=1}^n \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2} \right\}$$

Man sieht, dass l_y 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-likelihood auffassen. Eine Komposition von Lipschitz Funktionen ist wieder Lipschitz mit dem Produkt der Lipschitzkonstanten als Lipschitzkonstante:

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

Weiterhin ist die Logarithmusfunktion auf einem beschränkten Intervall lipschitz. Eine Dichte nach Proposition 2.2 (S.118) in *Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions* [HR10] lipschitz, wenn sie beschränkt ist. Die Normalverteilung ist beschränkt, also hat man Lipschitz-Stetigkeit der log-likelihood.

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 Whittaker-Henderson Verfahrens

In diesem Abschnitt betrachten wir das Whittaker-Henderson Verfahren. Das Whittaker-Henderson Verfahren macht bei der Bestimmung seiner theoretischen Eigenschaften einige Probleme. Zum einen ist es nicht parametrisch und zum anderen handelt es sich nicht um einen Maximum-Likelihood Schätzer. Genaugenommen handelt es sich nicht einmal um einen Schätzer, da kein statistisches Modell angenommen. Es handelt sich nur um ein numerisches Verfahren. Das heißt, dass auf alle Fälle nicht das Theorem von Cramér angewandt werden kann. Um die Problematik zu veranschaulichen, rechnen wir Identifizierbarkeit nach. Dazu braucht man zuerst ein statistisches Modell. Gehen wir deshalb von dem folgenden Modell aus, welches mehr Annahmen hat als das bisher betrachtete Whittaker-Modell (14):

$$Q = \tilde{Q} + \epsilon \quad (33)$$

Dabei sind \tilde{Q} die wahren Werte von Q und ϵ zufällige Fehler. Setzt man wieder die Definition der Identifizierbarkeit an erhält man:

$$\begin{aligned} P_{\theta_1} = P_{\theta_2} &\iff \tilde{Q}_1 + \epsilon = \tilde{Q}_2 + \epsilon \\ &\iff \tilde{Q}_1 = \tilde{Q}_2 \\ &\iff \theta_1 = \theta_2 \end{aligned}$$

Damit ist das Modell identifizierbar, unabhängig davon, welche Annahmen man an ϵ macht. Will man für das nichtparametrische Modell (14) Identifizierbarkeit zeigen, braucht man etwas andere Definitionen. Eine Möglichkeit ist eine Definition ähnlich der in "Conditions for Identification in Nonparametric and Parametric Models" (1988) von Charles S. Roehrig, [Roe88] zu verwenden:

Definition 3. Seien X und U Vektoren wobei X beobachtbar ist und U nicht. Sei weiterhin $Y = \pi(X, U)$ und Ψ die Verteilungsfunktion zu Y . Dann ist das Modell S identifizierbar, genau dann wenn es kein Modell S^* mit $\Psi = \Psi^*$ und $S \neq S^*$ gibt.

Wendet man diese Definition auf das Modell (14) an, erhält man $Y = \pi(X, U) = f(X) + \epsilon \sim N(f(X), \sigma_\epsilon^2)$. Hat man ein anderes Modell mit $Y \sim N(f(X), \sigma_\epsilon^2)$, da eine Zufallsvariable durch ihre Verteilung vollkommen charakterisiert ist.

Das heißt allein Identifizierbarkeit unter den geringen Annahmen des Whittaker-Modells zu zeigen ist bereits Problematik und deshalb werden keine weiteren Eigenschaften nachgerechnet.

4.4 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_a und ν getrennt.

4.4.1 Der Schätzer für β

Wir geben zuerst das Modell, in dem β geschätzt wird, an, berechnen den Maximum-Likelihood Schätzer und zeigen dann die fünf Eigenschaften.

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

Entsprechend erhalten wir die Dichte:

$$\Pi_{t=1}^n f(\tilde{m}_{a,t} | \beta_a, \sigma_{\epsilon_{a,t}}) = \Pi_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{a,t}}^2}} \exp\left(-\frac{(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t})^2}{2\sigma_{\epsilon_{a,t}}^2}\right)$$

Für die log-likelihood erhalten wir:

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

Wir erhalten als Ableitung in β_a Richtung

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

und erhalten als Nullstelle:

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

Das heißt, wir erhalten als Maximum-Likelihood-Schätzer:

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

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

$$\begin{aligned}
\sum_{a=1}^n \hat{\beta}_a &= \sum_{a=1}^n \left(\sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1} \\
&= \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1} \cdot \sum_{a=1}^n \left(\sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t \right) \\
&= \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1} \cdot \left(\sum_{t=1}^n \sum_{a=1}^n \tilde{m}_{a,t} \hat{\gamma}_t \right) \\
&= \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1} \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \\
&= 1
\end{aligned}$$

Berechnung wir als nächstes die zweite Ableitung:

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

Die zweite Ableitung ist immer negativ. Also haben wir in der Tat ein Maximum gefunden. Jetzt werden die fünf Eigenschaften 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 $\beta_a \in (c_1, c_2)$ 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 $\theta = (\alpha_a, \beta_a, \gamma_t, \nu)$. Wir setzen wieder die Definition an:

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

betrachten wir zuerst den ersten Summanden und setzen die Nebenbedingungen für $\alpha_a = \bar{m}_a = \alpha'_a$ und für $\gamma_t = \sum_a (m_{a,t} - \bar{m}_a)$ ein:

$$\begin{aligned}
\alpha_a + \beta_a \gamma_t + \epsilon = \alpha'_a + \beta'_a \gamma'_t + \epsilon &\iff \alpha_a + \beta_a \gamma_t = \alpha'_a + \beta'_a \gamma'_t \\
&\iff \bar{m}_a + \beta_a \sum_a (m_{a,t} - \bar{m}_a) = \bar{m}_a + \beta'_a \sum_a (m_{a,t} - \bar{m}_a) \\
&\iff \beta_a = \beta'_a
\end{aligned}$$

betrachten wir nun den zweiten Summenden

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

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

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

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

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

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

Man kann also in diesem Fall einfach $l''_\theta(x) = K(x)$ setzen, da die Summe über t endlich ist.

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

Die zweite Ableitung wurde als der folgender Ausdruck bestimmt:

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

Damit ist die Fisher-Informationsmatrix:

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

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

4.4.2 Schätzer für ν

Zeigen wir als nächstes das Cramér Theorem für den Schätzer für 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:

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

Man beachte, dass die Summe bei $t = S + 1$ beginnt, da für $t = S$ der random walk beginnt. Die Log-Likelihood ist gegeben durch:

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

Die erste Ableitung der Log-Likelihood ist gegeben durch:

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

Damit ist der Maximum-Likelihood Schätzer gegeben durch:

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

Dieser Ausdruck vereinfacht sich zu:

$$\begin{aligned}
\frac{\sum_{t=S+1}^E \hat{\gamma}_t - \sum_{t=S+1}^E \hat{\gamma}_{t-1}}{E-S} &= \frac{\hat{\gamma}_{S+1} + \hat{\gamma}_{S+2} + \dots + \hat{\gamma}_E - \hat{\gamma}_S - \hat{\gamma}_{S+1} - \hat{\gamma}_{S+2} - \dots - \hat{\gamma}_{E-1}}{E-S} \\
&= \frac{\hat{\gamma}_E - \hat{\gamma}_S}{E-S}
\end{aligned}$$

Man erhält als zweite Ableitung:

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

Es bleibt Eigenschaften drei, vier und fünf zu zeigen, da man Eigenschaft 1 durch passende Definition von Θ sicherstellen kann und Eigenschaft 2 über die Identifizierbarkeit bereits bei der Identifizierbarkeit von β mitbewiesen wurde. Siehe dazu Abschnitt 4.4.1.

- Die zweite Ableitung von f_θ existiert und ist stetig für alle $x \in \chi$ 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 die $K(x) < \infty$:

$$\begin{aligned}
\frac{d^2}{d\nu^2} l_{\hat{\gamma}_t, \hat{\gamma}_{t-1}, \sigma_{\xi_t}^2}(\nu) &= \frac{\hat{\gamma}_E - \hat{\gamma}_S}{2\sigma_{\xi_t}^2} \\
&= \frac{1}{2\sigma_{\xi_t}^2} \left(\sum_a (m_{a,E} - \bar{m}_a) - \sum_a (m_{a,S} - \bar{m}_a) \right) \\
&= \frac{1}{2\sigma_{\xi_t}^2} \left(\sum_a m_{a,E} - \sum_a m_{a,S} \right) \\
&= \frac{1}{2\sigma_{\xi_t}^2} \left(\sum_a \beta_a \gamma_E - \sum_a \beta_a \gamma_S \right) \\
&= \frac{1}{2\sigma_{\xi_t}^2} (\gamma_E - \gamma_S) < \infty
\end{aligned}$$

In der letzten Gleichung wurde $\sum_a \beta_a = 1$ und $\gamma_E, \gamma_S < \infty$ nach Annahme benutzt.

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

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

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

4.5 Eigenschaften des einfachen Trendmodells

Wir haben in Abschnitt 3.6 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 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

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

und 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

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

und somit Identifizierbarkeit. Konsistenz und asymptotische Verteilung folgend ebenfalls einfach, indem man die gleichen Rechnungen wie in den entsprechenden Abschnitten durchführt.

4.6 Qualitätsvergleich durch Simulation

In diesem Unterkapitel werden Simulationen vorgestellt, um die verwendeten Methoden zu vergleichen. Dazu werden zwei Arten von Simulationen durchgeführt: In der ersten Art Simulation werden die Parameter des einfachen Modells und des Lee-Carter Modells in ihrem Modell geschätzt. In der zweiten Art wird der historische Datensatz 1956 Westdeutschland halbiert und die frühere Hälfte benutzt um die spätere Hälfte vorher zu sagen. Dabei wird für das einfache Modell und das Whittaker Modell der Datensatz als Generationstafel benutzt und für das Lee-Carter Modell dieselben Daten im Format einer Periodentafel. Die Größe des Fehlers wird jeweils in der L_2 Norm bestimmt. Das heißt, es wird $\sum_i (\hat{\theta}_i - \theta_i)^2$ für die Parameter und $\sum_i (\hat{q}_i - q_i)^2$ für den Datensatz betrachtet. Beginnen wir damit, die Parameter des einfachen Modells zu diskutieren. Es ist $X = (0, \dots, 95)$ Es werden die Parameter $\mu = 77$, $\sigma = 12$ und $\sigma_\epsilon = \sqrt{0.000006}$ 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 $\hat{\sigma}_\epsilon^2 = \frac{1}{96-2} \sum_{i=1}^{96} (\hat{Y}_i - Y_i)^2$ geschätzt. Man erhält also das Modell 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 $m = 100$ Datensätze erzeugt. Diese Werten in der Matrix *simple_data* nach dem Format Alter, Daten ohne Fehler, Datensatz 1, ..., Datensatz 100 abgelegt. Dieses Format ist in Tabelle (2) veranschaulicht.

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

Tabelle 2: Der Datensatz *simple_data*.

In Graphik (22) sieht man ersten Datensatz. Die durchgezogene Linie zeigt $f(x)$ und die Kreuze sind der erste Datensatz. Das heißt die Kreuze sind $f(x) + \epsilon_{x,1}$. Man beachte, dass negative Sterbewahrscheinlichkeiten auf Null gesetzt werden. Dadurch wird die Annahme des Modells geringfügig verletzt, allerdings hat man sonst, gerade bei dem Lee-Carter Modell das Problem, dass man den Logarithmus einer negativen Zahl verwenden müsste. Deshalb sind die Fehler nicht gleichmäßig um den wahren Datensatz verteilt.

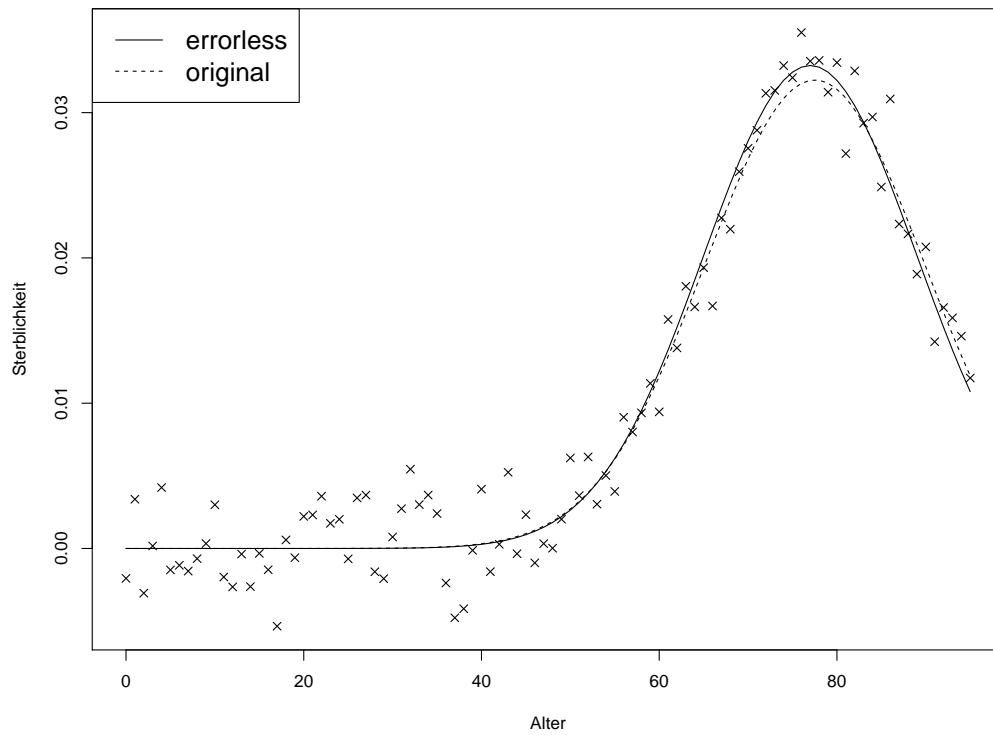


Abbildung 22: Ein Beispiel der Simulationsdaten für das einfache Modell

Als nächstes sieht man in Graphik (23) den Abstand zwischen (m, s) und (\hat{m}, \hat{s}) pro Simulation $i \in (1, \dots, 100)$. Die durchgezogene Linie zeigt den Mittelwert.

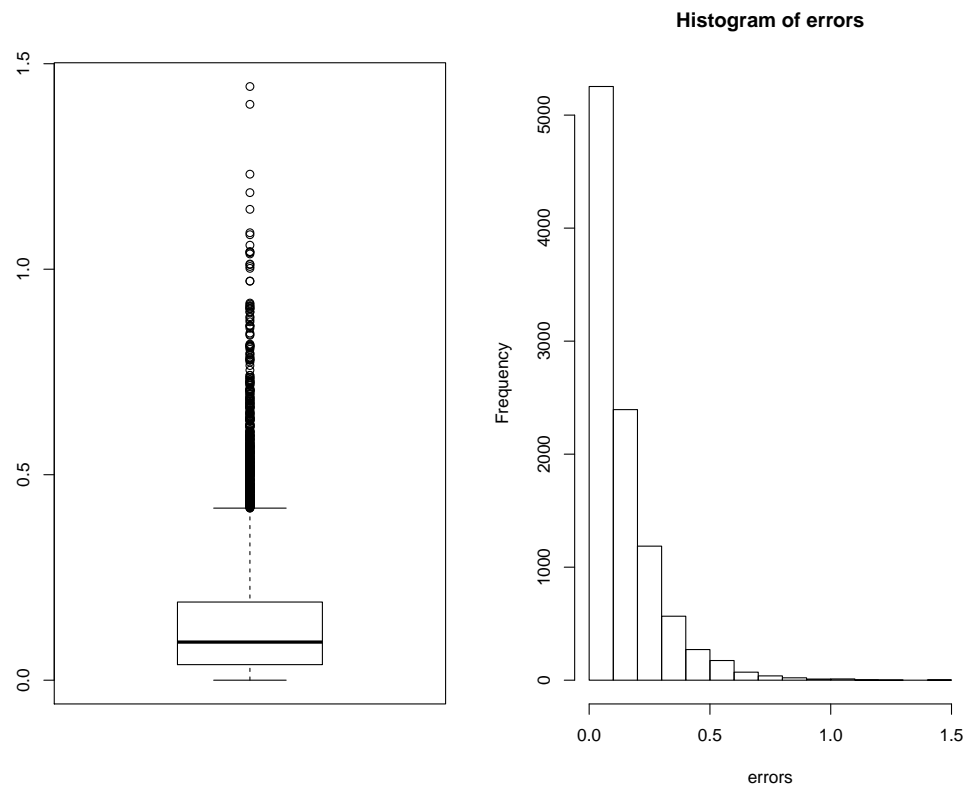


Abbildung 23: Fehler bei der Parameterschätzung beim einfachen Modell

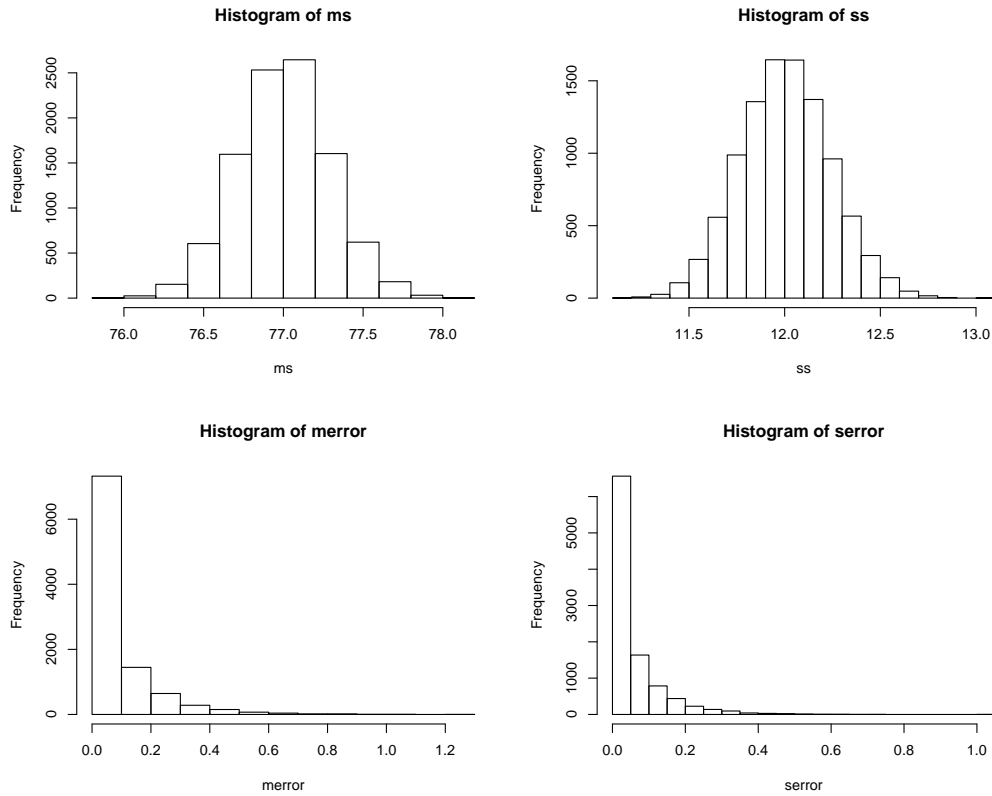


Abbildung 24: TODO

Betrachten wir nun das Lee-Carter Modell und schätzen wieder den Abstand zwischen den Parametern $(\alpha_a, \beta_a, \gamma_t, \nu)$, $a \in (0, 95)$, $t \in (0, 40)$ und ihren Schätzern. Da wir in diesem Modell sehr viele Parameter haben, sind die Werte der Parameter in der Graphik (24) zusammengefasst. Man sieht oben links die α Parameter, oben rechts die β Parameter und unten links die γ Parameter. 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 $\sigma_\xi = 2.344917$. Diese Parameter werden in der Funktion *GoodneesLeeData* in der Datei *EstimateGoodness* mittels $\hat{\sigma}_\xi^2 = (\sum (\hat{\gamma}_{t+1} - \hat{\gamma}_{t-1} - \nu)^2) / (\text{Anzahl Geburtsjahre} - 2)$ bestimmt. Auch für dieses Modell wurde eine Tabelle wie in (2) erzeugt. Allerdings musste vor der Spalte Alter noch ein Spalte mit dem Kalenderjahr eingefügt werden. Kalenderjahre sind von 0 bis 40 durchnummeriert. Der Fehler für die Datensätze wurde ebenfalls aus der Periodentafel 1965 Westdeutschland geschätzt und beträgt 0.001317417. Dieser Wert wird in der Funktion *GoodneesLeeData* in der Datei *EstimateGoodness* als die Standardabweichung der Abstände der ersten Hälfte des Datensatzes 1956 Westdeutschland zwischen der Schätzung und dem letzten Beobachtungsjahr bestimmt.

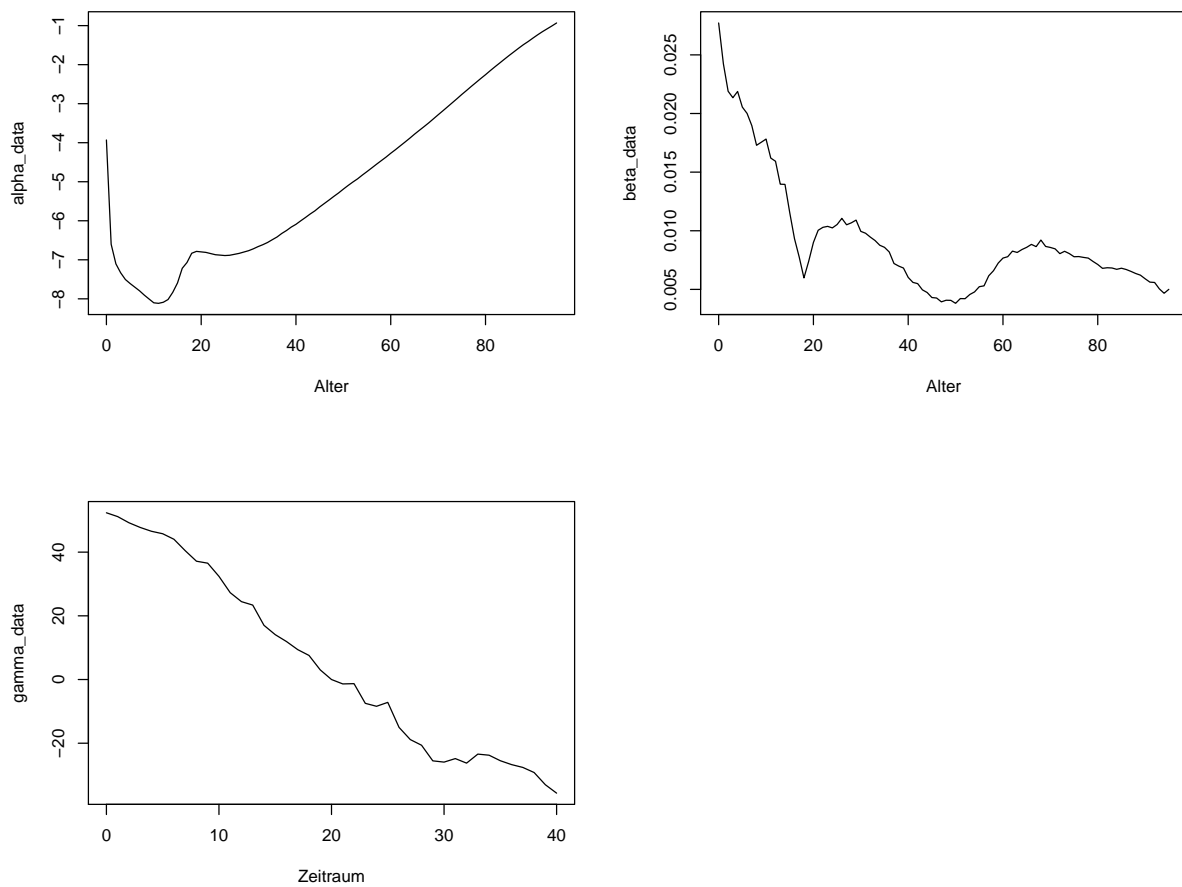


Abbildung 25: Parameters Lee-Carter

In Graphik (25) sieht man wieder ein Beispieldatensatz. Die durchgezogene Linie ist das wahre Modell, die Kreuze sind das wahre Modell Plus white Noise und die gestrichelte Linie zeigt die Periodentafel 1965 Westdeutschland im entsprechenden Kalenderjahr seit Start der Periodentafel im Jahr 1965. Es handelt sich diesmal um den zweiten Datensatz.

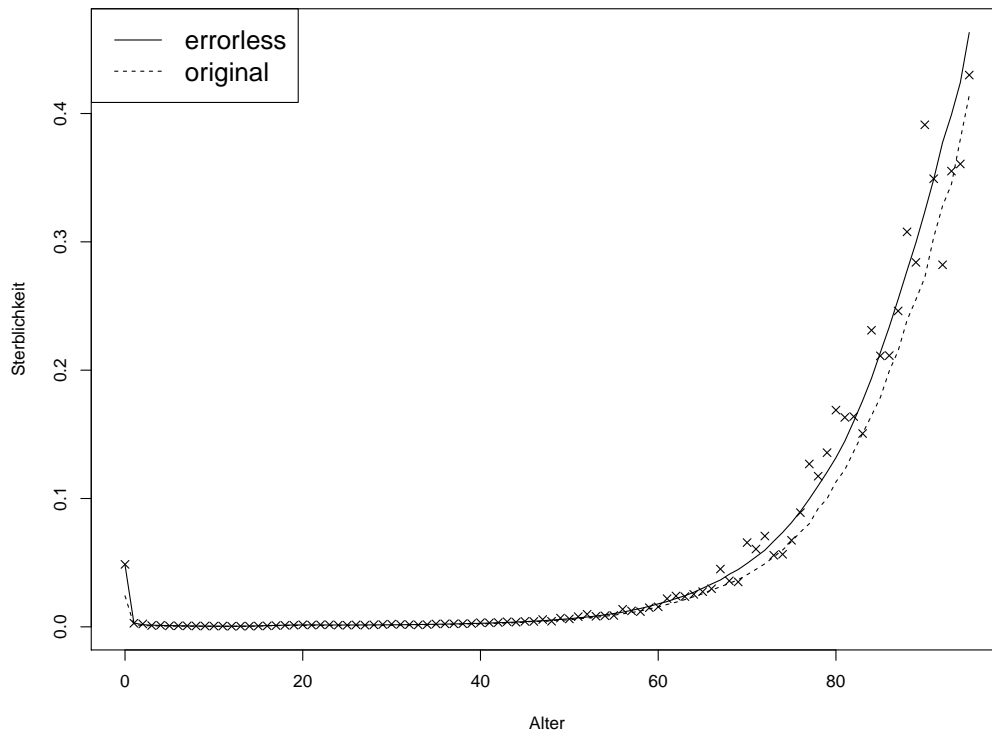


Abbildung 26: Beispiel der Simulationsdaten im Lee-Carter Modell

In Graphik (27) sieht man die Summe der Abstände der Parameter und ihrer Schätzer über die 100 Simulationen. Die durchgezogene Linie ist der Mittelwert der Fehlersummen. Man sieht, dass die Fehler keine Struktur aufweisen. An dieser Stelle sei noch bemerkt, dass ich auch das Singulärwertverfahren, das Lee und Carter vorschlagen, implementiert habe und auch dafür die Simulation durchgeführt habe. Bei dem Singulärwertverfahren gab es immer wieder recht starke Ausreißer bei denen der Fehler bis zum 10-fachen des Maximum-Likelihood Verfahrens betrug. Ich vermute, dass das Singulärwertverfahren numerisch weniger stabil ist, als das Maximum-Likelihood Verfahren.

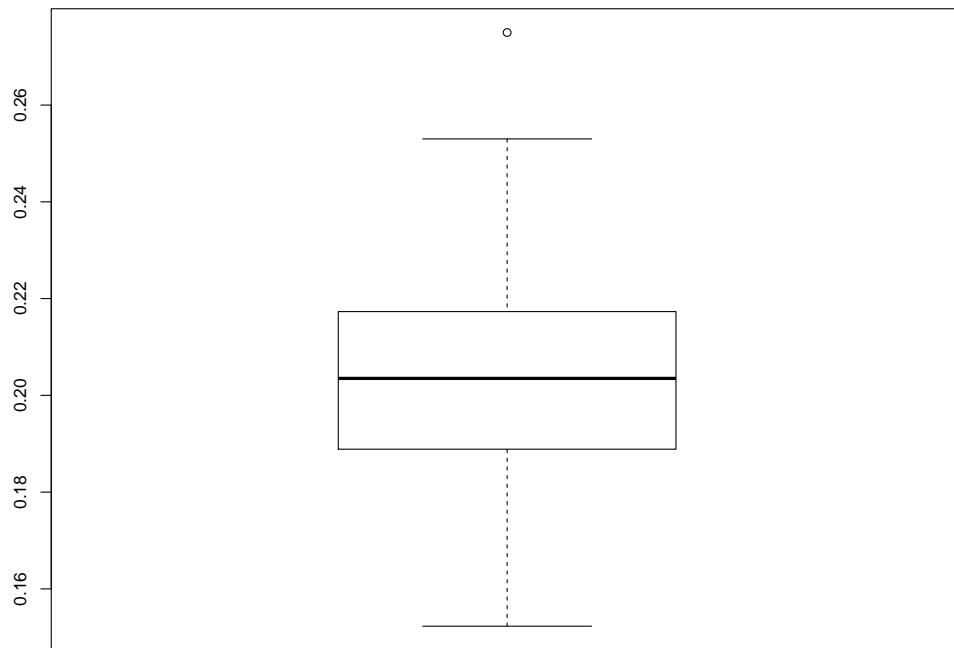


Abbildung 27: Plot der Fehler im Lee-Carter Modell für die Schätzung der Parameter

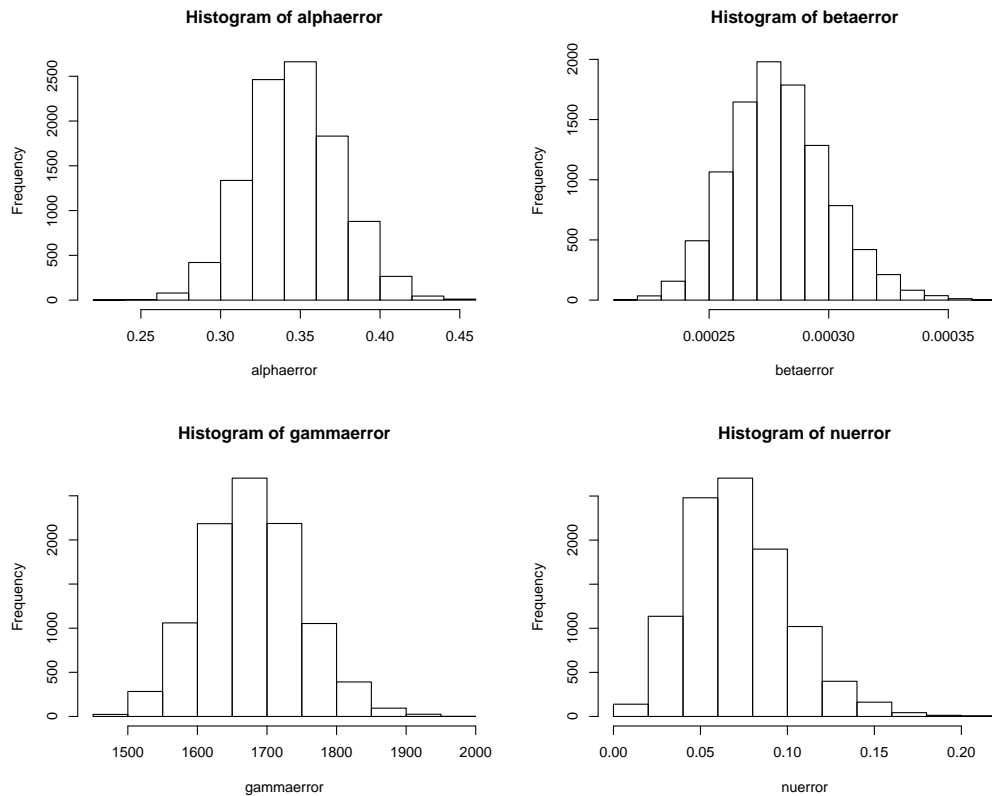


Abbildung 28: TODO

Nachdem das Schätzen der Parameter diskutiert worden ist, betrachten wir nun die Ergebnisse aller Berechnungen, die in Tabelle (3) zusammengefasst sind:

Methode	Einfaches Modell	Whittaker-Henderson	Lee-Carter
Parameter	0.1355731	Nicht parametrisch	7.599837
Fehler pro Parameter	0.06778656	Nicht parametrisch	0.03289973
Datensatz	0.02340596	0.02304943	0.001995316

Tabelle 3: Ergebnisse der Simulationen

Man sieht, dass die Fehler für das Lee-Carter Modell um einen Faktor 80 größer sind, als die Fehler im Einfachen Modell. Dies ist auch nicht weiter erstaunlich, da man $2 * a + t + 1$ Parameter schätzt, während man im einfachen Modell nur 2 Parameter schätzt. Deshalb sind in der mittleren Zeile die durchschnittlichen Fehler über die 100 Durchgänge pro Parameter angegeben. Der Fehler im einfachen Modell ist jetzt um einen Faktor 2 größer, als der Fehler im Lee-Carter Modell. Ich denke, dass die Ergebnisse das angeben, was man erwarten würde. Bei dem Whittaker Modell gibt es keine Parameter und deshalb auch keine Ergebnisse. Betrachten wir nun die Abweichung zwischen dem Schätzer für die zweite Hälfte des Datensatzes und den beobachteten Daten. Man sieht, dass dies für das

einfache Modell und das Whittaker Modell ähnlich sind. Bei dem Lee-Carter Modell sind die Fehler dagegen um eine Faktor 10 kleiner. Entsprechend kann man sagen, dass das Lee-Carter Modell für Sterbetafeln aufgrund dieser Berechnung besser ist. Dies ist genau das Ergebnis, was man erwarten würde. Allerdings war diese Berechnung eventuell ein wenig unfair. Zum einen hat man in der Generationstafel, wenn man sie halbiert wieder Probleme mit der Zensur. Deshalb wurde für das Lee-Carter Modell die Vorhersage mit dem letzten Kalenderjahr verglichen, während bei dem einfachen und dem Whittaker Modell die erste mit der zweiten Hälfte verglichen wurde. Man erhält trotzdem eine etwas seltsame Situation, die in Graphik (29) veranschaulicht ist.

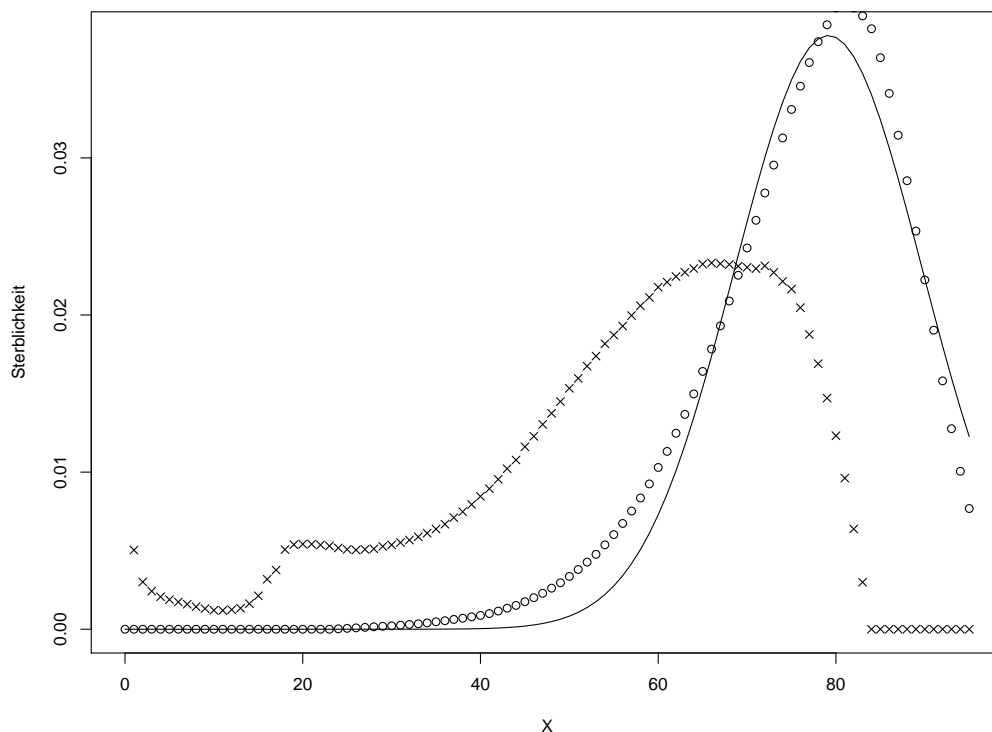


Abbildung 29: Der Schätzer für die Sterblichkeit in der Generationstafel 1956 Westdeutschland ist die durchgezogene Linie. Die erste Hälfte des Datensatzes, auf dem der Schätzer beruht sind die Kreise. Die zweite Hälfte, zu der der Abstand bestimmt wird, sind die Kreuze.

In Graphik (29) 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 circa 16 und 20 Jahren, 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 späteren Datensatz in frühen Jahren höher. Der relativ große Anteil an Personen, die in hohen Altern sterben wurde im späteren 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.

Um diesen Abschnitt abzuschließen hier noch einige Informationen 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*, *GoodneesLeeData* für den Datensatz und *GoodnessEinfachEinfachSim*, *GoodneesLeeSim* für die Parameter. 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 und das Lee-Carter Modell befinden sich in der Datei *DAV*.

Literatur

- [Beh85] H. Behrens, *Methodik von Sterblichkeitsuntersuchungen*, Ser. Schriftenreihe angewandte Versicherungsmathematik. Verlag Versicherungswirtschaft, 1985. Adresse: <https://books.google.de/books?id=GZvRSgAACAAJ>.
- [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 DGVM*, 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.
- [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.
- [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.
- [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>.
- [Roe88] C. S. Roehrig, »Conditions for identification in nonparametric and parametric models«, *Econometrica*, Bd. 56, Nr. 2, S. 433–447, 1988, ISSN: 00129682, 14680262. Adresse: <http://www.jstor.org/stable/1911080>.
- [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.

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