

Trend Estimation of Mortality Rates

In den vergangenen Jahren wurde der demographische Wandel zu einem immer wichtigeren Thema. Viele westliche Länder, darunter auch Deutschland, haben mit Überalterung und den daraus resultierenden Problemen bezüglich des Generationenvertrages zu kämpfen. Besonders das Konzept der Sozialversicherungssysteme sind dabei bedroht. Abgesehen von der gesellschaftlichen Problematik, bringt das Risiko der Langlebigkeit enorme Herausforderungen für die finanzielle Stabilität von Versicherungsunternehmen mit sich. Die Konditionen von Lebens- und Rentenversicherungsverträgen basieren auf den zu dieser Zeit vorhandenen Sterblichkeitsannahmen. Werden die künftigen Entwicklungen unterschätzt, kann das zu großen Verlusten für die Versicherungen führen. Für deutsche Lebensversicherer bildet das Rentengeschäft einen großen Bestandteil des Gesamtumsatzes, es ist daher besonders wichtig, sich bei den Berechnungen auf bestmögliche Annahmen zu stützen. Bevor fundierte Vorhersagen für die Zukunft getroffen werden können, ist es notwendig die historische Entwicklung anhand von empirischen Daten zu untersuchen.

Zielsetzung des Forschungsprojektes

Durch die Modellierung von empirischen Daten sollen zu Grunde liegende Trendstrukturen geschätzt werden. Ziel ist es hierbei so wenige Einflussgrößen wie möglich zu benutzen, unter der Prämisse, die grundlegende Struktur beizubehalten. Dadurch ist man fähig auch ohne die Komplexität, die viele Parameter mit sich bringen würden, präzise Vorhersagen zu treffen. Bisher sind die Algorithmen, mit deren Hilfe derartige Strukturkurven erstellt werden, sehr kompliziert und haben lange Laufzeiten. In diesem Projekt haben wir ein Vorgehen entwickeln, welches eine Berechnungszeit von unter einer Stunde hat und die bestmögliche Trendschätzung mit maximal zehn Regressionsparametern für einen Datensatz berechnen kann, der ursprünglich durch 356 Koeffizienten beschrieben wurde.

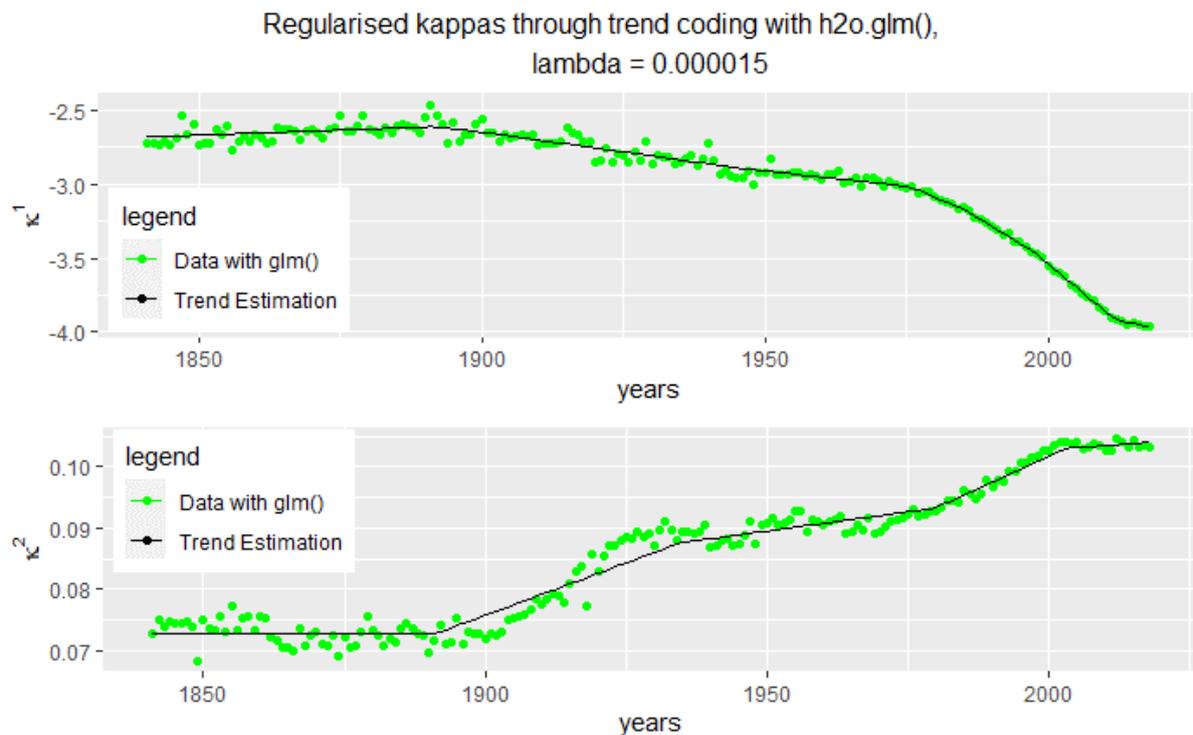
Struktur der Arbeit und die zugrunde liegenden Modelle

Wir verwenden Daten über die männliche Zivilbevölkerung von England und Wales aus den Jahren 1841 bis 2018. Dieser Datensatz wird verwendet, um die Ergebnisse vergleichbar zu der Arbeit von Börger und Schupp (2018) zu machen. Die in diesem Forschungsprojekt entwickelte Vorgehensweise kann allerdings auch auf andere Daten angepasst werden.

Grundlage der Analysen ist das Sterblichkeitsmodell von Cairns, Blake und Dowd (2006). Es verwendet zwei stochastische Parameter. κ_t^1 beschreibt dabei das allgemeine Sterblichkeitsniveau und κ_t^2 die unterschiedliche Entwicklung in verschiedenen Altersgruppen für ein ausgewähltes Jahr t . Dieser Ansatz, dargestellt als verallgemeinerte lineare Modelle, wird in der Software R implementiert. Die Ausgabe bildet eine Graphik, welche den Verlauf von κ_t^1 und κ_t^2 zeigt. Auf Grund des Datensatzes handelt es sich hierbei um insgesamt 356 Werte. Die Richtigkeit der Ergebnisse kann mit der Methode `cbd()` des `StMoMo` Paketes in R sichergestellt werden. Außerdem wird dabei die Trendcodierung eingeführt. Es handelt sich dabei um die Idee, nicht jede Merkmalsausprägung einer Datenreihe selber zu betrachten, sondern die Trendänderung zum vorangegangenen Wert zu untersuchen. Diese Herangehensweise erweist sich als besonders hilfreich, erfordert allerdings eine weitere Bearbeitung durch Anwendung der Techniken „Regularisierung“ und „Filterung“.

Regularisierung

Wir verwenden eine Lasso Regularisierung, die mittels der `h2o.glm()` Funktion in R umgesetzt wird. Dadurch wird ein Modell ermittelt, bei dem die Struktur des Datensatzes mit 26 Koeffizienten beschrieben werden kann. Hat man den passenden Regularisierungsparameter λ gefunden, kann das Modell in deutlich weniger als einer Minute berechnet werden (mit einem handelsüblichen Laptop).



Nachbearbeitung und Filterung

Visuell bildet die entwickelte Kurve den Trend sehr gut ab. Um die Anzahl der Parameter zu verringern, wenden wir anschließend eine Filterung an. Die Idee ist dabei, für eine festgelegte Anzahl an aktiven Parametern, das bestmögliche Modell zu ermitteln. Dafür berechnet das beschriebene R Skript für jede Kombinationsmöglichkeit das entsprechende Regressionsmodell. Dabei wird das entstandene Modell anhand von Gütekriterien wie dem Akaike-Informationskriterium (AIC) bewertet, um am Ende die bestmögliche Trendabschätzung auswählen zu können.

Fazit

Im Rahmen dieses Forschungsprojektes ist es gelungen, eine fundierte Trendschätzung für die Daten der männlichen Zivilbevölkerung von England und Wales zu erstellen. Eine solche Analyse bildet den Grundstein für präzise Sterblichkeitsvorhersagen, welche von großer Bedeutung für Versicherungsunternehmen sind. Durch die Filterung wird die Ermittlung einer Trendstruktur mit lediglich zehn Einflussgrößen ermöglicht.