



DAV

DEUTSCHE  
AKTUARVEREINIGUNG e.V.

Ergebnisbericht des Ausschusses Schadenversicherung

## **Auswirkungen der COVID-19-Pandemie – Modellierung und Prognose mit Hilfe von Betroffenheitsvariablen**

Köln, 01.06.2022

## **Präambel**

Der Ausschuss Schadenversicherung der Deutschen Aktuarvereinigung e. V. hat den vorliegenden Ergebnisbericht erstellt.<sup>1</sup>

## **Zusammenfassung**

Der Ergebnisbericht behandelt Fragestellungen zur Modellierung und Prognose bei der Tarifgestaltung unter Berücksichtigung der Auswirkungen der COVID-19-Pandemie. Er betrifft Aktuar\*innen in der Schaden-Unfall-Versicherung im Rahmen der Ausführung aktuarieller Aufgaben im Rahmen der Gestaltung von Tarifen.

Der Ergebnisbericht ist an die Mitglieder und Gremien der DAV zur Information über den Stand der Diskussion und die erzielten Erkenntnisse gerichtet und stellt keine berufsständisch legitimierte Position der DAV dar.<sup>2</sup>

## **Verabschiedung**

Der Ergebnisbericht ist durch den Ausschuss Schadenversicherung am 01.06.2022 verabschiedet worden. Er ersetzt die gleichnamige Vorversion vom 21. Februar 2022, die um das Kapitel 5 erweitert wurde.

---

<sup>1</sup> Der Ausschuss dankt der Arbeitsgruppe Tarifierungsmethodik ausdrücklich für die geleistete Arbeit, namentlich Dr. Stefan W. Wetzel (Leitung), Jochen Kneiphof, Daniel König, Ina Kühnel, Dr. Michael Schüte und Dr. Wiltrud Weidner. Weiterhin dankt der Ausschuss der Gen Re, sowie allen am Gen Re Kraftfahrt-Datenpool beteiligten Unternehmen für die Ermöglichung des Use Cases durch das Bereitstellen der Daten bzw. Analysen.

<sup>2</sup> Die sachgemäße Anwendung des Ergebnisberichts erfordert actuarielle Fachkenntnisse. Dieser Ergebnisbericht stellt deshalb keinen Ersatz für entsprechende professionelle actuarielle Dienstleistungen dar. Actuarielle Entscheidungen mit Auswirkungen auf persönliche Vorsorge und Absicherung, Kapitalanlage oder geschäftliche Aktivitäten sollten ausschließlich auf Basis der Beurteilung durch eine(n) qualifizierte(n) Aktuar DAV/Aktuarin DAV getroffen werden.

## Inhaltsverzeichnis

<b>1. Einführung in die Idee .....</b>	<b>5</b>
1.1. Ausgangssituation .....	5
1.2. Bekannte Lösungsansätze des Prognoseproblems .....	6
1.3. Prognose über Betroffenheitsvariablen.....	6
<b>2. Datenanforderungen &amp; Umgang mit Betroffenheitsvariablen .....</b>	<b>8</b>
2.1. Fortschreibung von Betroffenheitsvariablen .....	9
<b>3. Betroffenheitsvariablen – Quellen und Beispiele .....</b>	<b>11</b>
3.1. Messdaten der Corona-Epidemie in Deutschland: .....	14
3.2. Staatliche Maßnahmen .....	15
3.3. Allgemeine Aktivitätsparameter .....	16
3.3.1. Mobilitätsdaten.....	17
3.3.2. Allgemeinere Aktivitätsdaten .....	18
3.3.3. Speziell: Arbeitsmarktstatistiken .....	18
<b>4. Nutzung der Betroffenheitsvariablen in der Tarifierung .....</b>	<b>19</b>
4.1. Modellierung mit Betroffenheitsvariablen.....	19
4.2. Prognose mit Betroffenheitsvariablen .....	20
<b>5. Use Case .....</b>	<b>22</b>
5.1. Datengrundlagen .....	22
5.1.1. Datenaufbereitung - Betroffenheitsvariable .....	23
5.2. Modellierung & Prognose der Betroffenheitsvariable .....	24
5.3. Modellierung mit Betroffenheit .....	25
5.3.1. Saisonale Effekte .....	25
5.3.2. Ergebnisse für die Betroffenheitsvariable .....	26
5.3.3. Auswirkungen auf das restliche Modell .....	28
5.3.4. Niveau und Prognose.....	28
5.4. Praktische Hinweise aus dem Use Case .....	29
<b>6. Fazit und Ausblick .....</b>	<b>31</b>
6.1. Modellierung mit Betroffenheitsvariablen in der COVID-19-Pandemie. 31	

6.2. Anwendungsmöglichkeiten über Corona hinaus .....	31
<b>7. Anhang .....</b>	<b>32</b>
7.1. Berechnung des R-Wertes .....	32
7.2. Beispiel-Programme zur Berechnung einiger Betroffenheitsvariablen .	32

## **1. Einführung in die Idee**

Bei der Tarifgestaltung in der Schaden- und Unfallversicherung sollten Aktuar\*innen für eine dem Untersuchungsgegenstand angemessene Datenbasis sorgen. Dazu gehört, auch die Angemessenheit der Daten im Hinblick auf das Untersuchungsziel zu überprüfen. In der Regel verfügen Aktuar\*innen über einen Datensatz, der mehrere Jahre umfasst. Hierbei ist allerdings zu beachten, dass dieser aus mehreren Jahren bestehende Datensatz Merkmale aufweisen kann, die selbst im Zeitverlauf einer Veränderung unterliegen und somit einen Effekt auf die Zielgrößen der zu erstellenden Modelle haben. In der Regel stellt es kein Problem dar, wenn sich die Verteilungen über die Ausprägungen von Tarifmerkmalen ändern. Schwieriger wird es jedoch, wenn Änderungen der modellierten Zielgröße auftreten, die nicht modellinhärent sind, sondern aufgrund eines zusätzlichen externen Effektes, wie bspw. der COVID-19-Pandemie, induziert werden. An dieser Stelle sollten neue Variablen (genannt: Betroffenheitsvariablen) bei der Modellierung berücksichtigt werden, mit denen sich die relevanten Zeiträume charakterisieren lassen. Die in diesem Bericht vorgestellten Lösungsansätze unter Verwendung von sogenannten Betroffenheitsvariablen können auch auf andere Situationen (wie z.B. Klimawandel etc.) mit vergleichbarer Ausgangssituation (d.h. verändertes Verhalten aufgrund eines externen Effektes) übertragen werden. Betroffenheitsvariablen sind dabei Variablen, bei denen man unterstellt, dass sie einen Einfluss auf eine (temporäre) Verhaltensänderung haben. Damit wird auch das Datenmaterial für die Modellierung und Prognose nutzbar, das von einer (temporären) Verhaltensänderung betroffen ist.

### **1.1. Ausgangssituation**

Durch die Wirkung der COVID-19-Pandemie inklusive der dagegen getroffenen Maßnahmen auf die Daten der Schaden-Unfall-Versicherer ist ein (möglicherweise akutes) Prognoseproblem entstanden. Die Extrapolation der vorhandenen, historischen Daten in die Pandemiephase hat sich bereits jetzt als ungeeignet, wenn auch nicht wirtschaftlich schädlich gezeigt. Eine Fortschreibung der heutigen Daten auf eine Zeit nach der Pandemie erweist sich bei näherer Betrachtung bereits als kritisch, da die bekannten Probleme mit den Daten zukünftig zu versicherungstechnischen Verlusten führen können. Es ist also auf die Besonderheiten der COVID-19-Pandemie Rücksicht zu nehmen. In diesem Ergebnisbericht stellen wir den Aktuar\*innen eine mögliche Methode vor, die einen bewussten Umgang mit den Effekten der COVID-19-Pandemie ermöglicht. Dieser Vorgehensvorschlag wurde im Artikel „Über den Umgang mit den Spuren der Corona-Pandemie in den Kalkulationsdaten der Schaden-Unfall-Versicherung“<sup>3</sup> erstmals beschrieben. Dieser Ergebnisbericht beschreibt nun die Methode im Detail und bietet konkrete Hinweise und Hilfestellungen für die praktische Anwendung.

---

<sup>3</sup> Vgl. Kneiphof, J., Wetzels, S., Weidner, W.: Telematik in der Kfz-Versicherung: Über den Umgang mit den Spuren der Corona-Pandemie in den Kalkulationsdaten der Schaden-Unfall-Versicherung, Der Aktuar 01.2021, 18-23 (2021).

## **1.2. Bekannte Lösungsansätze des Prognoseproblems**

Es gibt mehrere bekannte Ansätze zur Lösung derartiger Prognoseprobleme.

Als eine gängige Strategie könnte der entsprechende Zeitraum aus den Daten ausgeschlossen werden. Die Modellierung wird sodann ohne Berücksichtigung der Daten dieses Zeitraums vorgenommen. Daran anschließend werden die Effekte des ausgeschlossenen Ereignisses bewertet und in der Prognose berücksichtigt.

Für die COVID-19-Pandemie bedeutet dies allerdings den Ausschluss des Jahres 2020 (spätestens ab März) und auch des kompletten Jahres 2021. Hier sind nicht nur die Schadendaten, sondern auch die Bestandszusammensetzung bis hin zur Bedeutung von Kovariablen, betroffen. Mittels multivariater Residualanalyse werden – auf der Basis der erstellten Modelle – die ausgeschlossenen Daten auf mögliche COVID-19-Effekte hin untersucht. Ziel dieser Untersuchung ist die korrekte Berücksichtigung der Effekte für eine angemessene Prognose. Die Modelle, die sich aus der Modellierung ohne Berücksichtigung des entsprechenden Zeitraums ergeben, werden als a-priori-Modelle und damit als Offset für die modellierte Zielgröße – jetzt mit dem vorher weggelassenen Zeitraum – verwendet. Als Ergebnis erhält man ein detailliertes Bild über die Abweichungen vom „Old Normal“.

Ein ähnliches Ergebnis könnte man auch dadurch erzielen, dass man die betroffenen Daten nicht ausschließt, die Statistikjahre 2020 und 2021 aber als Kovariablen und in Interaktionen mit anderen Variablen in der eigentlichen Modellierung berücksichtigt. Dieses Vorgehen hat den Vorteil, dass auf eine mehrstufige Modellkaskade verzichtet werden kann, welche nicht unbedingt zum optimalen Ergebnis führen muss. Gleichzeitig sei aber angemerkt, dass die Suche nach den notwendigen Interaktionen bei manueller Durchführung sehr zeitaufwändig sein kann.

Beide Ansätze weisen jedoch den Nachteil auf, dass die zugrundeliegenden Effekte nicht separiert werden können, was eine Prognose für die kommenden Jahre im Sinne eines „New Normal“ erschwert. Zudem wird hierbei eine Betrachtung auf Jahresbasis vorgenommen, die bei unterjährigen Verhaltensänderungen möglicherweise grob vereinfachend ist und auf vorhandene Effekte glättend wirkt.

## **1.3. Prognose über Betroffenheitsvariablen**

Aufgrund der durchaus vorhandenen Unzulänglichkeiten der im vorherigen Abschnitt genannten Ansätze sollte geprüft werden, die Betroffenheit durch mehrere Variablen direkt am Exposure zu berücksichtigen und explizit in die Modellierung einzubeziehen. Dieses Vorgehen setzt eine andere, komplexere Datenvorbereitung voraus, verspricht aber im Nachgang explizite, ursachenorientierte Ergebnisse, mit denen auch eine Prognose direkter möglich wird. Das Vorgehen besteht dabei aus sieben Schritten:

1. Identifikation und Modellierung neuer Betroffenheitsvariablen;
2. Ergänzung der Bestandsdatensätze um diese Informationen;
3. Korrektur der Schadenzahlungen um Effekte, die Einfluss auf die Schadenhöhe haben (bspw. Bei COVID-19: Mehrwertsteuersenkung);

4. Ergänzung der Schadendaten um zugehörige Bestandsinformationen;
5. Ergänzung der Bestandsdaten um zugehörige Schadeninformationen;
6. Modellierung der abhängigen Variablen;
7. Prognose mithilfe der Modellergebnisse der Betroffenheitsvariablen.

Die Schritte 4-6 entsprechen dem bekannten Vorgehen in der Tarifierung in der Schaden-Unfall-Versicherung, wobei lediglich neue Kovariablen für die Modellierung aufgenommen werden. Die Kalkulationsdaten haben bei diesem Vorgehen einen Zeitbezug, weswegen saisonale Effekte separat berücksichtigt werden müssen.

Da das Ziel einer aktuariellen Analyse bei der Tarifgestaltung nicht nur die bestmögliche Beschreibung der Vergangenheit, sondern auch eine Prognose der zukünftigen Entwicklung des Schadengeschehens ist, ist bei der Prognose in die Zukunft auch das zukünftige Verhalten dieser Betroffenheitsvariable ins Kalkül einzubeziehen. Hierbei ist es wichtig zu differenzieren, ob die Betroffenheitsvariablen in direkter oder nur mittelbarer Verbindung zu der beobachteten Verhaltensänderung stehen. So hat im Beispiel COVID-19 die Verhängung des Lockdowns sicherlich direkten Einfluss auf die Fahrleistung der PKW und damit die Schadenhäufigkeit, während bspw. die Tankmenge pro Tag nur ein Resultat aus der beobachteten Verhaltensänderung "Verstärkte Nutzung von Home-Office", verbunden mit geringerer PKW-Nutzung und geringerer Fahrleistung, ist. Hier ist bei einer Prognose sehr sorgfältig abzuwägen, welchen Einfluss die Betroffenheitsvariablen tatsächlich haben.

Zudem kann der Datensatz unter Berücksichtigung von Betroffenheitsvariablen deutlich größer werden, da bspw. aufgrund der Zeitabhängigkeit der Betroffenheitsvariablen die bei der Modellierung zu berücksichtigenden Zeitscheiben vielfältig werden. Daher sei an dieser Stelle die Warnung vor großen Datenmengen mit allen – durchaus auch statistisch unangenehmen – Begleiterscheinungen angebracht.

## **2. Datenanforderungen & Umgang mit Betroffenheitsvariablen**

Die Abbildung der vielfältigen und segmentspezifischen Effekte der COVID-19-Pandemie wird über die Modellierung von Betroffenheitsvariablen empfohlen. Eine große Rolle spielt der richtige Grad der Granularität an verfügbaren Informationen. Einerseits gilt es, systematische Effekte möglichst akkurat von zufälligen Effekten trennen und abbilden zu können, andererseits hat ein progressiv wachsendes Datenvolumen massive Auswirkungen auf die Rechenlaufzeit während der Tarifkalkulation und mittelbar auch auf Interpretationsmöglichkeiten der Kalkulationsergebnisse.

Die Modellierung einer Betroffenheitsvariablen setzt zunächst eine umfangreiche Datenvorbereitung, d.h. eine saubere Zusammenstellung, Bereinigung und Konsolidierung von Rohdaten, voraus. Während das Zusammenführen der Informationen aus verschiedenen internen und externen Datenquellen zeitaufwendig ist, ist die Datenbereinigung und -konsolidierung entscheidend für die Prognosegüte. Auf die Datenerfassung erfolgt dabei zunächst eine Datenkontrolle, vornehmlich zur Überprüfung der Vollständigkeit und Plausibilität. Im Anschluss müssen die Daten in eine für die Modellierung taugliche Form gebracht werden.

Um die Eignung sicherzustellen, ist folgende, grundlegende Form der notwendigen Variablen anzustreben:

- Name: Kürzel / Label der Variablen
- Datentyp der Variablen: qualitative oder quantitative Merkmalsausprägungen
- Wertebereich: Einordnung der beobachtbaren und messbaren Werte, Abschluss/Ersetzung fehlender Werte
- n-dimensionale Hierarchie: Zeitraumbezug, Regionalität, Berufsbezug, usw.

Dabei ist es ggf. notwendig, die Daten zu transformieren, zu erweitern oder zu verdichten. Zusätzlich kann die Notwendigkeit der Neudefinition von Ausprägungen einer oder mehrerer Variablen bestehen. Zur Gruppierung können sich aufgrund der semantischen Zugänglichkeit agglomerative Clusterverfahren anbieten. Über die Verdichtung von Daten sollten sich Aktuar\*innen frühzeitig Gedanken machen, da aufgrund der Systematik sonst schnell sehr große Datensätze entstehen. Andererseits kann eine Zusammenfassung von Ausprägungen zu einer ungeeigneten Glättung der Daten führen. Deshalb ist immer eine Abwägung zwischen dem Datenvolumen und dem mit der Verdichtung einhergehenden Informationsverlust zu treffen.

Analog der üblichen Sorgsamkeit im Hinblick auf die Datenqualität im Rahmen der Tarifkalkulation ist zu regelmäßigen Zeitpunkten stets die Aktualität der Datenbasis zu überprüfen und bei Bedarf zu aktualisieren.

Folgende Kriterien sind idealerweise zu erfüllen, um modellierte Betroffenheitsvariablen geeignet in der Tarifkalkulation berücksichtigen zu können:



- Ursachenorientierter Zusammenhang der Betroffenheit von Corona auf das Exposure. Der Signifikanznachweis geeigneter Hypothesen erfolgt zunächst individuell für jede Variable
- Granulare Erfassung und saubere Trennung von Ereignissen und politischen Entscheidungen sowie deren Umsetzung
- Qualitative und soweit möglich quantitative Erfassbarkeit der Betroffenheit von Corona und die dagegen getroffenen Maßnahmen
- Zuordenbarkeit des jeweiligen Zeitabschnitts in der Entwicklung der COVID-19-Pandemie an den entsprechenden Zeitabschnitt des Exposures
- Statistische Eigenschaften: Stabilität, Objektivität, Reliabilität, Validität

Neben dem ursachenbezogenen Zusammenhang zwischen dem Effekt einer Betroffenheitsvariablen und dem Exposure kann es entscheidend sein, sich einen Überblick über die Wechselwirkungen zwischen den modellierten Variablen zu verschaffen. Durch Kollinearitäten kann die Modellschätzbarkeit möglicherweise abhandenkommen bzw. damit eine starke Erhöhung der Varianzen der Schätzer einhergehen. Im Zweifel gilt es hier, weniger Betroffenheitsvariablen aufzunehmen, um Korrelationen, Kollinearitäten und Kovarianzen zwischen ihnen zu reduzieren, um später keine aufwendigere Regularisierung durchführen zu müssen.

### **2.1. Fortschreibung von Betroffenheitsvariablen**

Die Erfassung von Betroffenheitsvariablen auf Basis von beobachteten bzw. gemessenen Daten der Vergangenheit ist lediglich ein Teilaspekt, zusätzlich muss die zukünftige Fortschreibung und Entwicklung betrachtet werden. Hierbei stellt sich allerdings heraus, dass die Entwicklung der grundlegenden Variablen in der Zukunft einfacher zu erfassen sein wird. Aufgrund eines jetzt geschafften Bewusstseins der Erforderlichkeit entsprechender Variablen kann die Erfassung i.d.R. recht zeitnah erfolgen. Zur Prognose eines zugrundeliegenden Trends genügt in der Regel die Anwendung einer Trendexploration als relativ einfaches Prognoseverfahren. Wie üblich, wird dazu je nach Verfügbarkeit empfohlen, auf Zeitreihen (z.B. Verkehrsentwicklung) zurückzugreifen und *nicht* lediglich die von der COVID-19-Pandemie beeinflussten Jahre 2020/21 zu betrachten. Darüber hinaus sind aktuelle Entwicklungen der Betroffenheitsvariablen zeitnaher verfügbar als übliche Kalkulationsdaten.

Wichtiger als die Verfahrensauswahl ist, wie oft in der aktuariellen Praxis, allerdings die Vorbereitung und Herangehensweise. Eine separate Fortschreibung verschiedener Werte scheint dabei mit Blick auf eine differenzierte Prognose unverzichtbar. In Bezug auf die Verwendung der Fortschreibung der Betroffenheitsvariablen für die Prognose der Modellierung gelten weitere Anforderungen, auf die im Abschnitt 4 näher eingegangen wird. So gibt es bereits konkrete Hinweise für einerseits *temporäre*, vgl. z.B. die zeitweise Mehrwertsteuersenkung, andererseits aber auch *dauerhafte* Veränderungen, vgl. z.B. die Entwicklung von Homeoffice-Quoten. Bei letzterem ist zunächst eine Annahme über das mögliche „New Normal“ zu treffen.

Ferner sind die „unterjährigen“ Variablenausprägungen im Hinblick auf zyklische Schwankungen in Form von saisonalen Komponenten (z.B. andere Verkehrsdichte/-aufkommen zu Ferienzeiten, Transportleistung auf Schiffen, aber auch hausgemachte Abwicklungs- und Meldeeffekte im Datensatz zum Jahresende) zu überprüfen. Liegen saisonale Effekte vor, sind diese durch die Bestimmung einer Saisonkomponente, welche periodendurchschnittliche Abweichungen aufzeichnet, zu behandeln. Durch sie lässt sich dann die Frage beantworten, wie die Zeitreihe verlaufen würde, wenn ihr Verlauf nicht durch saisonale Faktoren superponiert worden wäre und es lässt sich auf den saisonal bereinigten Daten eine Trendanalyse anwenden.

Zu guter Letzt ist es zweckmäßig zu verdeutlichen, dass Prognosen mit Hilfe eines Trends zwar Vorhersagen über längere Zeiträume hinweg ermöglichen, sich Trends aber auch von einem Jahr aufs andere ändern können und gerade bei linearen oder quasi-linearen Fortschreibungen Prognosen, je weiter sie sich von dem letzten beobachteten Wert entfernen, immer unsicherer werden.

### **3. Betroffenheitsvariablen – Quellen und Beispiele**

Die für sinnvoll erachteten Betroffenheitsvariablen liegen üblicherweise nicht im unternehmensinternen Datenhaushalt vor, sondern stammen in der Regel aus externen Datenquellen. Diese Datenquellen können sowohl frei zugänglich sein als auch von kommerziellen Anbietern kommen.<sup>4</sup> In beiden Fällen sollten Aktuar\*innen vor Verwendung der Daten immer die fachliche Eignung, deren Aktualität und vor allem deren Güte prüfen.

Die Datenlage zu Betroffenheitsvariablen zu COVID-19 erweist sich, sowohl was den Inhalt als auch Struktur und Qualität der Daten betrifft, als extrem dynamisch. Daher ist bei den folgenden Ausführungen zu beachten, dass sich die Beschreibung der Datenquellen auf den Stand zu Oktober 2021 bezieht und jede Quelle vor der Verwendung auf etwaige Veränderungen in Struktur und Inhalt zu überprüfen ist. Sicherlich werden Aktuar\*innen bei der Suche nach Betroffenheitsvariablen in der Zukunft weitere aufschlussreiche Quellen finden.

Eine gute Quelle für frei zugängliche und vielfältige Daten ist das statistische Bundesamt, wobei die Aktualität und Struktur einer Statistik stark vom Thema abhängt. Bezogen auf Corona erweist sich unter anderem das RKI als gute kostenfreie Datenquelle mit hoher Aktualität.

Im Folgenden diskutieren wir einige Quellen unter verschiedenen Aspekten ohne Anspruch auf Vollständigkeit. Es sei darauf hingewiesen, dass sich – in der Folge der COVID-19-Pandemie, aber auch allgemein – die verfügbare Datenwelt dynamisch weiterentwickelt, und dass daher diese Zusammenstellung nur den gegenwertigen Stand widerspiegeln kann:

---

<sup>4</sup> Wir beschränken uns im Folgenden auf die Betrachtung frei zugänglicher Datenquellen.

Quelle	Inhalt	Update	Granularität	Link/Referenz
<b>RKI</b>	COVID-19 Datenhub: Messzahlen zur Pandemie, Erkrankungen, Genesen, Todesfälle	täglich	Lokal: KGS5 sonst	<a href="https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0/explore">https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0/explore</a>
RKI	Informationen zu COVID-19-Tests	wöchentlich	Lokal: Deutschland	<a href="https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Testzahl.html">https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Testzahl.html</a> ;
RKI	Impfmonitor	wochentäglich	Lokal: KGS5, sonst	<a href="https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Impfquoten-Tab.html">https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Impfquoten-Tab.html</a> ; <a href="https://github.com/robert-koch-institut/COVID-19-Impfungen_in_Deutschland/">https://github.com/robert-koch-institut/COVID-19-Impfungen_in_Deutschland/</a>
Divi – Intensivregister	Messreihen zur Belegung von Intensivbetten	Täglich	Lokal: KGS5, sonst	<a href="https://www.intensivregister.de/#/aktuelle-lage/downloads">https://www.intensivregister.de/#/aktuelle-lage/downloads</a>
<b>Our World in Data</b>	Policy Responses to the Coronavirus Pandemic - diverse Daten zur Pandemie und den Maßnahmen (free to use) nur Deutschlandweit	täglich	lokal: Deutschland im internationalen Vergleich	<a href="https://ourworldindata.org/policy-responses-covid">https://ourworldindata.org/policy-responses-covid</a>
<b>Destatis</b>	Mobilitätsdaten (experimentell) im Vergleich zum VJ	Täglich	Lokal: Bundesland/ Deutschland	<a href="https://www.destatis.de/DE/Themen/Querschnitt/Corona/Gesellschaft/kontextinformationen-gesellschaft.html#Mobi">https://www.destatis.de/DE/Themen/Querschnitt/Corona/Gesellschaft/kontextinformationen-gesellschaft.html#Mobi</a>
<b>COVID-19 Mobility Projekt</b>	Allgemeine Forschungsergebnisse zur Mobilität auf der Basis von Mobilfunkdaten	diverse	diverse (u.a. lokal mit NUTS3)	<a href="https://www.covid-19-mobility.org/de/">https://www.covid-19-mobility.org/de/</a> GitHub - rocs-org/covid-mobility-data: The daily mobility change in Germany as depicted in the covid mobility project
Destatis	Diverse Statistiken als Zeitreihen für wirtschaftliche und gesellschaftliche Indikatoren	Monatlich / quartärllich	Deutschland	<a href="https://www.destatis.de/DE/Themen/Querschnitt/Corona/_inhalt.html">https://www.destatis.de/DE/Themen/Querschnitt/Corona/_inhalt.html</a>

Quelle	Inhalt	Update	Granularität	Link/Referenz
	Experimentelle Daten: "neue, innovative Projektergebnisse"		Deutschland	<a href="https://www.destatis.de/DE/Service/EXDAT/inhalt.html;jsessionid=29B672BA0193AE4B438D6D7A740D39CB.internet711">https://www.destatis.de/DE/Service/EXDAT/inhalt.html;jsessionid=29B672BA0193AE4B438D6D7A740D39CB.internet711</a>
	Experimentelle Daten: Kreditvergaben in der Kreditwirtschaft und Auskünfte für Online-Transaktionen	wöchentlich	Deutschland	<a href="https://www.destatis.de/DE/Service/EXDAT/Datensaetze/kredite-onlinetransaktionen.html">https://www.destatis.de/DE/Service/EXDAT/Datensaetze/kredite-onlinetransaktionen.html</a>
	Wirtschaftliche Auswirkungen	diverse	Deutschland	<a href="https://www.destatis.de/DE/Themen/Querschnitt/Corona/Wirtschaft/kontextinformationen-wirtschaft.html#einzelhandel">https://www.destatis.de/DE/Themen/Querschnitt/Corona/Wirtschaft/kontextinformationen-wirtschaft.html#einzelhandel</a>
	Statistik Dossier: Daten zur COVID-19-Pandemie, Sammlung von Verweisen auf Zeitreihen, die unter COVID-19 beeinflusst wurden.	Halbjährlich	i.a. Deutschland	<a href="https://www.destatis.de/DE/Themen/Querschnitt/Corona/Downloads/dossier-covid-19.html">https://www.destatis.de/DE/Themen/Querschnitt/Corona/Downloads/dossier-covid-19.html</a>
	Projekt im Auftrag vom stat. Bundesamt: diverse Statistiken	diverse	diverse	<a href="https://www.corona-daten-deutschland.de/">https://www.corona-daten-deutschland.de/</a>
<b>Agentur für Arbeit</b>	Arbeitsamtdaten (Kurzarbeit, Arbeitslosigkeit etc.)			<a href="https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Themen-im-Fokus/Corona/Corona-Nav.html">https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Themen-im-Fokus/Corona/Corona-Nav.html</a>
<b>BAST</b>	Verkehrsdaten			<a href="https://www.bast.de/BASt_2017/DE/Statistik/Verkehrsdaten/Verkehrsbarometer.html">https://www.bast.de/BASt_2017/DE/Statistik/Verkehrsdaten/Verkehrsbarometer.html</a>
<b>RMWI und In-fas</b>	Datensammlungen auch lokal zu verschiedenen Aspekten			<a href="https://www.bmwi.de/Redaktion/DE/Coronavirus/corona-datenplattform.html">https://www.bmwi.de/Redaktion/DE/Coronavirus/corona-datenplattform.html</a>

Tabelle 1

### **3.1. Messdaten der Corona-Epidemie in Deutschland:**

Als Kandidaten für Betroffenheitsvariablen bieten sich zunächst Daten an, die direkt an die Intensität der Epidemie gekoppelt sind, wie die Anzahl der (Neu-) Infektionen, der Hospitalisationen oder Verstorbenen, ferner daraus abgeleitete KPIs wie die 7-Tage Inzidenz pro 100.000 Einwohner oder der R-Wert. Eine ausgezeichnete Quelle für solche Daten sind die Veröffentlichungen des RKI.<sup>5</sup> Detaildaten zu Fallzahlen sind im „COVID-19 Datenhub“ zusammengefasst. Qualität und Aktualität sind abhängig von der Datenerfassung und Meldung durch die Gesundheitsämter. Die Veröffentlichung ist tagesaktuell, die Granularität umfasst die regionalen Ebenen Bundesland und Kreis (KGS5), sowie Altersgruppe und Geschlecht. Die Daten können direkt als csv-file heruntergeladen werden. Auf Grund der Granularität übersteigt die Zeilenzahl mittlerweile die Möglichkeiten von Excel. Eine detaillierte Datensatzbeschreibung findet sich unter dem in der Tabelle 1 angegebenen Link.

Es empfiehlt sich die Messdaten von den Bevölkerungsgrößen zu dekorrelieren. Dazu sollten Bevölkerungszahlen nach Destatis/Genesis<sup>6</sup> angereichert werden und die Messwerte per 100.000 Einwohner indiziert werden, um z.B. die für Entscheidungsprozesse wichtige 7-Tages Inzidenz zu gewinnen. Als Maß für den Trend der Pandemie kann ebenfalls der R-Wert ermittelt werden. Die so gewonnenen Kennzahlen auf Kreisebene (KGS5) beschreiben die lokale Intensität der Pandemie und müssen geeignet geclustert werden, um ein zu großes Anschwellen der Menge der Datensätze zu verhindern. Die lokale Granularität spielt hierbei keine Rolle, da in der Regel ein Risiko genau einem Kreis zugeordnet werden kann. Wir geben im Anhang (Kapitel 7.2) ein SAS bzw. Python-Programm an, das diese Kennzahlen berechnet. Das Ergebnis dieses Programms ist eine Datei mit Kennzahlen und Gültigkeitszeitscheiben.

Die Basisseite des RKI (s.o.) verweist ebenfalls auf Messdaten zu COVID-19-Tests und zum Impfmonitor.

Der Umgang mit den Messdaten aus der Pandemie weist einige Besonderheiten auf: schon während der Pandemie zeigte sich, dass mit längerer Dauer der Pandemie und steigender Impfquote die 7-Tages-Inzidenz unterschiedlich interpretiert wurde. Dies führt schon in der Modellierung zu Schwierigkeiten bzw. sollte bei Anwendung der aus den 2020er Daten gewonnenen Modellergebnisse auf das Ist 2021 zu Abweichungen im Backtest führen. Ein Ausweg könnte eine Modellierung der Interaktion von Inzidenz und Pandemiedauer bzw. Impfquote sein. Als stabiler könnten sich die Messzahlen zur Hospitalisation („Lage auf den Intensivstationen“)

---

<sup>5</sup> Eine Übersicht über die gesamten Veröffentlichungen des RKI zum COVID-19-Virus findet sich unter [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/nCoV.html](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/nCoV.html) [abgerufen 10.12.2021]. Diese Quelle bietet neben der 7-Tages-Inzidenz die für die Modellierung geeignete Granularität.

<sup>6</sup> Vergl. Bevölkerungsstatistiken nach Kreis etc., z.B. <https://www-genesis.destatis.de/genesis//online?operation=table&code=12411-0015&bypass=true&levelindex=0&levelid=1640783578163#ab-readcrumb>, wird in diesem Format dauernd gepflegt.

erweisen, welche auf den Internetseiten der DIVI zu finden sind (Link, vergl. obige Tabelle). Je nach Granularität sind auch hier indizierte Kennzahlen zu bilden, etwa „Anteil der COVID-19-Patienten“ an der Gesamtbelegung oder Anteil freier Betten am gemeldeten Gesamtbestand. Die sehr detaillierte Segmentierung der RKI-Messzahlen bis auf KGS5-Ebene führt nicht zu einem Anwachsen der Anzahl der Modellierungsdatensätze, da jedes einzelne Risiko örtlich definiert ist. Allerdings sollte bei Verwendung mehrerer i.a. stark korrelierter Kennzahlen die Clusterung u.U. mit Koordination der Clustergrenze sorgfältig geplant werden, was eine Begrenzung der nötigen Teilung von Risikozeitscheiben ermöglicht.

Alle direkt auf die Pandemie bezogenen Kennzahlen haben den Nachteil, dass sie keine direkte Extrapolation auf die „Zeit nach Corona“ bzw. das „New Normal“ ermöglichen, da diese Zeit gerade durch das Absinken der Messwerte auf einen unauffälligen Durchschnitt definiert sein wird. Es ist auch davon auszugehen, dass dann auch der zurzeit enge Takt der Aktualisierung aufgegeben werden wird.

Für den ersten Schritt der Normalisierung der Messwerte der Pandemiejahre, sollten die direkten Messzahlen gute Dienste leisten. Allerdings sind Saisoneffekte sorgfältig zu trennen. Ein Beispiel für einen möglichen Saisoneffekt ist die Überlagerung des 2. Lockdowns 2020 mit der Weihnachtszeit. Eine Trennung wäre durch Übertragung der groben zeitlichen Struktur auf die Vorjahre möglich. Besondere Aufmerksamkeit erfordern nicht regelmäßige Saisoneffekte, wie z.B. die Frostperiode Februar 2021 während der Lockdownzeit Q1 2021.

### **3.2. Staatliche Maßnahmen**

Im Allgemeinen ist davon auszugehen, dass nicht die gemessenen und zeitnah veröffentlichten Pandemie-Kennzahlen das Verhalten der Risiken bzw. der Kunden bestimmt haben, sondern die daraus abgeleiteten Maßnahmen, Ge- und Verbote. Das Portfolio dieser Maßnahmen in Deutschland war gekennzeichnet durch eine wechselnde Mischung föderaler und zentraler Regelungen, zum Teil noch durch lokale Durchführungsbestimmungen ergänzt.

Eine einfache Bewertung des Maßnahmenkatalogs im Zeitverlauf bietet der Stringenzindex der Website "Our world in data".<sup>7</sup> Diese Datensammlung bietet auf Ebene des Nationalstaats – im Vergleich mit anderen Staaten – die Messwerte der Infektionen, Verstorbenen, Hospitalisierten und Intensivbetreuten, Tests, davon positive, sowie den Fortschritt der Impfkampagne auch als Index per 1 Million Einwohner.

---

<sup>7</sup> Eine ausführliche Beschreibung findet sich auf <https://github.com/owid/covid-19-data/tree/master/public/data>.

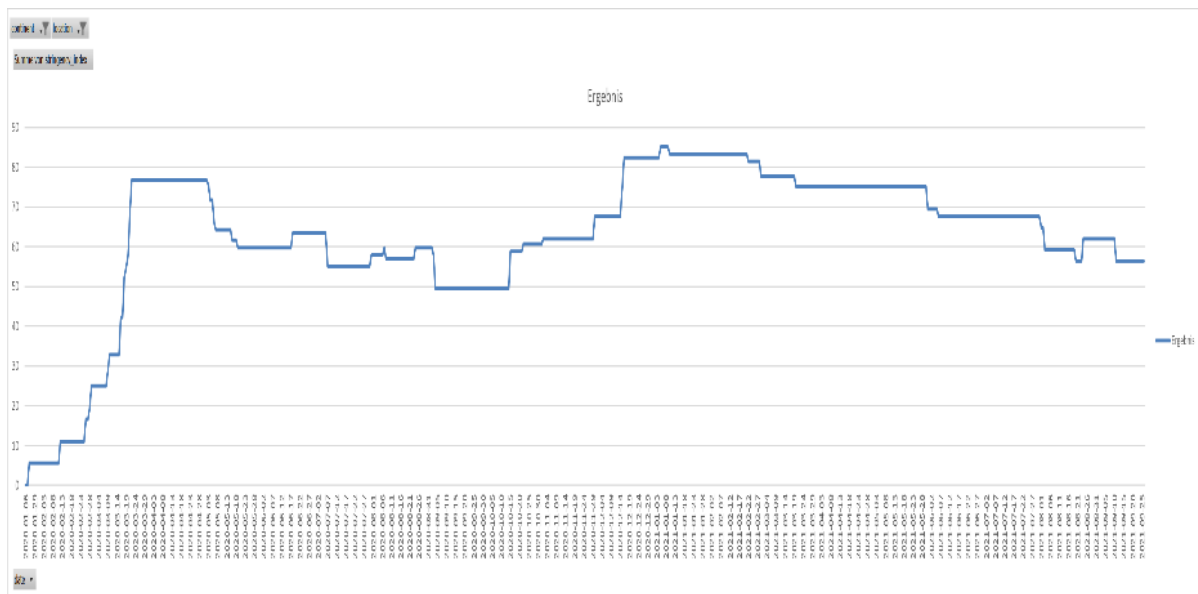


Abbildung 1: Die Maßnahmen der Regierungen, durch Einschränkungen eine Begrenzung der Pandemie zu erreichen, sind in einem „Stringenzindex“ zusammengefasst.

Die Abbildung 1 zeigt den Stringenzindex für Deutschland in seinem Verlauf in 2020 bis September 2021. Auch dieser Index sollte unter Beachtung des Zusammenhangs und der Datumsgrenzen anderer KPIs in wenige Klassen geclustert werden. Eine Entsprechung auf Bundeslandebene existiert hier nicht, so dass lokale Unterschiede in den Lockdownmaßnahmen nicht berücksichtigt werden können.

Die am internationalen Vergleich ausgerichtete Methodik der Ableitung des Index verspricht eine angemessene Vereinfachung zur Bewertung des Maßnahmenkatalogs. Es fällt aber auf, dass die Lockerungsmaßnahmen von August / September 2021 sich noch nicht in einer deutlichen Senkung des Index niederschlagen. Außerdem bietet der Index keinen Ausblick auf eine Zeit „nach der Pandemie“, die genau durch den Wegfall der beschriebenen Maßnahmen geprägt sein wird.

### 3.3. Allgemeine Aktivitätsparameter

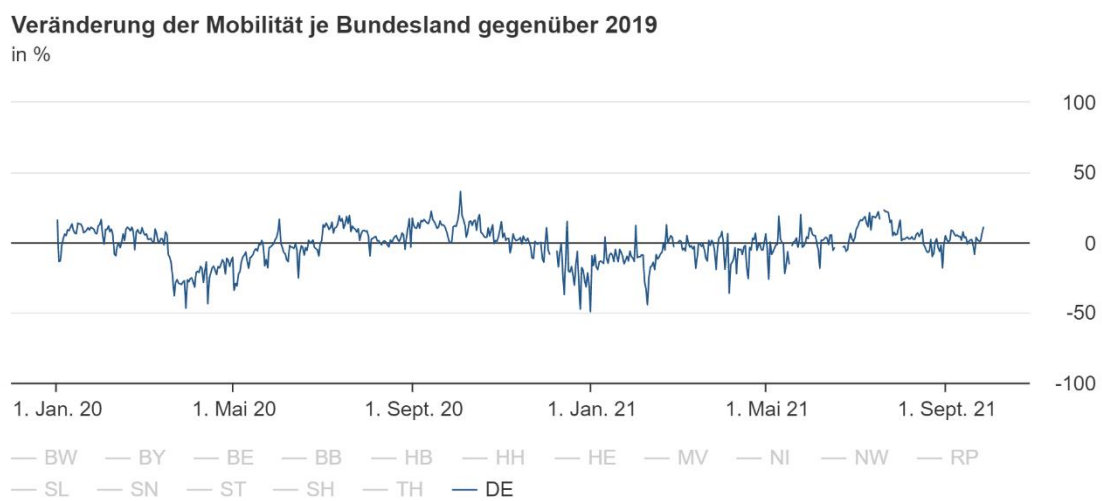
Die bislang diskutierten Datenquellen zeigen zwar einen engen Bezug zum Pandemiegeschehen bzw. den staatlichen Maßnahmen auf und versprechen daher einen Nutzen in der Bereinigung der Pandemieeffekte in den Daten der Statistikjahre 2020 und 2021, sie sind aber schwierig für Prognosezwecke in die Zukunft fortzuschreiben, wenn man sich nicht auf Szenarien einlassen will. Es lässt sich aus diesen Kennzahlen auch kein „New Normal“ ableiten.

Dies ist bei der dritten Art von Kennzahlen, die als Betroffenheitsvariablen dienen können, eher möglich. Da wir einen direkten Zusammenhang zwischen Lockdownmaßnahmen und gewissen betriebs- und volkswirtschaftlichen Kennzahlen unterstellen, können diese auf ihre Eignung als Betroffenheitsvariablen untersucht werden. Am naheliegendsten ist der Zusammenhang zwischen Lockdownmaßnahmen wie Betriebsschließungen und Reiseverboten mit Produktions-, Transport- und Mobilitätsindizes.



### 3.3.1. Mobilitätsdaten

Mobilitätsdaten werden im Allgemeinen aus Mobilfunkdaten gewonnen, d.h. sie zeigen nicht die Aktivitäten der Bewohner einer bestimmten Region, sondern die Aktivitäten von Menschen in dieser Region.<sup>8</sup> Der Grunddatensatz liefert die Veränderung des täglichen Mobilitätsgeschehens zum Vorjahr. Die Daten stehen auf Tagesbasis auf Bundes- wie Landesebene zur Verfügung und lassen sich direkt im csv-Format herunterladen. Ab dem 15. Juli 2020 wurde die Datenaufbereitung dahingehend geändert, dass Verfälschungen auf Grund saisonaler Effekte (Ferien- und Feiertage) herausgerechnet wurden. Für eine Anwendung ist sicherlich die Korrelation mit Nicht-Pandemie Effekten (insbesondere der Wetterlage) zu beachten.



Die Mobilitätszahlen – in der Graphik für Deutschland im Vergleich zu 2019 – lassen klar die Lockdown-Phasen, insbesondere den 1. Lockdown im Frühjahr 2020 erkennen. Während für die Kraftfahrtversicherung ein direkter Zusammenhang mit dem Risiko auf Grund der bekannten Tarifvariablen „Fahrleistung“ auf der Hand liegt, kann für andere Risiken nur auf einen indirekten Zusammenhang zwischen Mobilität und allgemeiner wirtschaftlicher Aktivität geschlossen werden.

Es ist zu beachten, dass die Mobilität bei lokaler Betrachtung sich auf den Ort der Mobilität und nicht den Zulassungsort des Fahrzeugs bezieht, sodass insbesondere Urlaubsverkehr – in Pandemiezeiten bevorzugt im Inland – scheinbare Aktivitäten vortäuschen kann. Das Problem wurde mittlerweile erkannt und angegangen.<sup>9</sup>

<sup>8</sup> Eine leicht zugängliche Datenbasis bietet das Statistische Bundesamt auf <https://www.destatis.de/DE/Themen/Querschnitt/Corona/Gesellschaft/kontextinformationen-gesellschaft.html#Mobi>. Ein umfangreicher Datensatz kann u.a. von Google oder Apple bezogen werden.

<sup>9</sup> Für weitere Informationen siehe: <https://www.destatis.de/DE/Service/EXDAT/Datensaetze/mobilitaetsindikatoren-mobilfunkdaten.html#Hinweis>

Allerdings lassen sich i.A. die Mobilitätsdaten nicht vor das Jahr 2020 zurückverfolgen, um längerfristige Zeitreihen zu erhalten. Insbesondere für die Kraftfahrtversicherung ergeben sich aus der Auswertung von Mobilitätsdaten umfangreiche Forschungsfelder in Zusammenhang mit Prognosen und Frühwarnalgorithmen, sodass hier auf eine tiefergehende Beschreibung verzichtet wird.

Insgesamt haben die Veröffentlichungen von Mobilitätsdaten durch Destatis noch „experimentellen“ Charakter, sodass die weitere Entwicklung sicherlich neue Strukturen und Erkenntnisse bringen wird. Granularere Daten auf Kreisebene finden sich z.B. auf Github<sup>10</sup> oder in den „Mobilitätsberichten zur Coronakrise“ von Google<sup>11</sup>. Das Thema „Mobilitätsdaten auf der Basis von Mobilfunkdaten“ entwickelt sich derzeit sehr dynamisch. Für konkrete Analysen ist daher der aktuelle Stand zu prüfen und eine oder mehrere Quellen geeignet auszuwählen.

### 3.3.2. *Allgemeinere Aktivitätsdaten*

Das statistische Bundesamt (Destatis) hat aus seinem Datenvorrat eine Reihe von volks- bzw. betriebswirtschaftlichen sowie gesellschaftlichen Indikatoren zusammengestellt, die einen Einfluss durch die Corona-Krise erfahren haben. Diese Zeitreihen haben oft den Vorteil, dass sie aus den Daten von Destatis/Genesis in die Vergangenheit weiterverfolgt und ggfls. in Prognosen mit Zukunftsschätzungen fortgeführt werden können. Eine Zusammenfassung liefert das Statistik Dossier „Daten zur COVID-19 Pandemie“<sup>12</sup>. Mit diesen umfangreichen Datenquellen verlassen wir den Bereich der speziell auf die COVID-19-Pandemie bezogenen Daten, da die hier vorgestellten KPIs allgemeine wirtschaftliche Trends beschreiben, deren Verlauf und Schwankungen über die Zeit beschrieben werden können. Die Menge der möglichen Datensätze ist so umfangreich, dass hier nur ein Hinweis auf die Quellen erfolgt. Bei der Nutzung dieser Daten ist insbesondere zu prüfen, ob eine direkte Korrelation mit dem zu untersuchenden Risiko vorliegt und ob eine valide Prognose in die Zukunft möglich ist.

### 3.3.3. *Speziell: Arbeitsmarktstatistiken*

Unter den allgemeinen Aktivitätsstatistiken seien noch die Zeitreihen der Bundesagentur für Arbeit erwähnt.<sup>13</sup> Sicher ist z.B. die Entwicklung der Kurzarbeit ein Indikator für den ersten Lockdown.

Ein Vorteil dieser Statistiken ist ebenfalls der längerfristige Zeithorizont. Allerdings ist auch hier die Vielfalt so groß, dass eine eigene Untersuchung nötig wäre, um die Eignung der möglichen Variablen als Betroffenheitsvariablen zu prüfen.

---

<sup>10</sup> Siehe [https://github.com/rocs-org/covid-mobility-data/blob/main/data/mobility\\_change/mobility\\_change\\_kreise.csv](https://github.com/rocs-org/covid-mobility-data/blob/main/data/mobility_change/mobility_change_kreise.csv).

<sup>11</sup> Siehe <https://www.google.com/covid19/mobility/>.

<sup>12</sup> Siehe [https://www.destatis.de/DE/Themen/Querschnitt/Corona/Downloads/dossier-covid-19.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Themen/Querschnitt/Corona/Downloads/dossier-covid-19.pdf?__blob=publicationFile), zuletzt aktualisiert Ausgabe 08/2021 vom 15.12.2021

<sup>13</sup> Siehe: <https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Themen-im-Fokus/Corona/Corona-Nav.html>, zuletzt aufgerufen 26.12.2021

## **4. Nutzung der Betroffenheitsvariablen in der Tarifierung**

Die Nutzung von Betroffenheitsvariablen ist für die Tarifierung insbesondere dann sinnvoll, wenn damit die Vergangenheit besser erklärt und das spezifische Prognoseproblem gelöst werden kann. In diesem Kapitel wird deshalb zuerst darauf eingegangen, wie diese Variablen bei der Modellierung verwendet werden können. Im Anschluss beschreiben wir, wie eine Prognose mit Hilfe dieser Information durchgeführt werden kann.

### **4.1. Modellierung mit Betroffenheitsvariablen**

Grundsätzlich lassen sich Betroffenheitsvariablen in derselben Form in der Modellierung verwenden wie andere Kovariablen auch. In der Praxis hat sich dabei bewährt, mögliche Variablen vorab univariat zu betrachten. Gerade bei den Betroffenheitsvariablen ist eine gewisse Vorsicht in der Interpretation der Ergebnisse angebracht, da vermutlich Abhängigkeiten mit anderen Kovariablen (nicht nur im Sinne von Interaktionen) vorliegen. Genau das ist auch ein wesentlicher Grund für ihre Einführung. Der Nutzen liegt in ihrer multivariaten Verwendung.

Als gängiges und transparentes Framework verwenden wir beispielhaft Generalisierte Additive Modelle (GAM)<sup>14</sup>. Daran soll erläutert werden, worauf es bei der Anwendung ankommt und welche Stolpersteine zu beachten sind. Auf die Besonderheiten anderer Verfahren wird am Ende kurz eingegangen.

Für die Diskussion wird angenommen, dass eine geeignete Verteilung und Linkfunktion für das GAM gewählt wurden. Wir konzentrieren uns also auf die Wahl des geeigneten linearen Prädiktors. Grundsätzlich stehen uns hierbei alle Mittel zur Variablenauswahl und zur Bewertung der Modellgüte zur Verfügung, die uns ein GAM bietet. Diese lassen sich ohne Weiteres auch auf die Modellierung mit Betroffenheitsvariablen anwenden. Allerdings treten mit Betroffenheitsvariablen einige spezifische Effekte auf.

Besonders zu beachten sind saisonale Effekte. Da die Betroffenheitsvariablen selbst einen Zeitbezug haben, muss in der Modellierung ausgeschlossen werden, dass durch sie temporäre Effekte abgebildet werden, die keinen COVID-19-Bezug haben, weil sie regelmäßig zyklisch z.B. in jedem Herbst auftreten. Der stringenste Weg dafür wäre die explizite Modellierung der saisonalen Effekte durch entsprechende Kovariablen. Da diese Vorgehensweise zu einem weiter erhöhten Datenvolumen und Aufwänden in der Datenaufbereitung führen kann, sollten Tarifierungsaktuar\*innen aber abwägen, inwieweit das für den konkreten Anwendungsfall relevant ist. Eine Überprüfung ist zumindest im Nachhinein stets anzuraten.

Ob Betroffenheitsvariablen in Interaktionen eine Rolle spielen, muss ebenfalls an den konkreten Daten überprüft werden. Die gewonnene Genauigkeit ist dabei auch immer gegen die Erhöhung der Modellkomplexität abzuwägen. Gerade für Inter-

---

<sup>14</sup> Hinweis: GAMs erweitern das Framework Generalisierter Linearer Modelle um die Möglichkeit funktionaler Terme für die Kovariablen. Wir beschränken uns hier gedanklich auf parametrische Terme.

aktionen sollte zudem überlegt werden, inwiefern ein solcher Effekt durch Erklärungsansätze zumindest plausibilisiert werden kann. Sollte das gelingen, erhöht sich neben der beschreibenden auch die prognostische Qualität des Modells.

Aufgrund der Verwendung der Betroffenheitsvariablen als Prognoseinstrument ist es notwendig, nicht nur bekannte bzw. aufgetretene Werte dieser Variablen bei der Prognose verwenden zu können. Daher müssen Inter- oder Extrapolationen dieser Werte vorgenommen werden, um diese in der Modellprognose zu verwenden. Das stellt Anforderungen an die Stetigkeit oder sogar die Monotonie der Modellierung. Im GAM-Framework bietet sich dafür die Modellierung über geeignet gewählte Polynome an. Inwiefern Monotonie-Eigenschaften tatsächlich realisiert werden müssen, sollte natürlich überprüft werden.

Ein linearer Prädiktor  $\eta_i$  des  $i$ -ten Satzes setzt sich für ein Modell mit der Betroffenheitsvariable  $b_i$  also beispielsweise so zusammen:

$$\eta_i = \sum_j \beta_j x_{ij} + \sum_k \gamma_k f_k(b_i) + \sum_l \delta_l g_l(s_i)$$

Dabei ist  $\sum_j \beta_j x_{ij}$  der lineare Prädiktor ohne die Berücksichtigung von Betroffenheitsvariablen,  $\sum_k \gamma_k f_k(b_i)$  modelliert den eigentlichen Betroffenheitseffekt und  $\sum_l \delta_l g_l(s_i)$  den Effekt der saisonalen Kovariable  $s_i$ . In einem GLM lassen sich diese Effekte entsprechend separieren und für die Prognose weiterverwenden.

Um die Wirkung des Einsatzes der Betroffenheitsvariablen auf das restliche Modell einschätzen zu können, schlagen wir vor, das finale Modell auch ohne Betroffenheitsvariablen anzupassen. Die resultierenden  $\beta$ -Schätzer und Prädiktoren der so entstandenen Modelle lassen sich recht einfach miteinander vergleichen. Daraus lässt sich auch erkennen, wie stark die Wirkung der Corona-Effekte z.B. auf einzelne Segmente ausfällt.

Natürlich ist die Verwendung der Betroffenheitsvariablen auch in anderen Verfahrensklassen möglich. Das gilt auch für die Verfahren des maschinellen Lernens. Es sollte aber vor der Verwendung abgeklärt werden, inwiefern sich die Anforderungen an die Beherrschbarkeit der Prognose mit dem jeweiligen Verfahren realisieren lassen. Das gilt zum Beispiel für die Anforderung an die Stetigkeit, wie wir sie oben erläutert haben. Da das Maß an Transparenz gerade in solchen Details oft nicht gewährleistet ist, sollte dies spätestens bei der Anwendung eines konkreten Verfahrens geprüft werden.

#### **4.2. Prognose mit Betroffenheitsvariablen**

Ist ein geeignetes Modell inklusive der notwendigen Betroffenheitsvariablen erstellt, kann die Prognose auf Basis der Fortschreibung dieser Variablen erfolgen. Dazu genügt es, den konkreten, prognostizierten Wert der jeweiligen Betroffenheitsvariable im Modell zu verwenden. Sollte dieser außerhalb des im Datensatz enthaltenen Wertebereichs liegen, muss zudem eine geeignete Extrapolation für die entsprechenden Bestandteile des Prädiktors erfolgen. Für diesen Fall sollte aber insbesondere die Frage beantwortet werden, inwiefern es durch die Fortschreibung zu einer Überbewertung der modellierten Effekte kommen kann.

Durch die Verwendung der hier beschriebenen Methodik wird in der Regel sichergestellt, dass der Einfluss auf die Wirkung anderer Kovariablen bereits angemessen abgebildet ist. Eine Besonderheit spielen dabei allerdings Interaktionen mit verwendeten Betroffenheitsvariablen. Während das bei üblichen Kovariablen nur für Extrapolationen zu Schwierigkeiten führen wird, ist von der Modellierung von Interaktionen mit saisonalen Variablen abzuraten.

Es sei an dieser Stelle nochmals darauf hingewiesen, dass eine sachgerechte, angemessene Fortschreibung der Betroffenheitsvariablen die Grundlage für eine gute Prognose darstellt. Auf diese ist ein entsprechend hohes Augenmerk zu legen. Gegebenenfalls kann es sinnvoll sein, die Wirkung der Prognose durch Szenarioberechnungen zu unterlegen, um ein besseres Gefühl für die Wirkung einzelner Annahmen zu erhalten. Eine sinnvolle Variante wäre ein Szenario mit der Rückkehr auf das „Old Normal“, da man hier anhand der Erfahrung oder mit Hilfe eines Hold-Out-Samples die erzielte Wirkung gut plausibilisieren kann. Falls eine direkte Prognose durch Fortschreibung der Betroffenheitsvariablen zu unsicher erscheint und das gewählte Modell die dafür nötige Transparenz ermöglicht, kann im ersten Schritt auch eine explizite Rekalibrierung des Modells durch Annahme einer zukünftigen Verteilung vorgenommen werden. Diese Vorgehensweise entspricht dem Standardvorgehen, nach der Modellierung und vor Anwendung eines Modells das Basisniveau um in den Daten wahrgenommene, aber für das Risiko als untypisch angesehene Effekte zu bereinigen.

Sollte die Wirkung der Betroffenheitsmodellierung aufgrund der Modellklassenwahl oder wegen der Ausgestaltung des Modells nicht unmittelbar ersichtlich sein, empfiehlt sich zudem eine segmentierte Analyse der Auswirkungen. Damit kann vermieden werden, dass in einzelnen Bestandsgruppen unplausible oder übermäßig große Effekte auftreten.

Grundsätzlich sollten die Ergebnisse der Prognose mit Betroffenheitsvariablen auf dem gesamten Bestand mit Hilfe einer separaten, pauschalen Schätzung plausibilisiert werden. Eine unerwartete, ungewöhnliche Veränderung sollte zur Ursachen-suche und, falls notwendig, zur Anpassung der Modellierung führen.

Da die Prognose letztendlich sehr einfach aus der Fortschreibung abzuleiten ist, ist eine sorgsame Wahl und ein sachgerechter Umgang mit den genutzten Betroffenheitsvariablen entscheidend für den erfolgreichen Einsatz der Methode.

## **5. Use Case**

Der Use Case soll exemplarisch ein mögliches Vorgehen bei der Betroffenheits-Modellierung anhand eines Kraftfahrt-Haftpflicht-Portfolios (Pkw) demonstrieren. Dabei liegt das Hauptaugenmerk auf der in vorangegangenen Kapiteln eingeführten Methodik und weniger auf exakt quantifizierten versicherungstechnischen Effekten, die aus der COVID-19-Pandemie resultieren. Der Use Case wurde in Zusammenarbeit mit Gen Re Consulting unter anderem auf Basis eines Teilsegmentes<sup>15</sup> des Gen Re Kraftfahrt-Datenpools (bestehend aus insgesamt 26 Unternehmen mit Pkw-Marktanteil von ca. 31% in 2020) erstellt.

### **5.1. Datengrundlagen**

Die Datenbasis ergibt sich im Wesentlichen aus zwei Datenquellen. Einerseits aus einer klassischen Meldejahres-Statistik der Kraftfahrt-Haftpflichtversicherung des präpandemischen Statistikjahres 2019 und des von Corona betroffenen Statistikjahres 2020. Andererseits werden zusätzlich eine oder mehrere geeignete Betroffenheitsvariablen zu COVID-19 benötigt. Hierzu wurden Mobilitätsdaten des COVID-19 Mobility Projektes, die die tägliche Änderung des allgemeinen Mobilitätsverhaltens im Jahr 2020 gegenüber des Vorjahres 2019 auf Kreisebene (KGS5) darstellen, herangezogen.

Vor der Aufbereitung für die Betroffenheitsmodellierung umfasst die Datenbasis für die Statistikjahre 2019 und 2020 jeweils ca. 1 Mio. Datensätze, was einem Exposure von ca. 700.000 Jahreseinheiten pro Statistikjahr entspricht.

Die zur Modellierung genutzten Kovariablen entsprechen den Risikomerkmale der unverbindlichen Netto-Risikokalkulation für Pkw (WKZ 112) in der Kraftfahrzeug-Haftpflichtversicherung des Gesamtverbands der Deutschen Versicherungswirtschaft e. V. und umfassen die folgenden Datenfelder:

- SF-Klasse
- Typklasse
- Tarifgruppe
- Jährliche km-Fahrleistung
- Nutzerkreis
- Differenziertes Nutzeralter
- Fahrzeugalter bei Erwerb
- Regionalklasse
- Fahrzeughalter

---

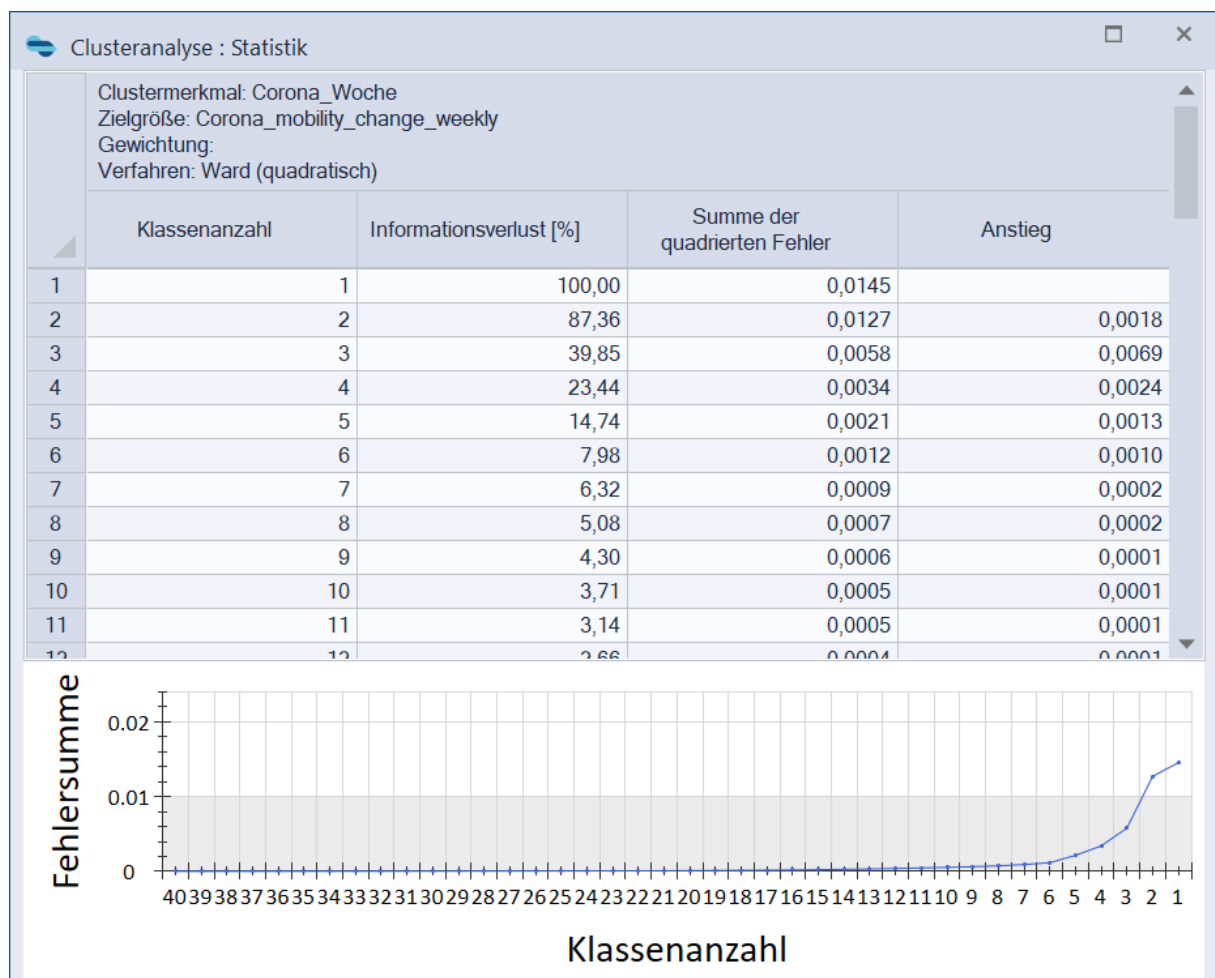
<sup>15</sup> Die konkreten Auswahlkriterien werden nicht offengelegt, um die inhaltliche Weiterverwendung der Ergebnisse zu verunmöglichen.

Des Weiteren enthält die Datenbasis Informationen zu Beginn und Ablauf des Vertrages, den Kreisgemeindeschlüssel zum Wohnort des Versicherungsnehmers sowie Schadenanzahl und -eintrittsdatum von gemeldeten Sachschäden zum jeweiligen Vertragsdatensatz.

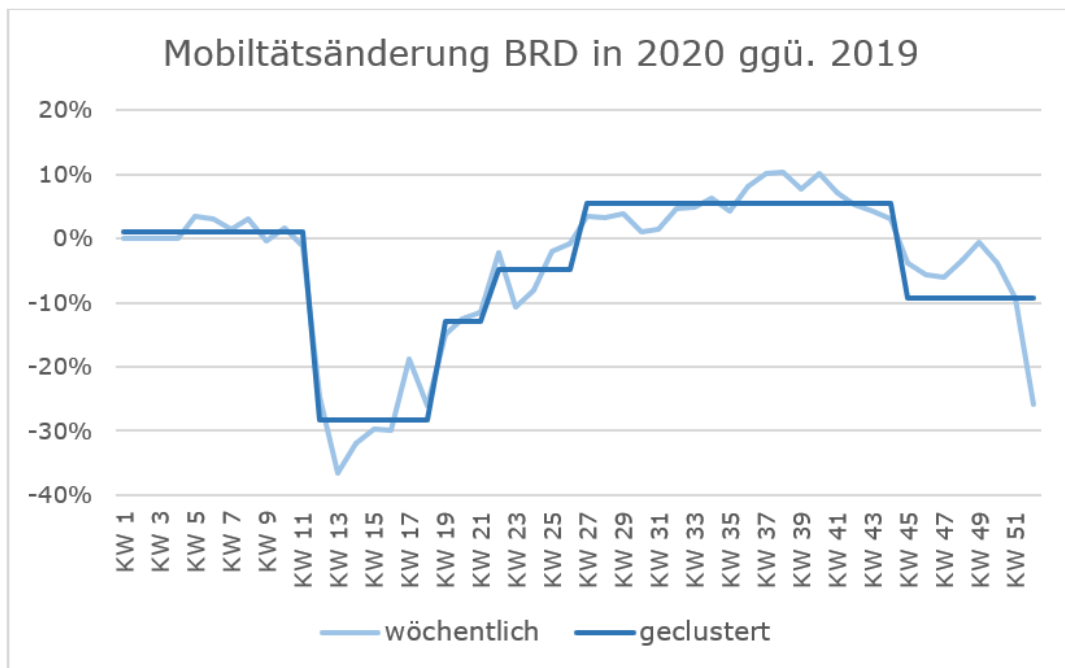
### 5.1.1. Datenaufbereitung - Betroffenheitsvariable

Die granular auf Kreisebene vorliegenden tagesaktuellen Änderungen der Mobilität im Jahr 2020 gegenüber des Basisjahres 2019 müssen zur weiteren Kalkulation geeignet zusammengefasst und/oder bereinigt werden. Da eine manuelle Nachbearbeitung von Ferien-/Feiertagen je Bundesland recht aufwändig erscheint, wird an dieser Stelle darauf verzichtet. Außerdem gehen wir, um Wochentageeffekte zu vermeiden, von einer täglichen zu einer wöchentlichen Mobilitätsbetrachtung über.

Per Definition hat die Mobilitätsänderung für das Jahr 2019 den konstanten Wert 0, für das Jahr 2020 ermitteln wir mittels des Ward'schen Clusteralgorithmus geeignete Zusammenfassungen der einzelnen Kalenderwochen zu größeren Zeitscheiben und mitteln die Mobilitätsänderung auf diesen Zeitscheiben. Die Anzahl der Klassen wird anhand des Elbow-Kriteriums ermittelt und in diesem Fall wird die Betrachtung des Jahres 2020 auf sechs zusammenhängende, vom Kreis unabhängige Zeitabschnitte reduziert.



Die Clusterung führt neben der angestrebten Reduktion der Klassenanzahl auch zu einer Glättung einzelner Spitzen in der Mobilitätsänderung, die beispielsweise durch Feiertageffekte zustande kommen (z.B. Ostermontag: 2019 in KW 16, 2020 in KW 17). Insgesamt erscheint die Clusterung in der grafischen Analyse plausibel zu sein und spiegelt sowohl die Lockdown-Maßnahmen im Frühjahr und zum Jahresende 2020 als auch das leicht erhöhte Reiseaufkommen im Inland durch substituierte Auslandsreisen im Sommer wider.



Durch Aufteilung der KH-Bestandssätze mit dem zugehörigen Exposure auf die 6 Zeitscheiben und gleichzeitiger Allokation der Schäden anhand des Schadendatums in die jeweilige Zeitscheibe, entsteht die für die Modellierung der Betroffenheitsvariablen benötigte Datenbasis.

Analog wird die Statistik des Jahres 2019 bearbeitet und in die aus 2020 bekannten sechs Zeitscheiben aufgeteilt. Die Aufteilung des Jahres 2019 wird ausschließlich für die folgende Abgrenzung von saisonalen Effekten zur Betroffenheitsvariablen benötigt. Um sicher zu stellen, dass Feiertageffekte die saisonale Betrachtung nicht (übermäßig) verzerren, ist es sinnvoll die Zeitscheiben so zu gestalten, dass die relevantesten Feiertage in die jeweils gleiche Zeitscheibe fallen. Weitere schadenbeeinflussende saisonale Ereignisse, wie z. B. Frostperioden wurden im Rahmen dieses Use Cases nicht weiter berücksichtigt.

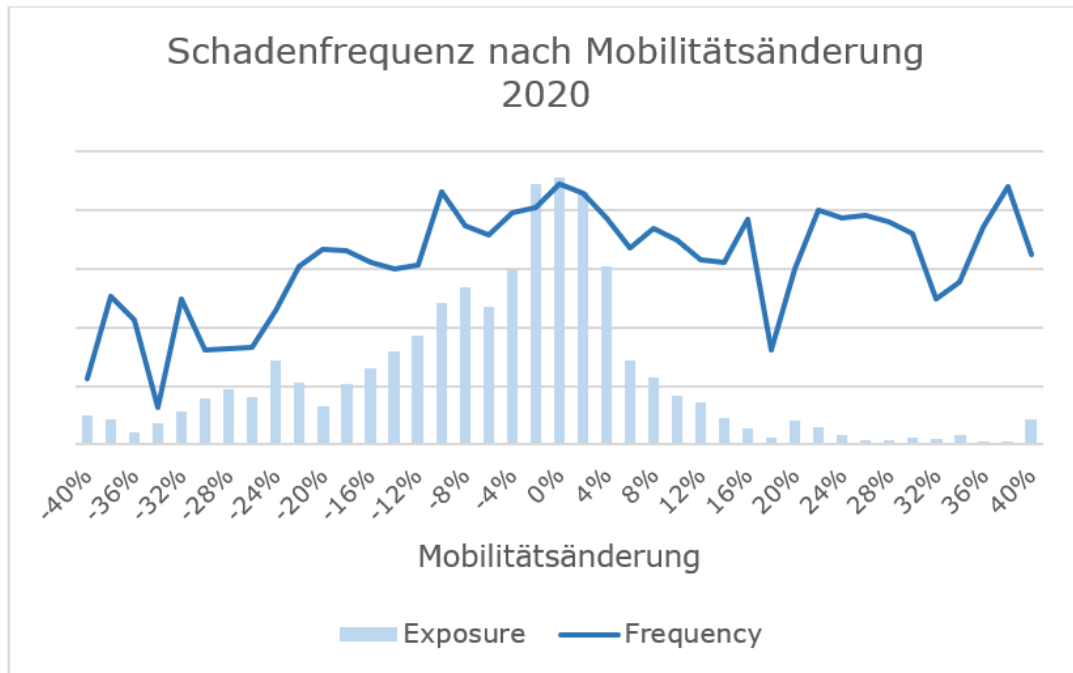
Die Datenbasis vervielfacht sich durch das beschriebene Verfahren von ursprünglich ca. 2 Mio. Datensätze auf insgesamt über 9 Mio. Datensätze bei selbstverständlich gleichbleibendem Exposure.

## 5.2. Modellierung & Prognose der Betroffenheitsvariable

Die gewählte Betroffenheitsvariable sollte vorab auf ihre Eignung geprüft werden. Unter anderem ist die angenommene Wirkung auf die zu modellierende Schadengröße zu verifizieren, in diesem Fall also der Effekt der Mobilitätsänderung auf die



Schadenfrequenz. Hierzu bieten sich für eine erste Analyse insbesondere grafische Methoden an.



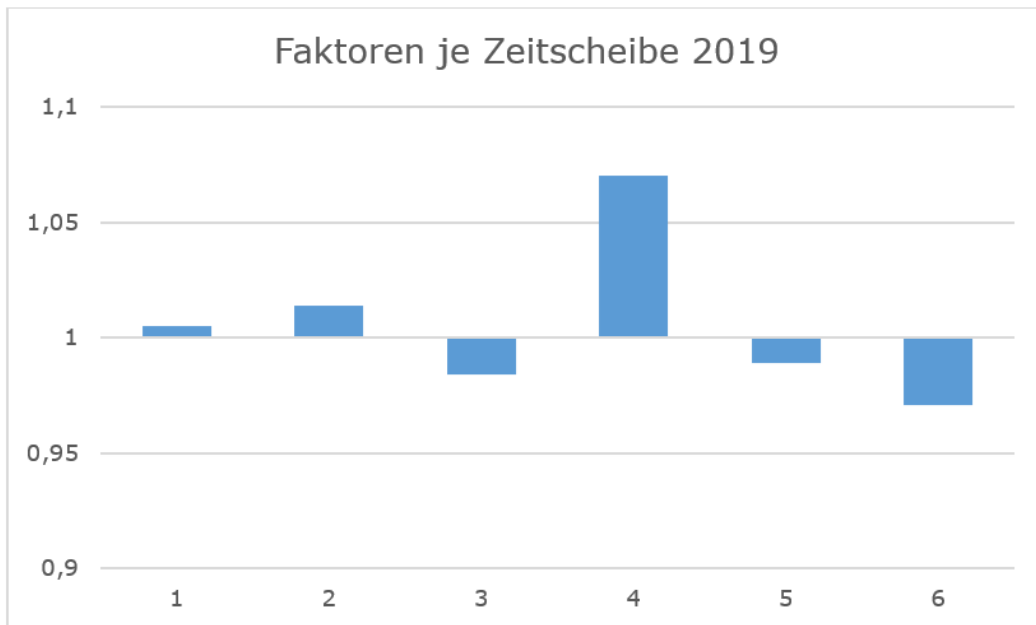
Deutlich zu erkennen in der univariaten Betrachtung ist die Abhängigkeit der Schadenfrequenz zur Mobilitätsänderung im negativen Bereich. Bei einer Mobilitätszunahme im Jahr 2020 gegenüber 2019 (positiver Bereich der Mobilitätsänderung) scheint der Einfluss nicht gegeben zu sein. Hier liegt die Vermutung nahe, dass es sich um regionale Besonderheiten im Corona-Jahr 2020 handelt, die keine oder kaum Auswirkungen auf das gesamte Pkw-Schadengeschehen haben. Beispielsweise seien hier mehr Fahrradmobilität oder auch veränderte Freizeitmobilität in touristischen Regionen (mehr Urlaub im Inland) genannt. Insbesondere werden die in Urlaubsregionen anfallende Schäden nicht dem Ort der erhöhten Mobilität (Schadenort), sondern dem Zulassungsort des Pkw zugeordnet. Für die weitere Modellierung nutzen wir deshalb zusätzlich eine bei 0% gekappte Mobilitätsänderung, nämlich  $\min(0\% ; \text{Mobilitätsänderung})$ . Damit ist gewährleistet, dass die Betroffenheitsvariable „gekappte Mobilitätsänderung“ in plausibler Weise, auch bei positiver Veränderung gegenüber 2019, fortgeschrieben werden kann.

### 5.3. Modellierung mit Betroffenheit

#### 5.3.1. Saisonale Effekte

Ziel ist es, den Einfluss zeitlicher Effekte, die nicht in Zusammenhang mit COVID-19 stehen, von der Betroffenheitsvariablen zu isolieren. Die oben beschriebene Datenaufbereitung ermöglicht es, saisonale Effekte in den sechs betrachteten Zeitscheiben im Jahr 2019 zu quantifizieren und anschließend auf das Jahr 2020 zu übertragen. Hierzu wird die Zielgröße Schadenfrequenz (mit Gewichtungsgroße Jahreseinheiten) mittels eines generalisierten additiven Modells (GAM) mit logarithmische Linkfunktion und Poisson-Verteilungsannahme mit den unter 5.1 beschriebenen Kovariablen zuzüglich der kategorialen Kovariablen „Zeitscheibe“

(Saisonalität) auf dem Statistikjahr 2019 modelliert. Die so ermittelten saisonalen Effekte werden in Form von Zu-/Abschlagsfaktoren auf den Jahresmittelwert der Schadenfrequenz zur späteren Verwendung gespeichert.



Im folgenden Schritt werden die ermittelten saisonalen Effekte als gegeben vorausgesetzt und als Offset im Modell verwendet. Dieses kaskadierende Vorgehen ist hier aufgrund der per Konstruktion hohen Abhängigkeit zwischen der Saisonvariablen „Zeitscheibe“ und der Betroffenheitsvariablen „Mobilitätsänderung“ erforderlich und weicht von der in **Fehler! Verweisquelle konnte nicht gefunden werden.** beschriebenen gemeinsamen Modellierung ab.

Eine mehrjährige Untersuchung von saisonalen Effekten findet aus Komplexitätsgründen im Rahmen des Use Cases nicht statt, sollte jedoch zur Stabilisierung und Verifizierung der Ergebnisse in Betracht gezogen werden.

### 5.3.2. Ergebnisse für die Betroffenheitsvariable

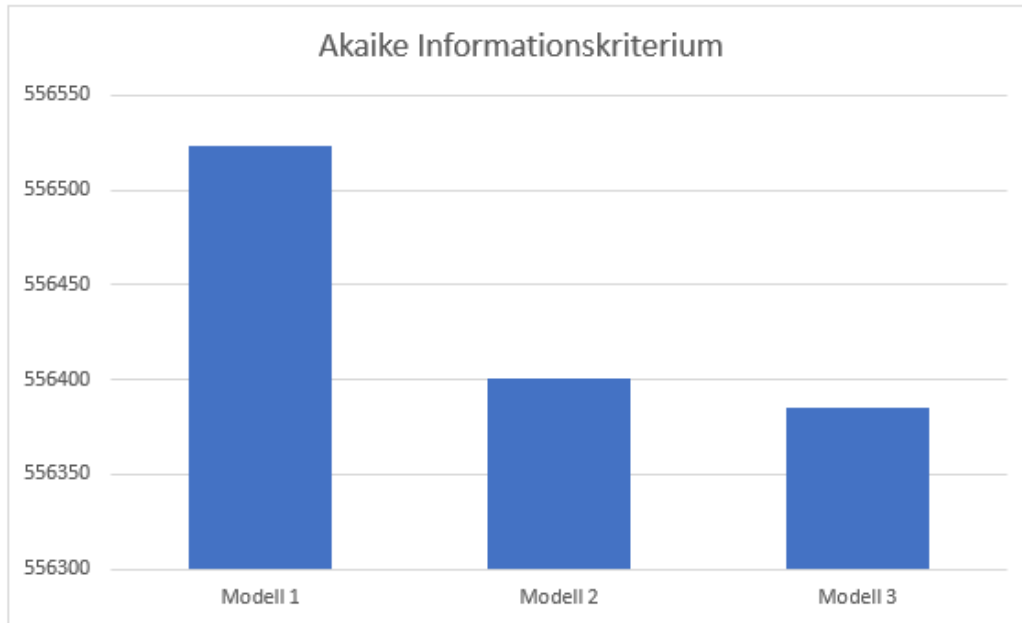
Mit den Vorbereitungen aus den vorangegangenen Abschnitten lässt sich nun die Betroffenheitsvariable „Mobilitätsänderung“ (bzw. „gekappte Mobilitätsänderung“) modellieren. Dazu passen wir wiederum ein generalisiertes additives Modell (GAM) mit logarithmischer Linkfunktion und Poisson-Verteilungsannahme an die Daten der Statistikjahre 2019 und 2020 an. Verwendet werden dazu die aus 5.1 bekannten Kovariablen, die Zielgröße Schadenfrequenz (Gewichtung: Jahreseinheiten), die saisonalen Effekte als Offset und als zusätzliche Kovariablen die (gekappte) Mobilitätsänderung sowie das Statistikjahr.

Nach einem Vergleich mehrerer Modellansätze für die Betroffenheitsvariable stellen wir hier drei vielversprechende Ansätze vor:

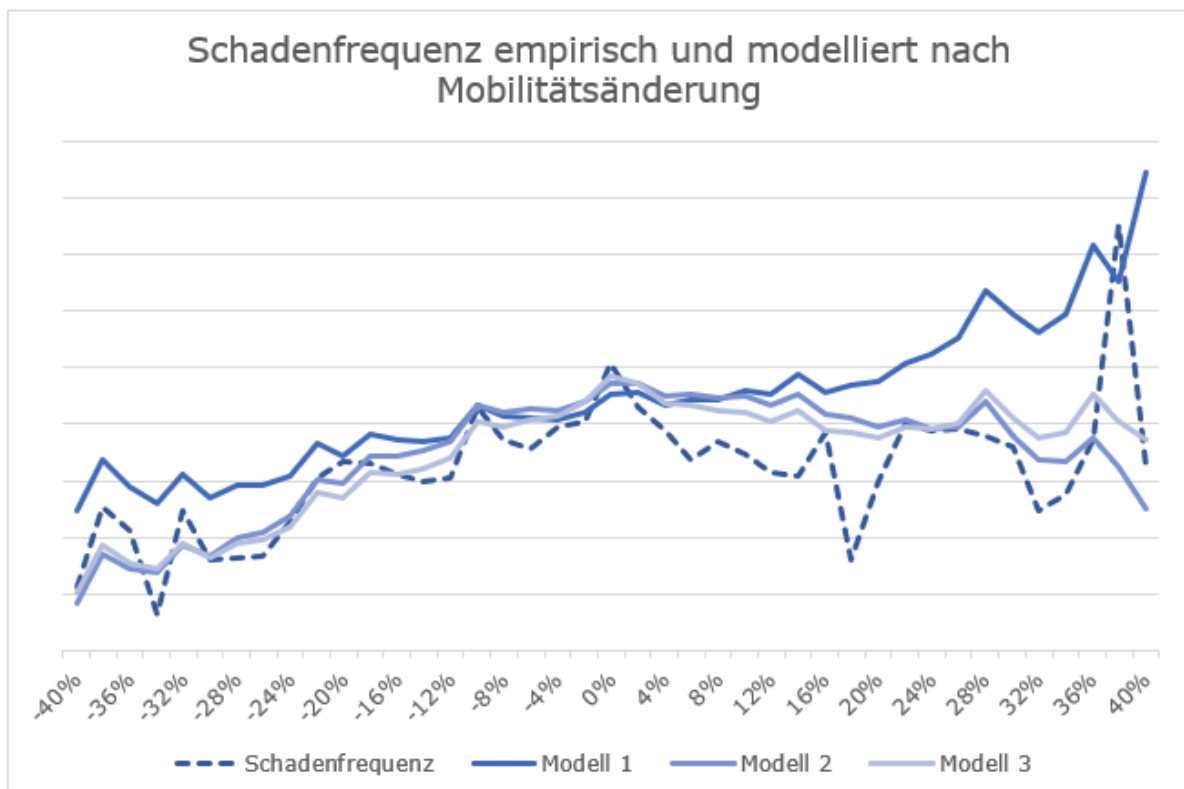
1. Die Kovariable „Mobilitätsänderung (ohne Kappung)“ wird linear im linearen Prädiktor modelliert.
2. Die Kovariable „Mobilitätsänderung (ohne Kappung)“ wird mittels Polynom 4. Grades im linearen Prädiktor modelliert.

3. Die Kovariable „Gekappte Mobilitätsänderung“ wird linear im linearen Prädiktor modelliert.

Der Vergleich der Modelle mittel Akaike-Informationskriterium zeigt deutlich, dass die Modelle 2 und 3 dem Modellansatz 1 überlegen sind Die Kappung der Mobilitätsänderung (Modell 3) zeigt sich sogar der Modellierung mittels polynomialer Struktur (Modell 2) überlegen.



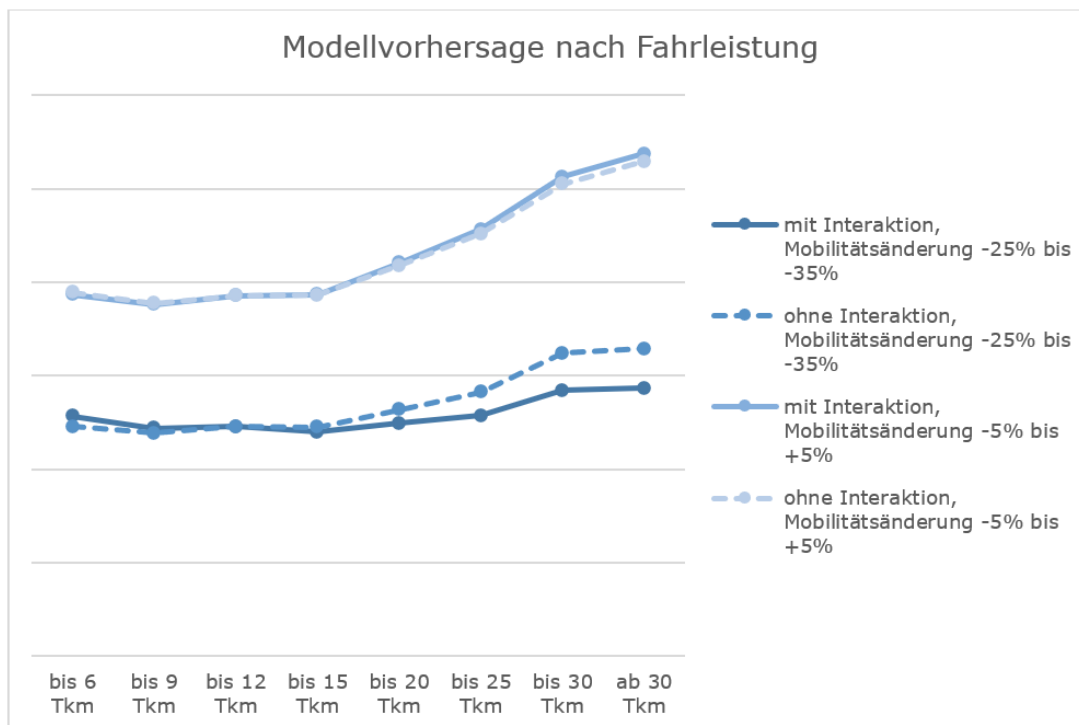
Betrachtet man die Vorhersagen der Modelle auf dem 1-dimensionalen Rand des Merkmals Mobilitätsänderung, zeigt sich die bessere Anpassung an die Empirie der Modelle 2 und 3.



### 5.3.3. Auswirkungen auf das restliche Modell

Wir untersuchen die Auswirkungen auf das restliche Modell anhand der Kovariable „Fahrleistung“ (jährliche km-Fahrleistung). Dazu vergleichen wir die Modelle mit und ohne die Kovariable „Mobilitätsänderung“ auf Basis des Modells 3 aus 5.3.2. und stellen keine nennenswerten Abweichungen in den Koeffizienten der Fahrleistung fest.

Betrachtet man hingegen die Interaktion zwischen den Kovariablen „Fahrleistung“ und „Mobilitätsänderung“ zeigt sich mittels Likelihood-Ratio-Test, dass das Modell mit der zusätzlichen Interaktion dem Modell ohne Interaktion vorzuziehen ist ( $p$ -Wert = 0,013). Die grafische Analyse der beiden Modellvorhersagen bestätigt das Testergebnis. Unterschiede zwischen den Modellen zeigen sich vor allem bei größerem Mobilitätsrückgang und werden in den höheren Fahrleistungsklassen deutlich. Exemplarisch sind Modellvorhersagen für die Bereiche der Mobilitätsänderungen von -5% bis +5% („keine Änderung“) und von -25% bis -35% („hoher Mobilitätsrückgang“) in der folgenden Graphik dargestellt:



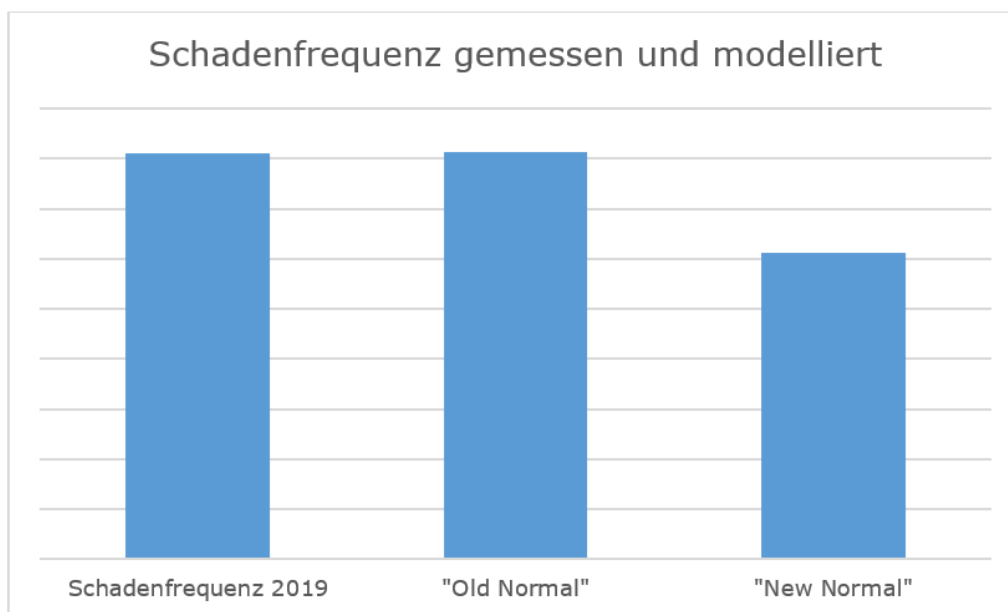
Da ein überproportionaler Rückgang der Schadenfrequenz bei verminderter Mobilität in den höheren Fahrleistungsklassen auch inhaltlich plausibel erscheint, nehmen wir die Interaktion aus Fahrleistung und Mobilitätsänderung in das finale Modell auf.

### 5.3.4. Niveau und Prognose

Durch die gewählte Modellierung mit der Betroffenheitsvariablen „Mobilitätsänderung“ lässt sich der Rückgang der Schadenfrequenz im Jahr 2020 gegenüber 2019 nur teilweise erklären. Auch unter Berücksichtigung des langfristigen Trends in der Pkw-Sachschadenfrequenz verbleibt ein Rückgang von ca. 8%, der nicht durch die Mobilitätsänderung erklärt werden kann und dementsprechend im vorliegenden

Modell allein durch das globale Niveau (also die Kovariable „Jahr“) beschrieben wird. Ob diese Effekte durch zusätzliche Betroffenheitsvariablen erklärt werden können, bleibt im Rahmen des Use Cases unbeantwortet.

Für die Prognose empfiehlt sich, mehrere Szenarien zu betrachten. Exemplarisch seien hier zwei Szenarien betrachtet. Das Szenario „New Normal“ mit einem Mobilitätsrückgang von 20% gegenüber 2019 verbunden mit der (sicherlich vereinfachenden) Annahme, dass das globale Niveau dem des Jahres 2019 entspricht. Und das Szenario „Old Normal“ ohne Änderungen gegenüber dem Jahr 2019. Dann ergeben sich auf einem Holdout-Sample von jeweils 100.000 Jahreseinheiten aus den Jahren 2019 und 2020 folgende modellierte und gemessene Schadenfrequenzen:



Die gemessene Schadenfrequenz 2019 unterscheidet sich wie erwartet nur marginal vom Szenario „Old Normal“. Der angenommene Mobilitäts-Rückgang von 20% beim Übergang vom Szenario „Old Normal“ zu „New Normal“ resultiert in einer prognostizierten Schadenfrequenzreduktion von fast 25%.

#### **5.4. Praktische Hinweise aus dem Use Case**

Wie in fast allen aktuariellen Tarifierungsaufgaben ist auch bei der Betroffenheitsmodellierung eine adäquate Datenaufbereitung der Schlüssel zum Erfolg und verursachte dementsprechend auch den größten Aufwand beim Erstellen des Use Cases. Trotz der sorgsamsten Datenaufbereitung sind einige sich überlagernde Effekte nicht vollständig zu trennen gewesen – beispielsweise die Trennung von Saisonalität und Betroffenheit oder auch die Trennung von Betroffenheit und zugrundeliegendem Trend.

Im vorliegenden Use Case konnte gezeigt werden, dass das durch COVID-19 geprägte Statistikjahr zwar zu einer verminderten Schadenfrequenz im gewählten Segment der Kraftfahrt-Haftpflicht geführt hat. Abgesehen von den hier aufgezeigten Änderungen im Merkmal „Fahrleistung“ wurden jedoch keine starken Verwerfungen in der zugrundeliegenden Risikostruktur beobachtet.

Die beobachteten Niveaueffekte konnten mit Hilfe der Betroffenheitsvariable modelliert werden und bereinigen weitgehend die Schätzung für das Jahr 2020 um die COVID-19-Effekte. Mit Hilfe der Modellierung der Betroffenheitseffekte ist damit eine fundiertere Prognose aus dem verwendeten Modell heraus möglich.

Der gemessenen Interaktion der Fahrleistung mit der Betroffenheitsvariable sollte in der Prognose angemessene Aufmerksamkeit geschenkt werden, da nicht ausgeschlossen werden kann, dass Effekte auf fehlende Meldungen der erwarteten Fahrleistung zurückzuführen sind. Das kann zu falschen Schlüssen in der Prognose führen.

## **6. Fazit und Ausblick**

### **6.1. Modellierung mit Betroffenheitsvariablen in der COVID-19-Pandemie**

Der offene Umgang mit den Messdaten der Pandemie führt zu frei verfügbaren Datenquellen mit direktem Bezug, aus denen recht bequem Betroffenheitsvariablen gewonnen werden können. Mobilitäts- und Kontakteinschränkungen als Standardmaßnahmen zur Eindämmung der Epidemie können in ihrer Wirkung durch Mobilitätsanalysen auf der Basis von Mobilfunkdaten beschrieben und analysiert werden. Zusammen mit übergreifenden Zeitreihen steht ein reichhaltiges Arsenal an potenziellen Kovariablen zur Modellierung von Betroffenheit zur Verfügung. Da die Betroffenheitsvariablen systematisch Zeitscheiben abgrenzen ist eine sorgfältige Planung ihrer Aufbereitung und Verwendung notwendig, um eine Aufblähung der Datensätze bis hin zum Zusammenbruch von Modellierungsstrukturen zu vermeiden. Die resultierende erhöhte Komplexität der Modelle muss immer im Verhältnis zum Informationsgewinn gesehen werden. Aber schon einfachste Variablen – z.B. „Lockdown Ja/Nein“ – schaffen ex post einen Zugang zum Verständnis der Auswirkungen der COVID-19-Effekte in Unterscheidung zu anderen Saisonereffekten. Variablen, die auch nach der Pandemie eine nachhaltige Verhaltensänderung im „New Normal“ beschreiben, ermöglichen auf der Basis fundierter Fortschreibungsszenarien darüber hinaus sachgerechtere und genauere Prognosen für die Zeit, in der das COVID-19-Virus ein dauerhafter Faktor in unserer Umwelt geworden sein wird. Insofern wurde z.B. in unserem Use-Case mit Hilfe der Betroffenheitsvariable „Mobilitätsänderung“ durch die COVID-19-Effekte der Einfluss der Veränderung von Mobilität auf die Schadenhäufigkeit messbar.

### **6.2. Anwendungsmöglichkeiten über Corona hinaus**

Der Ansatz eignet sich aber nicht nur für die spezielle Situation der COVID-19-Pandemie. Vielmehr empfiehlt er sich immer, wenn externe, temporale Faktoren die Kalkulationsdaten merklich beeinflussen. Das kann zum Beispiel bei zeitlich begrenzten (bspw. Finanzkrise), aber auch allmählichen Verhaltensänderungen (bspw. Klimawandel) der Fall sein. Bei der Übertragung ist darauf zu achten, dass die Betroffenheitsvariablen mit dem Effekt in einem Zusammenhang stehen. Möglicherweise müssen hier mehrere geeignete Kandidaten von Variablen in Betracht gezogen werden. Im Falle von allmählichen Änderungen ist dann eine geeignete Prognose der Betroffenheitsvariablen vorzunehmen. Schwieriger wird es, wenn es sich bei den Änderungen um Strukturbrüche (Leute ziehen auf das Land, durch Home-Office zukünftig weniger Mobilität nötig ...) handelt, d. h. einer dauerhaften Änderung des Verhaltens. Aber auch hier kann eine Analyse mit der Verwendung von Betroffenheitsvariablen wertvolle Hinweise geben.

## 7. Anhang

### 7.1. Berechnung des R-Wertes

Bei der Berechnung verwendet das RKI<sup>16</sup> ein bestimmtes Zeitintervall, die sogenannte Generationszeit. Diese besagt, wie lange es dauert, bis ein Infizierter die nächste Person ansteckt. Das RKI geht dabei von vier Tagen aus. Will man jetzt R berechnen, addiert man die Infektionszahlen einer Generationszeit und teilt sie durch die Infektionszahlen der vorherigen Generationszeit. Damit aber Meldeverzug die Rechnung nicht verzerrt, werden die zurückliegenden drei Tage nicht mit einbezogen. Wenn das RKI also R für den 14. Mai. berechnet, ignoriert es die Infektionszahlen vom 11. bis zum 13. Mai. Als erste Generationszeit werden die Zahlen vom 7. bis zum 10. Mai addiert und geteilt durch die Infektionszahlen vom 3. bis zum 6. Mai. So käme man auf die Rechnung  $(557 + 611 + 680 + 752) / (824 + 884 + 931 + 819) = 0,75$ .

Das Problem bei dieser Rechnung ist, dass lokale Ausbrüche wie z.B. bei Schlachtbetrieben in Nordrhein-Westfalen und Baden-Württemberg R kurzzeitig nach oben treiben können. Das verursacht tägliche Schwankungen, die wenig über den langfristigen Trend aussagen.

Daher wird versucht, bei der geglätteten Reproduktionszahl diese Schwankungen auszugleichen. Das Prinzip ist grundsätzlich das gleiche wie zuvor. Der einzige Unterschied: Das RKI verwendet hier nicht vier Tage als Intervall, sondern sieben Tage. Wenn also an einem Tag die Infektionszahlen deutlich über dem Rest der Woche liegen, fällt das nicht so sehr ins Gewicht. Ausbrüche wie in den Fleischbetrieben können die Reproduktionszahl also nicht so sehr verzerren.

### 7.2. Beispiel-Programme zur Berechnung einiger Betroffenheitsvariablen

Für die Berechnung einiger Betroffenheitsvariablen steht auf dem [GitHub-Konto der DAV](#) in den Programmiersprachen SAS und Python beispielhaft Programmcode zur Verfügung.

---

<sup>16</sup> Siehe: [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Projekte\\_RKI/R-Wert-Erlaeuterung.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/R-Wert-Erlaeuterung.pdf?__blob=publicationFile), zuletzt aufgerufen am 29.12.2021