

1. Einleitung

1.1 Ausgangspunkt

Die KOF Konjunkturforschungsstelle der ETH Zürich berechnet jährlich eine Prognose für die Kostenentwicklung im Bereich der obligatorischen Kranken- und Pflegeversicherung (OKP). Die Prognose beruht auf der Machbarkeitsstudie der KOF von Köthenbürger und Sandqvist (2018), welche die Datenlage und mögliche Methoden zur Modellierung der Zeitreihen eingehend geprüft haben. Der vorliegende Revisionsbericht prüft die Qualität der bestehenden kantonalen Prognosemodelle und verbreitert das Modellspektrum auf zuvor noch nicht getestete Modelle. Die Revision ist Teil des normalen Evaluierungsprozesses, aber auch Antwort auf die im Vergleich zur Pilotstudie verbesserte Datenlage. Die Prognosegüte wurde analog der Pilotstudie von 2018 mittels «pseudo-out-of-sample»-Prognosen evaluiert, was einen direkten Vergleich der Ergebnisse ermöglicht.

2. Evaluierung

2.1 Modellspektrum

Bei der Pilotstudie von Köthenbürger und Sandqvist (2018) erzielte das gepoolte autoregressive Modell den kleinsten durchschnittlichen Fehler über alle Kantone hinweg. Obwohl die Prognosegüte in der Vergangenheit hoch war, bringen die Voraussagen auf Basis dieses Modells auch Nachteile mit sich. Der Vorteil des Poolens ist denn auch der grösste Nachteil. Durch die Verbindung von Längs- und Querschnitt wird das Sample zwar vergrössert (was v.a. bei den teils hoch korrelierten kantonalen Daten durchaus Sinn ergeben kann). Die kantonale Heterogenität geht bei der Schätzung dadurch aber natürlich verloren. Letzteres führt zu quasi identischen Konfidenzbändern wie auch ähnlichen Punktschätzern, da nur der autoregressive Teil des Modells (und nicht die Konstante) die jeweiligen Kantonsverhältnisse in die Prognose einfließen lässt. Aus diesem Grund prüft der vorliegende Bericht die Gruppierung von Kantonen ("Clustering"), welche ähnliche Dynamiken und Durchschnittswachstumsraten umfassen. Durch das Poolen von statistisch gesehen homogenen Kantonen könnte eine Balance zwischen kantonomer Heterogenität und Vergrößerung des sonst kleinen Samples erreicht werden. Die so gepoolten Gruppen werden dann mittels dem in der Pilotstudie am besten abschneidenden Modell sowie einem generelleren, autoregressiv gleitenden Durchschnittsmodell (ARMA) geschätzt und prognostiziert. Das zu evaluierende Modellspektrum wird zudem erweitert um ein dynamisches Faktormodell, also ein klassisches Werkzeug der multivariaten Statistik. Da die kantonalen Reihen wie bereits erwähnt z.T. hoch korreliert sind, lohnt es sich, das Faktormodell und dessen Prognosegüte eingehend zu prüfen, da dieses in vielen Fällen gemeinsame Dynamiken über die Zeit hinweg extrahieren kann. Zusammen mit den univariaten Modellen der Pilotstudie ergibt sich somit folgendes Modellspektrum:

Modelle	Abkürzung
Gepooltes autoregressives Modell	Baseline
Gepooltes autoregressives Modell mit Gruppen	Baseline Gruppen
Gepooltes, autoregressiv gleitendes Durchschnittsmodell mit Gruppen	ARMA Pooled Gruppen
Kantonsspezifisches, autoregressiv gleitendes Durchschnittsmodell	ARMA
Kantonsspezifisches, autoregressiv gleitendes Durchschnittsmodell mit Ausreisserbereinigung	ARMA Dummy
Faktormodell	Faktor

Die zweite Spalte in der obigen Tabelle zeigt die nachfolgend verwendeten Abkürzungen für die jeweilige Modellart. Der erste Eintrag – das gepoolte autoregressive Modell – ist mit der Abkürzung «Baseline» ausgestattet, da dieses in der Pilotstudie die beste Prognosegüte aufwies und darum im Prognosebericht von 2019 (Anderes & Sturm, 2019) für die kantonalen Voraussagen benutzt wurde. In

diesem Sinne stellt es einen natürlichen Vergleichsrahmen für die Ergebnisse der anderen Modelle. Da beim zweiten Eintrag die gleiche Modellspezifikation benutzt wird (ein autoregressiver Term mit gepoolten Daten), wobei der einzige Unterschied im individuellen Schätzen der Gruppen liegt, wurde als Abkürzung «Baseline Gruppen» gewählt. Weil Ausreisser im vorliegenden Bericht anhand von sogenannten Dummy-Variablen bereinigt werden, wird das ARMA-Modell mit Ausreisserbereinigung als «ARMA Dummy» abgekürzt.

2.2 Ergebnisse

Jedes Modell wurde auf die Fähigkeit hin überprüft, die kantonalen Wachstumsraten der OKP-Kosten bestmöglich prognostizieren zu können. Dabei wurde wie in der Pilotstudie mittels "out-of-sample"-Prognosen gearbeitet. Nach dieser Methodik werden die Daten künstlich gekürzt, im vorliegenden Fall bis 2015, und die Genauigkeit der Prognose für 2016 und 2017 evaluiert. Dieser Schritt wird bis zum letzten verfügbaren Zeitpunkt wiederholt (2019). Auf diese Weise wird geprüft, wie das Modell unter realistischen Bedingungen in den letzten Jahren prognostiziert hätte. Mit den berechneten Prognosefehlern werden schliesslich Fehlermasse gebildet, z.B. der "root mean square error" (RMSE, Wurzel der mittleren Fehlerquadratsumme) oder der "mean absolute error" (MAE, mittlerer absoluter Fehler). Diese Masszahlen vermitteln ein Bild der erreichten Prognosegüte, indem sie die vergangenen Prognosefehler in einer Zahl destillieren. Der vorliegende Bericht fokussiert auf den oft verwendeten RMSE.

Auf eine Diskussion der einzelnen Kantonsresultate wird an dieser Stelle verzichtet. Grundsätzlich beweist das Faktormodell seine Stärke, da es für zwölf von 26 Kantonen den tiefsten RMSE aufweist. Vor allem bei den Grenzkantonen scheint das Faktormodell aber schlecht abgeschnitten zu haben, was wohl am Strukturbruch der Zeitreihen im Jahr 2018 liegt¹. Zudem ist bei acht Kantonen die univariate Modellierung am genauesten, also entweder ein ARMA-Modell mit oder ohne Ausreisserbereinigung (Dummies). In fünf Kantonen schliesst das Baseline-Modell aus der Pilotstudie am besten ab und nur in einem Fall das gepoolte gruppierte ARMA-Modell. Das gepoolte gruppierte Baseline-Modell erzielt in keinem Kanton den kleinsten Prognosefehler. Aufgrund der mangelhaften Ergebnisse der beiden gruppierten Modelle wird bei der weiteren Diskussion nicht weiter darauf eingegangen. Wird die Modellvielfalt so eingeschränkt, bleiben die folgenden vier Modelle übrig: Das Baseline-Modell, das Faktormodell sowie das ARMA-Modell mit und ohne Ausreisserbereinigung.

Da wie im letztjährigen Prognosebericht alle Kantone mit demselben Modelltypus prognostiziert werden sollen², gibt die nachfolgende Tabelle einen Überblick über die durchschnittlichen Prognosefehler pro Modell über alle Kantone hinweg.

Modell	Baseline	ARMA	ARMA Dummy	Faktor
RMSE	2.54	2.66	2.52	2.97

Da der RMSE die ursprüngliche Skala beibehält, erfolgen alle Angaben in Prozent. Die Ergebnisse bestätigen das gute Abschneiden des Baseline-Modells, das bereits in der früheren Pilotstudie zu beobachten war. Der Grund hierfür stellt aber wie erwähnt gleichzeitig das Problem dieses Modells dar: Die Prognosen streuen relativ stark um die Durchschnittswachstumsraten aller Kantone. Aufgrund der hohen jährlichen Stabilität einer solchen Prognose (da sich der Durchschnitt nicht schnell ändert) ergeben sich denn auch selten grosse Ausreisser, welche das über alle Kantone hinweg betrachtete Fehlermass hochtreiben. Knapp besser schneidet nur das univariate ARMA-Modell mit Ausreisserbereinigung ab, wobei dieses auch über alle Kantone hinweg relativ stabil ist. Obwohl das Faktormodell für einige Kantone eine hohe Präzision aufweist, sind die Prognosen bei anderen Kantonen mit einem grossen Prognosefehler behaftet. Im Durchschnitt ergibt sich so ein höheres Fehlermass als bei den anderen Modellen.

¹ Seit 2018 werden die Grenzgänger nicht mehr den Kantonen zugerechnet, was zu einem einmaligen Anstieg der Wachstumsraten aufgrund einer Niveaueverschiebung führt, da Grenzgänger tendenziell jung sind und wenig Kosten verursachen.

² Die Spezifikation der univariaten Modelle, also z.B. die Anzahl autoregressiver Terme, wird natürlich für jeden Kanton individuell angepasst.

Der Unterschied zwischen der Prognosegüte des Baseline-Modells und des ARMA-Modells mit Ausreisserbereinigung ist gering und kaum signifikant. Da die Zeitreihen kurz sind, ist es nicht möglich, eine grosse Anzahl von historischen Prognosefehlern zu berechnen. Denn die Modelle selbst können sonst nicht mehr geschätzt werden. Dadurch stellen die Fehlermasse eine Funktion von relativ wenigen Prognosefehlern dar, womit der Unterschied im RMSE von rund 0.02% kein ausschlaggebendes Kriterium für die Modellwahl ist. Ein weiteres Kriterium ist aber bei den Unsicherheitsmassen zu finden. Da die Prognoseintervalle bei den univariaten ARMA-Modellen kantonsspezifisch sind, ist diese Modellklasse bei gleicher Prognosegüte den Baseline-Modellen gegenüber zu bevorzugen. Wie in der Einleitung beschrieben, bringt das gepoolte Modell (Baseline) den Nachteil des Verlustes der kantonalen Heterogenität mit sich, was bei der Prognose zu gleichen Unsicherheitsmassen über die Kantone hinweg führt. Dies ist bei der Verwendung der ARMA-Modelle nicht der Fall, da die Prognoseunsicherheit bei jedem Kanton mit der historischen Varianz verflochten ist, was ein realistischeres Bild der Unsicherheit abgeben dürfte. Aus diesen Gründen ist zu empfehlen, dass die kantonalen Prognosen anhand kantonsspezifischer, autoregressiv gleitender Durchschnittsmodelle erstellt werden.

3. Fazit

Grundsätzlich stützt der diesjährige Revisionsbericht die Resultate der Pilotstudie. So gibt es für die Kantone immer noch kein einzelnes Modell, welches für alle Zeitreihen den kleinsten Prognosefehler erzielt. Zudem überzeugt das gepoolte autoregressive Modell weiterhin durch eine hohe Genauigkeit, wobei die univariate Modellierung mittels ARMA-Modelle mit Ausreisserbereinigung die knapp beste Prognosegüte aufweist. Da dieses Modell im Gegensatz zum Modell der Pilotstudie (Köthenbürger und Sandqvist, 2018) zudem die zukünftige an die historische Unsicherheit knüpft, ist eine Verwendung in kommenden Prognoseberichten grundsätzlich zu empfehlen.

4. Literaturverzeichnis

Anderes, M. und Sturm, J. E. (2019) *Prognose der Kostenentwicklung in der obligatorischen Krankenpflegeversicherung (OKP)*. KOF Studien, Nr. 138.

Köthenbürger, M. und Sandqvist, A. P. (2018) *Prognoseverfahren für die Gesundheitsausgaben in der obligatorischen Krankenversicherung*. KOF Studien, Nr. 125.