

Prognosemodelle

Arndt Leininger

Abstract: *Prognosen stellen in der Politikwissenschaft ein zwar noch kleines, aber stetig wachsendes Forschungsfeld dar, welches in verschiedenen Teilbereichen der Disziplin Anwendung findet. Gemeint sind hiermit statistische Modelle, mit denen explizit politikwissenschaftlich relevante Phänomene vor ihrem Eintreten vorhergesagt werden. Dabei folgen sie den wissenschaftlichen Leitlinien der intersubjektiven Nachvollziehbarkeit und Reproduzierbarkeit. Dieser Beitrag führt ein in die Grundlagen politikwissenschaftlicher Prognosen. Den Schwerpunkt der Darstellung bilden Wahlprognosen, insbesondere strukturelle Modelle, welche beispielhaft anhand eines kanonischen Wahlprognosemodells erläutert werden. Daneben werden synthetische Modelle, Aggregationsmodelle, “Wisdom of the crowd”-Ansätze und Prognosemärkte diskutiert.*

Schlüsselwörter: Forecasting, Prognosen, Quantitative Methoden, Vergleichende Politikwissenschaft, Wahlprognosen, Konfliktforschung

1 Einleitung

Prognosemodelle stellen in der Politikwissenschaft ein zwar noch vergleichsweise kleines, aber wachsendes Forschungsfeld dar. Gemeint sind hiermit statistische Modelle, mit denen explizit politikwissenschaftlich relevante Phänomene wie Wahlergebnisse, Bürgerkriege oder Verfassungsgerichtsentscheidungen, um nur einige Beispiele zu nennen, Wochen, Monate oder gar Jahre im Voraus vorhergesagt werden. In der englischsprachigen Politikwissenschaft werden sie *forecasting models*, welche *forecasts* in Abgrenzung zu *predictions* generieren, genannt. Gemeint ist hiermit, dass Datenpunkte vorhergesagt werden, die nicht für die Modellschätzung berücksichtigt werden. Dabei folgen sie den wissenschaftlichen Leitlinien der intersubjektiven Nachvollziehbarkeit und Reproduzierbarkeit.

Ziel dieses Beitrags ist es, eine praxisorientierte Einführung in politikwissenschaftliche Prognosemodelle zu geben. Der Leserin¹ sollen die Grundlagen politikwissenschaftlicher Prognosemodelle vermittelt und ein Überblick über das Forschungsfeld, wie es sich momentan darstellt, gegeben werden. Sie soll zudem befähigt werden, Prognosemodelle kritisch bewerten zu können sowie eigene Prognosemodelle aufzustellen.

Im Folgenden stelle ich zunächst kurz dar, *was* politikwissenschaftliche Prognosemodelle sind. Dem schließt sich unmittelbar eine kurze Diskussion des *Warum* an, während ich dann im folgenden Abschnitt das *Wie* exemplarisch erläutere. Hierfür wähle ich bewusst ein etabliertes und denkbar einfaches Modell aus der Wahlforschung. Im sich daran anschließenden Abschnitt stelle ich weitere Ansätze vor, um die Vielfalt und die neueren Entwicklungen im Bereich der Prognosemodelle darzustellen. Besondere Aufmerksamkeit schenke ich dabei Wahlprognosen. Diese gehören zu den am verbreitetsten und am weitesten entwickelten Prognosemodellen in der Politikwissenschaft. Auch in anderen Teilbereichen der Politikwissenschaft finden sich Prognosemodelle - insbesondere zur Prognose inner- und zwischenstaatlichen Konflikten (Gleditsch und Ward 2013) sowie von Gerichtsurteilen (Kastellec 2010). Daher werde ich vereinzelt auch auf diese und andere Anwendungsfälle Bezug nehmen. Abschließend diskutiere ich die Potenziale und Problemfelder des Forschungsfeldes. Der letzte Abschnitt fasst zusammen und bietet einen Ausblick.

Die Erläuterung der Grundlagen politikwissenschaftlicher Prognosemodelle erfolgt in diesem Beitrag wesentlich am Beispiel von Wahlprognosen. Was sind Wahlprognosemodelle und warum hat die Wahlforschung neben Wahlumfragen noch eine Vielfalt an Prognosemodellen entwickelt? Wahlumfragen mögen zwar den Ausgang von Wahlen kurz vor deren Stattfinden genauer vorhersagen, aber mit zunehmendem zeitlichen Abstand zur Wahl und damit auch größerem Desinteresse der Wählerin gegenüber der bevorstehenden Wahl verringert sich diese Überlegenheit so weit, dass Umfragen schließlich ungenauer sind als Prognosemodelle (Erikson und Wlezien 2014). Obwohl Wahlumfragen und deren Aggregation häufiger als solche wahrgenommen werden (Blumenthal 2014), handelt es sich bei ihnen eher um prinzipiell vergängliche Momentaufnahmen. Prognosemodelle sind nicht nur bessere Prädiktoren, wenn die Wahl, deren Ausgang prognostiziert werden soll, zeitlich noch weit entfernt ist, sie liefern auch echte Prognosen.

¹Aufgrund der besseren Lesbarkeit wird in diesem Beitrag der Einfachheit halber nur die weibliche Form verwendet. Die männliche Form ist selbstverständlich immer mit eingeschlossen.

Es lassen sich Leininger (2015) folgend folgend grob drei verschiedene Ansätze unterscheiden: strukturelle Modelle, Aggregationsmodelle und synthetische Modelle. Strukturelle Modelle ermitteln Zusammenhänge zwischen Wahlergebnissen und erklärenden Faktoren und schreiben diese Zusammenhänge in die Zukunft fort. Der konkrete Aufbau eines strukturellen Modells leitet sich aus Theorie und Empirie der Wahlforschung ab. Aggregationsmodelle fußen primär auf statistischer Theorie und fassen mehrere Umfragen zusammen, um zu einer gegenüber einzelnen Umfragen verbesserten Prognose zu gelangen. Synthetische Modelle schließlich kombinieren strukturelle und Aggregationsmodelle. Diese Kategorisierung - theoretisch motivierte strukturelle Modelle, primär statistisch getriebene Aggregationsmodelle und deren Kombination, synthetische Modelle - ist im Prinzip auch auf Prognosemodelle außerhalb der Wahlforschung anwendbar. In diesem Beitrag erweitere ich die Perspektive und stelle zusätzlich auch “Wisdom of the Crowd”-Ansätze und Prognosemärkte kurz vor.

Prognosemodelle sind also statistische Modelle, die, statt, wie sonst üblich, zur Erklärung, zur Prospektion genutzt werden. Warum aber dieser Wechsel in der Perspektive der Analyse? So wie in der Meteorologie oder der Volkswirtschaftslehre Prognosen ein fester Bestandteil der Disziplin sind, so hat auch die Politikwissenschaft ein naheliegendes, wenn auch noch vergleichsweise weniger ausgeprägtes, Interesse an Prognosen. Politikwissenschaftliche Analysen zielen in der Regel nicht darauf ab, Besonderheiten einzelner Ereignisse, seien es Wahlen oder Bürgerkriege, hervorzuheben, sondern allgemeine Muster und damit letztendlich kausale Mechanismen aufzudecken. Auf der Basis kausaler Theorien und sie abbildender empirischer Modelle sollten Fortschreibungen auch in die Zukunft grundsätzlich möglich sein. Prognostische Validität ist jedoch schwieriger zu erreichen als die (retrospektive) Erklärung und wird daher nach Meinung mancher Autorinnen in den Sozialwissenschaften (noch) vernachlässigt (Schrodt 2013).

Die Wahlforschung eignet sich aus verschiedenen Gründen besonders gut als Anwendungsfeld für Prognosen. Wahlen finden regelmäßig statt und ihr Ausgang ist exakt messbar. Zugleich kann auf einen umfangreichen Fundus an Theoriebildung und empirisch-quantitativer Forschung zurückgegriffen werden. Wissenschaftliche Wahlprognosen nahmen ihren Anfang Ende der 1970er Jahre in den USA. Seitdem wurde eine Vielzahl verschiedener Modelle entwickelt, jenseits der USA vor allem im Vereinigten Königreich, Frankreich und Deutschland (Lewis-Beck 2005). Die beispielhafte Darstellung eines klassischen Prognosemodells folgenden

Abschnitt soll in die Grundlagen politikwissenschaftlicher Prognosen einführen.

2 Ein Praxisbeispiel: das “Time for Change”-Modell

Um diese praxisorientierte Einführung überschaubar zu halten greife ich auf ein möglichst einfaches Modell zurück, welches für den für diese Zwecke denkbar einfachsten politischen Kontext, US-amerikanische Präsidentschaftswahlen, entwickelt wurde. Im Folgenden stelle ich das sogenannte “Time for Change”-Modell des Politikwissenschaftlers Alan I. Abramowitz dar. Dieses Modell prognostiziert den auf die Kandidatin der amtsinhabenden Partei entfallenden Stimmanteil. Das Modell wurde erstmals 1988 publiziert und auch für darauffolgende Wahlen, mit leichten Modifikationen, genutzt. Die hier verwendete Darstellung stützt sich auf die Version des Modells, welches zur Prognose der Wahl 2012 zum Einsatz kam (Abramowitz 2012).

Das Modell basiert auf einer Zeitreihe vergangener US-Präsidentschaftswahlen seit 1948 und einem OLS-Regressionsmodell. Modelliert werden der Stimmanteil eines Präsidentschaftskandidaten als Funktion verschiedener erklärender Variablen. Im hier vorgestellten Modell wird wie in den meisten solcher Modelle der sogenannte *two party vote share* unter Ausschluss der sogenannten *third party candidates* geschätzt. Somit wird der Stimmanteil für den Herausforderer, als Differenz des Schätzwerts für den Amtsinhaber zu 100%, immer mit prognostiziert. Das “Time for Change”-Modell ist wie folgt spezifiziert:

$$y_t = \beta_0 + \beta_1 NETAPP_t + \beta_2 Q2GDP_t + \beta_3 TERM1INC_t + \epsilon_t$$

Der auf den Amtsinhaber entfallende Stimmanteil y im Jahr t wird zurückgeführt auf drei Variablen. $NETAPP$ bezeichnet die Differenz zwischen den Zustimmung- und Ablehnungswerten (“approval” und “disapproval”) aus der letzten im Juni des Wahljahres durchgeführten Umfrage des Instituts Gallup. $Q2GDP$ bezeichnet die Wachstumsrate im zweiten Quartal des Wahljahres und $TERM1INC$ ist eine Dummy-Variable welche gleich eins ist, wenn die Amtsinhaber erneut antritt.² Wie viele andere strukturelle Modelle enthält das “Time for Change”-Modell

²Andere Versionen dieses Modells enthalten statt dieser Variablen eine Dummy-Variable die anzeigt, ob die amtierende Präsidentin bereits zwei Amtszeiten im Amt war.

ein Maß der Wirtschaftslage sowie ein Maß für die Popularität der Regierung - Variablen, die oft als *fundamentals* bezeichnet werden (Erikson und Wlezien 2012). Schätzt man das Modell über Wahlen von 1948 bis einschließlich 2008, so erhält man folgende Koeffizienten:

$$\hat{y}_i = 47,78 + 0,12 \times NETAPP + 0,46 \times GROWTH + 3,02 * TERM1INC$$

In die geschätzte Regressionsgleichung lassen sich sodann die aktuellen Werte, in unserem Beispiel für 2012, für die erklärenden Variablen einsetzen, um den zu erwartenden Stimmanteil in einer bevorstehenden Wahl zu schätzen. Die in das Modell einfließenden Werte für die unabhängigen Variablen können selbst Schätzwerte oder Prognosen sein. Ist dies der Fall, so lässt sich das Modell aktualisieren, wann immer sich die in die Modellgleichung einzusetzenden Werte ändern.

Setzt man nun in die oben stehende Gleichung die Werte für 2012 ein, so erhält man die folgende Prognose: Auf den Amtsinhaber Barack Obama werden 51,63 % der Stimmen entfallen. Die Differenz zum tatsächlichen Ergebnis, 52,01% des “two-party vote share”, fällt mit 0,38 Prozentpunkten sehr klein aus. Diese Abweichung der Prognose vom tatsächlichen Wahlergebnis wird als Prognosefehler bezeichnet.

Die Validität eines solchen Modells, wie es hier zur Anwendung kommt, beruht im Kern auf dessen Fähigkeit, Ergebnisse von Wahlen vorherzusagen, die nicht selbst Teil der Stichprobe sind (auf Englisch: *out-of-sample prediction* oder *forecast*). Der ultimative Test ist die Vorhersage einer noch bevorstehenden Wahl. Doch muss man nicht bis zur nächsten Wahl warten, um zu prüfen, ob ein Modell geeignet ist, gute Prognosen zu generieren. Indem man eine Wahl aus der Stichprobe ausschließt und das Modell auf Basis der reduzierten Stichprobe erneut schätzt, kann man mit den Werten der nicht berücksichtigten Beobachtungen eine Prognose für diese erstellen.

Diese Unterscheidung von Trainings- und Testdatensätzen stellt eines der Grundprinzipien wissenschaftlicher Prognosemodelle dar und findet sich, wenn auch nicht immer mit diesen Begriffen belegt, in allen Anwendungen politikwissenschaftlicher Prognosemodelle. Für eine echte Prognose ist es essentiell, dass die vorherzusagenden Datenpunkte nicht Bestandteil der Stichprobe sind, auf

der das statistische Modell geschätzt wurde. Im Englischen wird daher explizit begrifflich zwischen *prediction* und *forecast* unterschieden. Der erste Begriff kann auch eine Vorhersage der Werte der abhängigen Variablen für Beobachtungen der Stichprobe, auf der das Modell geschätzt wurde, bedeuten. Letzterer bezeichnet exklusiv die Vorhersage von Beobachtungen, die nicht in die Modellschätzung einfließen.

Bei der Vorhersage von in der Zukunft liegenden Ereignissen ergibt sich diese Unterscheidung zwangsläufig. Der Wert der abhängigen Variable ist noch unbekannt kann aber auf Basis der bereits bekannten Werte der unabhängigen Variablen geschätzt werden. In der Entwicklung eines Prognosemodells ist es üblich, dass die Forschende ihren Datensatz in Trainings- und Testdatensatz unterteilt. In der Regel werden eine oder mehrere Beobachtungseinheiten über mehrere Zeiträume hinweg beobachtet. Die Forschende setzt dann einen Zeitpunkt T und behandelt alle vor diesem Zeitpunkt T ($t < T$) gemessenen Beobachtungen als Trainingsdatensatz und alle Beobachtungen zum Zeitpunkt T oder später ($t \geq T$) als Testdatensatz. Das Modell wird auf dem Trainingsdatensatz geschätzt und über die Vorhersagen für den Testdatensatz evaluiert.

In meinem eben eingeführten Beispiel bedeutet dies, das Modell zunächst auf den Zeitreihe von 1948 bis 2008 zu schätzen, um das Wahlergebnis 2012 vorherzusagen. Im Vorfeld der Wahl 2012 hat Abramowitz das Modell auf den Zeitreihen 1948 bis 2004 geschätzt, um die Wahl 2008 vorherzusagen. Ich berechne nach gleicher Vorgehensweise Prognosen für die Wahlen 1988, der ersten Wahl die Abramowitz mit seinem Modell vorhersagte, bis einschließlich 2008 (siehe Tabelle 1).

Tabelle 1: *Out-of-Sample-Prognosen 1988 bis 2008 und Zusammenfassung der Fehler als Mean Absolute Error (MAE) und Root Mean Squared Error (RMSE)*

Jahr	1988	1992	1996	2000	2004	2008	MAE	RMSE
Ergebnis	53,83	46,38	54,74	50,26	51,23	46,31		
Prognose	50,73	52,05	55,58	53,84	51,76	42,50		
Fehler	-3,10	5,67	0,84	3,58	0,53	-3,81	2,92	3,42

Für jede Iteration wird der Prognosefehler berechnet. Anschließend gilt es, diese zusammenzufassen, um die Genauigkeit des Modells, insbesondere im Vergleich zu alternativen Spezifikationen, zu beurteilen. Die im Bereich der Wahlprognosen am häufigsten dafür genutzten Maße sind der Mean Absolute Error (MAE) und

der Root Mean Squared Error (RMSE). Der MAE ist, wie der Name schon sagt, der Durchschnitt der absoluten Prognosefehler. Es wird mit absoluten Werten gerechnet damit sich positive, der vorhergesagte Stimmanteil ist höher als das tatsächliche Ergebnis, und negative, der vorhergesagte Stimmanteil ist niedriger als das tatsächliche Ergebnis, Fehler nicht gegenseitig ausgleichen. Ein andere Möglichkeit ist es, die Prognosefehler zu quadrieren. Daraus wird dann der Durchschnitt berechnet und anschließend noch die Wurzel genommen, um ein Maß zu erhalten, das in der gleichen Einheit wie die Prognosefehler, in unserem Fall Prozentpunkte, notiert.

Die eben beschriebenen Maße setzen eine metrisch skalierte abhängige Variable voraus und sind daher nicht für die Evaluation von Prognosemodellen für binäre Ereignisse geeignet. Außerhalb der Wahlforschung sind die vorherzusagenden Variablen jedoch oft binär: beispielsweise die An- oder Abwesenheit von Konflikten oder ein Gerichtsentscheid für oder gegen die Klägerin. Die in diesem Bereich eingesetzten Prognosemodelle resultieren oft in vorhersagten Wahrscheinlichkeiten. Ist dies der Fall, müssen diese zunächst durch Setzung eines Schwellenwerts, meist 50%, in binäre Vorhersagen übersetzt werden. Sodann lässt sich prüfen wie viele der vorhergesagten Beobachtungen korrekt klassifiziert werden. Dazu können die Maße Genauigkeit (“Accuracy”), Ausbeute (“Recall”) und Präzision (“Precision”) berechnet werden (O’Brien 2010).

Zur Illustration dieser Maße sei angenommen, dass mittels eines statistischen Modells das Auftreten von Bürgerkriegen prognostiziert wurde. Die Genauigkeit bezeichnet den Anteil der korrekten Vorhersagen an der Gesamtzahl der Vorhersagen. Die Ausbeute, auch Richtig-positiv-Rate, stellt die Zahl der korrekt vorhergesagten Bürgerkriege relativ zur Gesamtzahl der Bürgerkriege dar. Dazu analog wäre die Falsch-positiv-Rate die Zahl der fälschlich vorhergesagten Bürgerkriege relativ zur Gesamtzahl der Beobachtungen ohne Bürgerkrieg. Die Präzision schließlich ist die Zahl der korrekt vorhergesagten Bürgerkriege relativ zur Zahl der insgesamt vorhergesagten Bürgerkriege. Letztere können auch als bedingte Wahrscheinlichkeit des Eintretens des vorhergesagten Ereignisses interpretiert werden. Alternativ kann die Güte eines Prognosemodells auch über die Receiver-Operating-Characteristic-Kurve (auch ROC-Kurve) bewertet werden. Im Beitrag Gleditsch und Wards (2013) lässt sich die Bedeutung und Anwendung der ROC-Kurve anschaulich nachvollziehen.

Über den Vergleich verschiedener Modellspezifikationen versucht man nun, das

bestmögliche Prognosemodell zu finden. Ziel ist es, in den eben beschriebenen, sogenannten Out-of-Sample-Prognosen, den durchschnittlichen Prognosefehler zu minimieren. Es ist explizit nicht das Ziel Fitmaße wie R^2 oder Pseudo- R^2 zu maximieren. Natürlich ist für Prognosemodelle eine bessere Varianzaufklärung als in theorietestenden Anwendungen anzustreben, jedoch besteht in Prognosen auch die Gefahr eines zu genauen Fits. In diesem Fall wird ein Teil des Rauschens in den Daten als Signal fehlinterpretiert. Ein Modell mag zwar die Varianz in den zur Modellschätzung herangezogenen Datenpunkten sehr genau erklären, kann aber gerade deshalb an der Vorhersage zukünftiger Datenpunkte scheitern. Um dies zu vermeiden sollte die Spezifikation eines Modells theoretisch begründbar sein. Als grobe Orientierung sei genannt, dass für Wahlprognosemodelle ein R^2 von 90% nicht unüblich ist und im Bereich der Konfliktforschung in der Regel 80% der Observationen korrekt binär klassifiziert werden können.

Es ist jedoch nicht ausreichend, nur Spezifikationen gegeneinander zu vergleichen, um ein relativ bestes Modell zu finden. Ein Prognosemodell muss natürlich auch genauere Prognosen liefern als deutlich einfachere Verfahren. Soll eine binäre Variable vorhergesagt werden, so bieten sich zufällige Ziehungen aus einer Binomialverteilung an, deren Parameter p sich aus der im Trainingsdatensatz beobachteten Häufigkeitsverteilung der abhängigen Variablen ergibt. In Bezug auf Wahlen kann dies das Setzen des vergangenen Wahlergebnisses oder des Mittelwerts aller vorangegangenen Wahlen als Prognose sein.

Campbell (2004) illustriert diesen Anspruch anschaulich anhand der US-Präsidentenwahlen: ein Prognosemodell sollte einen geringeren MAE als 4,8 Prozentpunkte haben, denn sonst könnte man die Präsidentenwahl auch einfach über den Durchschnitt aller Wahlergebnisse seit 1948 schätzen. Ein Modell muss also besser funktionieren als ein naiver atheoretischer Ansatz.

Zur Spezifikation eines Modells gehört nicht nur die Auswahl relevanter Variablen, sondern auch eine Entscheidung über die Länge der Zeitreihe auf der das Modell geschätzt wird. So weist Sanders (1996) darauf hin, dass der Zusammenhang zwischen gegebenen abhängigen und unabhängigen Variablen in den letzten dreißig Jahren möglicherweise nicht so ein guter Anhaltspunkt für ihre wahrscheinliche Beziehung in den nächsten drei Jahren ist, wie ihre neuere Beziehung. Mit Blick auf das in diesem Abschnitt dargestellte "Time for Change"-Modell bedeutet dies, die bisher im Prinzip unhinterfragte Praxis aus Gründen der Datenverfügbarkeit auf Zeitreihen ab der Präsidentenwahl 1948 zu setzen und zu hinterfragen. Die Wahl des Schätzzeitraums muss jedoch vom konkreten Anwendungsfall und

von der Einschätzung der Analystin abhängen, welche vergangene Zeitperioden der kurz- bis mittelfristigen Zukunft am ehesten ähneln werden.

Prognosewerte sind grundsätzlich mit Unsicherheit behaftet, die sich jedoch mit den zu Grunde liegenden statistischen Methoden quantifizieren lässt. Wird in retrospektiven Analysen viel Wert auf Signifikanztests gelegt, so muss in Prognosen Wert auf die Unsicherheit der Prognose gelegt werden. Die Signifikanz von Koeffizienten ist in Prognosemodellen kein guter Indikator, ob eine Variable in einem Modell inkludiert werden sollte (Armstrong 2012; Ward, Greenhill, und Bakke 2010). Entscheidend ist allein, ob eine Variable zur Verbesserung der Genauigkeit der Out-of-Sample-Prognosen beiträgt.

Die Unsicherheit der Prognose speist sich im Wesentlichen aus zwei Quellen. Zum einen sind die geschätzten Parameter der Regressionsgleichung mit Ungenauigkeit behaftet, welche in Standardfehlern auch quantifiziert werden kann. Diese Unsicherheit in den Koeffizientenschätzer überträgt sich auch in unsichere Prognosen. Zum anderen weichen die aus dem Modell generierten Schätzwerte für vergangene Wahlen natürlich vom tatsächlichen Ergebnis ab - der Fehlerterm des Modells. Beide Quellen der Unsicherheit sind zu berücksichtigen, wenn die Unsicherheit einer Prognose quantifiziert werden soll. Konfidenzintervalle für Prognosewerte lassen sich am einfachsten für eine Vielzahl an verschiedenen Modelltypen per Simulation bestimmen (vgl. King, Tomz, und Wittenberg 2000). Auf dieser Grundlage lässt sich auch einfach die Bandbreite der möglichen Ergebnisse visualisieren. Aus diesen lassen sich wiederum Wahrscheinlichkeiten für verschiedene Ereignisse ableiten.

Bei der Kommunikation des Prognosemodells gegenüber Dritten sollte neben den vorhergesagten Werten auch die Unsicherheit, beispielsweise in Form von Schwankungsbreiten, kommuniziert werden und das dahinterstehende Modell transparent dargestellt werden. Letzteres kann in Form einer klassischen Regressionsstabelle oder auch in Form einer Gleichung geschehen, wie der vorliegende Text zeigt. Außerdem muss die Performanz des Modells in Out-of-Sample-Schätzungen offengelegt werden. Je nach Veröffentlichungsformat, insbesondere bei Veröffentlichungen welche sich nicht an ein Fachpublikum richten, kann es geboten sein, die Zahl der Tabellen oder Visualisierungen auf ein Minimum zu reduzieren. Dennoch sollten diese in geeigneter Form öffentlich zugänglich gemacht werden. Bei wissenschaftlichen Veröffentlichungen ist es im Sinne der intersubjektiven Nachvollziehbarkeit und Reproduzierbarkeit außerdem wünschenswert, wenn die

Daten und der Code zur Analyse öffentlich zugänglich gemacht werden.

Schon aus dieser bewusst sehr einfach gehaltenen beispielhaften Darstellung lassen sich einige Grundprinzipien politikwissenschaftlicher Prognosemodelle festhalten, die der Leserin Orientierung für die Bewertung von Prognosemodellen Dritter und die Entwicklung eigener Modelle geben sollen. Grundlegend für Prognosemodelle ist die Unterscheidung zwischen einem Trainings- und einem Testdatensatz und die damit einhergehende konzeptionelle Unterscheidung eines *forecast* von einer *prediction*. Nur wenn zum Zeitpunkt der Prognose nicht beobachtbare oder bewusst von der Modellschätzung ausgeschlossene Beobachtungen vorhergesagt werden, haben wir es mit einer Prognose zu tun. Dieses Grundprinzip hat auch Implikationen für die Kalibrierung eines Prognosemodells. So ist bei der Auswahl von Variablen einzig ausschlaggebend, ob eine Variable die Genauigkeit der Out-of-Sample-Prognosen verbessert. Ob einzelne Koeffizienten eines Regressionsmodells statistisch signifikant sind, ist dabei nicht von Belang. Ebenso von begrenzter Aussagekraft ist der Fit eines Modells auf dem Trainingsdatensatz. Entscheidend ist allein die Prognose, das heißt die Vorhersage von Variablenwerten im Testdatensatz. Dabei muss ein Prognosemodell bessere Vorhersagen liefern als atheoretische naive Ansätze.

3 Strukturelle Prognosemodelle

Strukturelle Modelle stellen nach wie vor die meistverbreitete Anwendung im Bereich der Prognosemodelle dar. Sie modellieren politische Ereignisse, beispielsweise Wahlergebnisse, als Funktion ausgewählter durch Theorien motivierte Variablen. Die meisten Modelle beschränken sich nicht auf einen einzigen theoretischen Ansatz, sondern kombinieren zwei oder mehrere Ansätze. Viele Wahlprognosemodelle berücksichtigen zudem in Anlehnung an das sogenannte “Michigan-Modell” der Wahlentscheidung zudem eine Mischung aus kurz- (beispielsweise die Wirtschaftslage im Quartal vor der Wahl), mittel- (beispielsweise die Popularität der Präsidentin) und langfristigen Faktoren (beispielsweise vergangene Wahlergebnisse).

Auch in der Konfliktforschung dominieren strukturelle Modelle, die durch verschiedene Theorien zur Erklärung inner- und zwischenstaatlicher Konflikte motiviert sind. Ebenso wie Wahlprognosen kommen in der Konfliktforschung primär Regressionsmodelle zum Einsatz. Bei der Vorhersage von Gerichtsentscheidungen

werden verstärkt auch Verfahren des maschinellen Lernens eingesetzt welche über Klassifikationsverfahren Muster in den Daten identifizieren. Kastellec (2010) erläutert die Anwendung von Entscheidungsbäumen auf Gerichtsentscheidungen anschaulich. Von zentralem Interesse ist in dieser Literatur, ob primär juristische oder auch politische Faktoren Gerichtsentscheidungen erklären und vorhersagen können. Noch beschränkt sich diese Literatur auf den Supreme Court. Sternberg (2017) zeigt aber, dass sich Prognosemodelle auch erfolgreich auf die europäische Verfassungsgerichtsbarkeit anwenden lassen. Sein Modell ist in der Lage, bis zu drei Viertel aller Entscheidungen des deutschen Bundesverfassungsgerichts korrekt vorauszusagen.

Das größte Problem für Wahlforscherinnen, die die Logik US-amerikanischer Modelle auf andere Länder wie beispielsweise Deutschland übertragen wollen, ist die Tatsache, dass diese Modelle für ein Zwei-Parteien-System entwickelt wurden. Die Pioniere für Wahlprognosen in Deutschland, Gschwend und Norpoth (2001), begegneten mit ihrem “Kanzlermodell” dem Problem, indem sie den kombinierten Stimmanteil der regierenden Parteien vorhersagten. Bei ihrer ersten Prognose 2002 war das die Summe der auf SPD und Bündnis 90/Die Grünen entfallenden Stimmen.

Auch in den USA ist die Prognose des *two party vote share* aber nur so lange aussagekräftig, wie die Siegerin des *popular vote* auch eine Mehrheit im *Electoral College* erzielt. Zuletzt gewann der im *popular vote* unterlegene republikanische Kandidat Trump die US-Präsidentschaftswahl 2016 durch eine Mehrheit im *electoral college*, obwohl die Gegenkandidatin Hillary Clinton landesweit mehr Stimmen erhielt. Insofern ist Norpoths (2016) korrekte Vorhersage des Wahlsiegs Trumps nicht ganz richtig, da er ihn fälschlich als Sieger des *popular vote* sah.

Um diesem Problem zu begegnen, entwickelten Politikwissenschaftlerinnen auch Modelle, die auf einem Panel an Staaten statt einer einzigen nationalen Zeitreihe beruhten (beispielsweise Klarner 2008). Mit diesem Ansatz lässt sich zugleich einem häufigen Problem struktureller Modelle begegnen: die geringe Größe der Stichprobe aufgrund der begrenzten Anzahl vorangegangener Wahlen. In einem solchen Modell lassen sich mehr Variablen berücksichtigen. Zugleich ist es damit nur ein kurzer Schritt hin zu Prognosen in Mehrparteiensystemen: statt die Stimmanteile einer Partei in mehreren Staaten, werden die Stimmanteile mehrerer Parteien prognostiziert (Foucault 2012).

4 Aggregationsmodelle

Aggregationsmodelle fassen verschiedene Datenquellen zu einer Prognose zusammen. Aggregationsmodelle der Wahlforschung beispielsweise fassen die von den Instituten berechneten Anteilswerte der in den Umfragen geäußerten Wahlintentionen zu einer Prognose zusammen. Im einfachsten Fall kann dies ein einfacher Mittelwert aller Umfragen zu einem Zeitpunkt sein. Dadurch können die, sich aus einer Zufallsstichprobe natürlich ergebenden, Variationen des Stimmanteils der Parteien ausgeglichen werden. Wenn verschiedene Institute jedoch den gleichen systematischen Biases gegenüber bestimmten Parteien unterliegen dann findet sich dieser auch in der Aggregation wieder. Der Beitrag Selb und Munzerts (2015) geht über eine einfache Mittelwertbildung hinaus. Sie betrachten ausschließlich den Zusammenhang zwischen Umfragewerten und Ergebnissen vorangegangener Wahlen für die großen Parteien mittels eines Mehrebenenmodells und extrapolieren diesen zu einer Prognose.

Eine interessante Anwendung stellt die “equally weighted regression” dar. Sie erlaubt es, alle relevanten unabhängigen Variablen in ein Modell aufzunehmen, selbst wenn die Zahl der Variablen die Zahl der Beobachtungen in der Stichprobe übersteigt. Dazu werden alle unabhängigen Variablen z -transformiert und so rekodiert, dass sie positiv mit der abhängigen Variablen korrelieren. Die Summe dieser Variablen fungiert dann als einzige unabhängige Variable in einem Regressionsmodell. Graefe (2015b) erläuterte dieses Verfahren ausführlicher und zeigt, dass es für US-Präsidentenwahlen genauere Vorhersagen liefert als einzelne strukturelle Modelle.

Die britische Zeitschrift *The Economist* fasste 1992 erstmals die Umfragen mehrerer Institute in einem sogenannten “poll of the polls” zusammen, was in der Folge von vielen Medien kopiert wurde (Blumenthal 2014). Diese Form der Umfrageaggregation wird in den Medien sogar häufiger betrieben als in der Wissenschaft. Die meisten Medien legen ihre zumeist recht simplen Methodologien und Datenquellen offen. Damit folgen sie den Prinzipien der intersubjektiven Nachvollziehbarkeit und Reproduzierbarkeit, auch wenn sie mit ihren Aggregationsmodellen keine wissenschaftlichen Zwecke verfolgen. Aggregationsmodelle haben gegenüber strukturellen Modellen, was die öffentliche Aufmerksamkeitsökonomie angeht, den Vorteil, dass sie quasi täglich mit jeder neu erscheinenden Umfrage aktualisiert werden können. Weltweite Bekanntheit erlangte die Umfrageaggregation insbesondere durch Nate Silvers Blog “FiveThirtyEight”, welcher

seit 2008 Prognosen für jede US-Präsidentschaftswahl liefert. Er fasste Umfragen aus allen Staaten zusammen und leitet daraus Wahrscheinlichkeiten für einen Wahlsieg der verschiedenen Kandidatinnen ab. Das dabei zur Anwendung kommende Verfahren legt Silver (2016) nicht komplett offen. Jedoch fließt in diese Berechnung auch ein strukturelles Modell ein, sodass es sich bei “FiveThirtyEight” strenggenommen um ein synthetisches Modell handelt, welche ich im nächsten Abschnitt bespreche.

5 Synthetische Modelle

Synthetische Modelle kombinieren verschiedene Prognosemodelle in einem Modell. An der Schnittstelle zwischen Aggregationsmodellen und synthetischen Modellen befindet sich Andreas Graefes (2017) *PollyVote* welcher vier Kategorien von Prognosen - Umfragen, Prognosemärkte, Expertenurteile und strukturelle Prognosemodelle - kombiniert. Die Idee hinter Aggregationsmodellen ist, dass keine einzelne Methode durchweg besser als andere Methoden ist (Graefe u. a. 2014). So zeigt Graefe (2015a) anhand des *PollVote*, dass der Prognosefehler eines Durchschnitts mehrerer Vorhersagen oftmals kleiner als die Abweichungen der einzelnen Prognosen vom tatsächlichen Ergebnis.

In synthetischen Modellen werden die Gewichte jedoch modellbasiert aus den zugrundeliegenden Daten geschätzt. In synthetischen Wahlprognosen verhält es sich oft so, dass ein strukturelles Modell quasi eine Basisprognose liefert, die dann über die Zeit durch die Umfragekomponente aktualisiert wird. In der Regel bekommen die Umfragen mehr Gewicht zugewiesen, je näher die Wahl rückt. Dies ist beispielsweise der Fall in Linzers (2013) Bayesschen Prognosemodell für US-Präsidentschaftswahlen, das ein strukturelles Modell mit Umfragen verbindet. Munzert u. a. (2017) adaptieren mit “Zweistimme.org” Linzers Ansatz für das deutsche Parteiensystem. Lewis-Beck, Nadeau, und Bélanger (2016) berechnen ein synthetisches Modell in drei Schritten. Zunächst wird ein einfaches strukturelles Modell und ein Modell, das den Zusammenhang zwischen Wahlergebnissen und Umfragewerten schätzt. Die vorhergesagten Werte beider Modelle fließen dann als unabhängige Variablen in ein drittes Regressionsmodell, dem synthetischen Modell, ein, das die Prognose generiert. Ein ähnlich zugängliches synthetisches Modell wurde von Theresa Küntzler (2017) zur Bundestagswahl 2013 entwickelt.

Die Vorzüge der Kombination vieler verschiedener Prognosen sind also genauere

Vorhersagen und die Möglichkeit der laufenden Aktualisierung dieser. Viele der neueren Entwicklungen im Bereich der Wahlprognosen finden deshalb in diesem Bereich statt. In anderen Bereichen der Politikwissenschaft ist die Kombination verschiedener Prognosemethoden noch nicht so verbreitet, was auch damit zusammenhängen mag, dass in diesen Bereichen eine natürliche alternative Quelle für Prognosen, wie sie Umfragen für die Wahlforschung darstellen, fehlt. Grundsätzlich ist aber die Kombination mehrerer struktureller Modelle überall möglich wo mehr als ein Prognosemodell existiert. In den folgenden Abschnitten stelle ich abschließend zwei weitere wissenschaftlich fundierte Prognosemethoden vor.

6 “Wisdom of the crowd”-Ansätze

Dies ist zum einen die Idee des “Wisdom of the crowd” welche anhand des Beispiels einer Wahlerwartungsbefragung erläutert werden soll. Anders als die klassische Sonntagsfrage fragt diese nicht nach der individuellen Wahlintention, sondern danach, welchen Wahlausgang, der oder die Befragte erwartet. Dieser Ansatz wurde in den 1980er Jahren ebenfalls im Kontext der US-Präsidentchaftswahlen entwickelt (Lewis-Beck und Skalaban 1989) und seither sehr erfolgreich genutzt. So wertet Graefe (2014) insgesamt 217 Wahlerwartungsumfragen zu US-Präsidentchaftswahlen aus und kommt zu dem Schluss, dass 193 von ihnen den Wahlsieger korrekt vorher sagten. Murr (2011) zeigt, dass die Wahlerwartungsumfrage ein erfolgreicher Anwendungsfall von Condorcets Jury Theorem³ ist: obwohl nur 69% der Befragten der “British Election Study Internet Survey” 2010 die Gewinnerin ihres Wahlkreises korrekt voraussagen, prognostiziert eine Mehrheit der Befragten aus demselben Wahlkreis in 86 Prozent aller Wahlkreise korrekt die Gewinnerin.

Eine Variante der Wahlerwartungsumfrage ist die Expertinnenbefragung, wie sie beispielsweise Graefe für seinen PollyVote durchführt. Die gegenüber Bevölkerungsumfragen geringere Größe der Stichprobe soll durch die besondere Expertise der Befragten überkompensiert werden. Als Expertinnenbefragung kommt der “Wisdom of the crowd”-Ansatz auch in anderen Teilbereichen der Politikwissenschaft zum Einsatz. So zum Beispiel bei der Vorhersage von Entscheidungen des US-amerikanischen Supreme Courts. Die befragten Expertinnen schnitten dabei mit 59,1% korrekt vorhergesagten Entscheidungen jedoch deutlich schlechter ab

³Für eine Erläuterung des Jury Theorems siehe Nurmi (2002).

als ein statistisches Modell, das 75% der Entscheidungen korrekt vorhersagte (Martin u. a. 2004).

7 Prognosemärkte

Politische Ereignisse sind auch Gegenstand von Wetten. So erlauben es insbesondere britische Wettanbieter der interessierten Öffentlichkeit, auf den Ausgang diverser politischer Ereignisse, zu wetten. In Bezug auf politische Wetten könnte man sagen, dass die aus Prognosemodellen generierten Wahrscheinlichkeiten Orientierung für Wetten bieten. So gesehen sind Wettquoten, welche üblicherweise als Odds⁴ angegeben werden, auch Prognosen des Wahlausgangs. Hierbei ist in der Regel jedoch völlig intransparent, wie Wettanbieter die Odds für die von ihnen angebotenen Wetten bestimmen.

Wettanbieter sind oft auch Anbieter von Prognosemärkten, die aber auch von Forscherinnen sowie anderen Akteurinnen, wie zum Beispiel Zeitungen, angeboten werden. Die Teilnehmerinnen an solchen Märkten setzen jedoch nicht auf durch Wettanbieter festgesetzte Quoten, sondern es findet eine freie Preisbildung über das Angebot- und Nachfrageverhalten der Nutzerinnen statt. Auf einem Prognosemarkt handeln die Teilnehmenden beispielsweise fiktive Anteilsscheine der zur Wahl stehenden Parteien oder Personen, deren Preis die Prognose des Marktes bildet. Nach der Wahl erfolgt die Auszahlung der Marktteilnehmenden auf der Basis des Wahlergebnisses.⁵ Die Erwartung ist, dass die Händlerinnen in ihren Entscheidungen alle relevanten Informationen berücksichtigen, da diese Geld gewinnen oder eben auch verlieren können. Deshalb sollte auch der Einfluss persönlicher politischer Überzeugungen gegenüber den Wahlerwartungsumfragen reduziert sein. Prognosemärkte wurden, nachdem der “Iowa Presidential Stock Market” den Sieg Bushs bei der US-Präsidentschaftswahlen 1988 erfolgreich vorhersagte (Forsythe u. a. 1989), auch zu einem Thema in der Politikwissenschaft.

Auch in anderen politikwissenschaftlichen Forschungsbereichen können Prognosemärkte Anwendung finden. Ein im Rahmen eines Forschungsprojektes des US-Verteidigungsministeriums entwickelter Prognosemarkt für politische Ereig-

⁴Die Odds einer Wette auf das Eintreten eines Ereignisses E entsprechen der Wahrscheinlichkeit, dass das Ergebnis nicht eintritt relativ zur Wahrscheinlichkeit, dass es eintritt: $\frac{1-p(E)}{p(E)}$. Hieraus ergibt sich auch der Auszahlungsbetrag.

⁵Für eine genauere Erläuterung der Funktionsweise von Prognosemärkten und einigen Beispielen siehe (Wolfers und Zitzewitz 2004).

nisse, unter anderem für Terroranschläge, im Nahen Osten wurde aber nach breiter öffentlicher Empörung über das Vorhaben vor Inbetriebnahme eingestellt (Meirowitz und Tucker 2004). Für Forscherinnen ist es relativ einfach, eigene Prognosemärkte zu entwickeln. Hierfür gibt es mehrere unter freien Lizenzen verfügbare Softwarelösungen, beispielsweise Augur (Peterson u. a. 2015).

8 Potenziale und Problemfelder politikwissenschaftlicher Prognosemodelle

Regressionsmodelle sind das vorherrschende Mittel zur Erstellung von Prognosen. Solche Modelle stoßen jedoch in Situationen mit wenigen Datenpunkten und vielen potenziellen Variablen an ihre Grenzen. Letzterer Fall ist typisch für den Bereich der Wahlprognose, aber auch in anderen Bereichen in denen es um kurze Zeitreihen geht (beispielsweise Bechtel und Leuffen 2010). Es stellt daher eine nicht zu unterschätzende Herausforderung dar, die richtige Balance zwischen theoretischer Spezifikation und empirischer Anpassung zu finden.

Weiterhin darf kritisch hinterfragt werden, ob die in einem statistischen Modell geschätzten Zusammenhänge sich so auch in Zukunft zeigen werden. Prognosemodelle unterliegen nämlich zwei wesentlichen Annahmen. Erstens, dass kein politikwissenschaftlich relevantes Einzelphänomen wirklich einzigartig ist, und zweitens, dass die Bedeutung verschiedener, dieses Phänomen beeinflussender, Faktoren über die Zeit hinweg konstant bleibt. Urheberinnen struktureller Modelle werden einwenden, dass ihre Modelle durch politikwissenschaftliche Theorien motiviert sind. Diese Theorien, gestützt durch empirische Forschung, begründen, dass es sich um dauerhaft geltende Zusammenhänge handelt.

In welchem Ausmaß Prognosemodelle theoretisch motiviert sein müssen ist nicht unumstritten. Lewis-Beck (2005) sieht keinen großen Unterschied zwischen der Anfertigung eines Prognosemodells und klassischer Forschungsarbeit: Es gehe darum, Theorien zu Rate zu ziehen, diese in einer Schätzgleichung auszudrücken, Daten zu sammeln, die Gleichung zu schätzen und die empirische Passung zu prüfen. Er zeigt sich überzeugt, dass ein theoretisch motivierter Modellansatz auf lange Sicht eine bessere Leistung erbringen wird. Van der Eijk (2005) stimmt ihm darin im Prinzip zu, kritisiert aber, dass nur die wenigsten Wahlprognosen diesen Anspruch einlösen. So sei beispielsweise eine der Schlüsselvariablen vieler Modelle,

die Popularität der Amtsinhaberin, eher ein Platzhalter für alle möglichen Phänomene als eine theoretisch wohl spezifizierte unabhängige Variable. Die Debatte, ob Prognosen in der Politikwissenschaftlich grundsätzlich sinnvoll sind, ist aber nach Meinung Bechtels und Leuffens (2010) zu Gunsten der Prognosemodelle beendet.

Umstritten ist, ob Modellanpassungen im Falle vom tatsächlichen Ergebnis stark abweichenden Prognosen, zulässig sind. Colomer (2007) sieht Modellanpassungen sehr kritisch, da diese meist weder theoretisch begründet seien, noch empirisch eine Verbesserung mit sich brächten. Ebenso sieht Campbell (2004) Modellstabilität, also das unveränderte Bestehen eines Modells über mehrere Wahlen, als ein wichtiges Bewertungskriterium für Prognosen an. Lewis-Beck und Rice (1992) hingegen sind der Meinung, dass mit jeder Wahl neue Fehlerquellen identifiziert werden können, und das Modell durch Revision verbessert werden kann.

Die stetig wachsende Zahl der Prognosemodelle führt zwangsläufig auch zu der Frage, welches Prognosemodell nun am besten sei. Diese Frage lässt sich nicht beantworten. Zum einen ist die Zahl der durch mehrere Modelle wiederholt vorhergesagten Ereignisse noch zu gering. Zum anderen ist die genaueste ist nicht notwendigerweise auch die beste Vorhersage. Ein Beispiel aus der Wahlforschung soll dies illustrieren: Eine Prognose, die drei Monate vor der Wahl abgegeben wurde, und im Durchschnitt drei Prozentpunkte vom Wahlergebnis abweicht, kann durchaus als besser bezeichnet werden im Vergleich zu einer Prognose, welche zwar nur um durchschnittlich zwei Prozentpunkte danebenliegt, aber nur zwei Wochen vor der Wahl getroffen wurde. In der Konfliktforschung werden Vorlaufzeiten von mindestens einem Jahr für Prognosemodelle angestrebt, da die Vorhersagen sonst für die Außen- und Sicherheitspolitik nur wenig Wert hätten. Auch wenn eine entsprechende Meta-Studie zu diesem Thema noch aussteht, so ist davon auszugehen, dass es einen Trade-Off zwischen der Vorlaufzeit und der Genauigkeit eines Modells gibt.

Lewis-Beck (2005) schlägt neben diesen beiden Kriterien auch Sparsamkeit und Reproduzierbarkeit als Kriterien zur Bewertung von Prognosemodellen vor. Sparsamkeit bedeutet, dass nur solche Variablen inkludiert werden, die zweifelsfrei einen Beitrag zur Verbesserung der Prognose liefern. Auch rein praktische Erwägungen sprechen für sparsame Modelle, wenn die einem Modell zugrunde liegende Stichproben sehr klein sind. Zu guter Letzt sollte jedes Modell reproduzierbar sein. Das heißt, die ursprüngliche Autorin, wie auch jede andere

methodisch hinreichend qualifizierte Person, sollte in der Lage sein, das Modell anzuwenden und zu den gleichen Ergebnissen kommen.

Die Genauigkeit eines Modells ist von zentraler Bedeutung. Es ist jedoch, gerade in Bezug auf strukturelle Modelle, anzumerken, dass diese auch dann noch nützlich sein können, wenn sie mal “daneben liegen”. Strukturelle Modelle erfüllen die wichtige Funktion, Erwartungen aufzustellen, mit denen die tatsächlichen Resultate schließlich verglichen werden können. Wahlprognosen beispielsweise sagen im Grunde voraus, wie ein durchschnittliche Kandidatin mit einer durchschnittlichen Kampagne und Opposition in der prognostizierten Wahl abschneiden wird. Prognosen helfen uns damit also, das Außergewöhnliche an einzelnen Ereignissen, sei es ein Wahlergebnis, ein Regierungsumsturz oder ein Gerichtsentscheid, zu erkennen.

Strukturelle Modelle können so auch zur Theoriebildung beitragen. Ein Beispiel soll dies illustrieren. Das Kanzlermodell wurde 2005 und 2013 durch das Aufkommen neuer Parteien besonders herausgefordert. 2005 korrigierten Gschwend und Norpoth (2005) für ihre Prognose die Variable Kanzlerpopularität um den erwartenden Stimmanteil der Partei Die Linke (damals noch Die Linke/PDS) und lieferten so eine ziemlich genaue Prognose. 2013 nahmen sie keine solche Korrektur vor und lagen mit ihrer Prognose erstmals deutlicher daneben was sie unter anderem mit dem Auftreten der AfD erklärten (Norpoth und Gschwend 2014). Daraus lässt sich ableiten, dass die Popularität des Regierungschefs in einem fraktionalisierten Parteiensystem weniger Vorhersagekraft für den Wahlentscheid besitzt. Diese Hypothese, abgeleitet aus einem Prognosemodell, lässt sich in einem klassischeren Forschungsdesign systematisch testen. Tatsächlich war die Motivation hinter den ersten Wahlprognosen, theoretische Erwartungen zum Effekt der Makroökonomie auf Wahlentscheidungen zu testen (Fair 1978; Hibbs 1982). Mayer (2014) hebt dementsprechend die Beiträge, die Prognosemodelle zu unserem Verständnis von Wahlen im Allgemeinen und der Wirkung der wirtschaftlichen Lage auf Wahlen im Besonderen geleistet haben, hervor.

Prognosemodelle in der Konfliktforschung werden explizit mit dem Ziel entwickelt, die vorhergesagten Ereignisse wie Kriege oder Genozide zu verhindern. Es liegt also die Frage nahe, ob Prognosen die Wahrscheinlichkeit des Eintretens der Phänomene, die sie vorhersagen, beeinflussen. Für Wahlprognosen ist von einer solchen Wirkung, zumindest in Deutschland, bisher noch nicht auszugehen - dafür finden sie noch zu wenig öffentliche Beachtung. In den USA kann über

einen möglichen Einfluss der 2016 besonders prominent wahrgenommenen Aggregationsmodelle, die einen vermeintlichen sicheren Sieg Clintons vorhersahen, nur spekuliert werden. Auch im Bereich der Konfliktforschung lassen sich zu diesem Aspekt nur Spekulationen anstellen, da Art, Umfang und Nutzung von Prognosen in der Sicherheits- und Außenpolitik in der Regel nicht öffentlich gemacht werden.

9 Konklusion

Prognosemodelle erfreuen sich in der Politikwissenschaft wachsender Verbreitung. Nachdem Wahlprognosemodelle erstmals in den 1980er Jahren in den USA im Kontext der Präsidentschaftswahlen, ebenso wie Jahrzehnte zuvor auch Meinungsumfragen, entwickelt wurden, haben sie sich mittlerweile stark verbreitet. Befanden sich Wissenschaftliche Wahlprognosen zunächst in der paradoxen Situation, dass sie in der Öffentlichkeit vielleicht stärkere Beachtung erfuhren als in der Fachöffentlichkeit, so haben sie sich langsam in der Politikwissenschaft etabliert. So wurde unlängst ein Artikel zu Wahlprognosen in dem hoch angesehenen interdisziplinären aber überwiegend naturwissenschaftlichen Fachjournal “Science” veröffentlicht (Kennedy, Wojcik, und Lazer 2017). Auch im Bereich der Konfliktforschung haben Prognosen seit Jahrzehnten eine Tradition.

Dieser Beitrag bot eine praxisorientierte Einführung in das Feld der politikwissenschaftlichen Prognosen. Dazu wurden zunächst deren Grundlagen anhand eines kanonischen Modells dargestellt und anschließend ein Überblick über verschiedenste Ansätze gegeben sowie einige Aspekte diskutiert, die Anhaltspunkte für die zukünftige Entwicklung des Feldes geben können. Neben Wahlen stellen politische Gewalt und Gerichtsentscheide die häufigsten Anwendungsfälle für Prognosemodelle in der Politikwissenschaft dar. Aber auch in anderen Bereichen der Politikwissenschaft sind Prognosen grundsätzlich möglich. Die Anwendungsfelder reichen von der Vorhersage der legislativen Aktivität des Europäischen Parlaments (Bechtel und Leuffen 2010) bis hin zur Zinspolitik der Zentralbanken (Krueger und Kuttner 1996).

Konzeptionell ist es kein weiter Schritt von der Erklärung zur Prädiktion. Erstere zielt darauf ab, allgemeine Muster, und damit letztendlich kausale Mechanismen, aufzudecken. Auf der Basis kausaler Theorien und sie abbildender empirischer Modelle sind Vorhersagen grundsätzlich möglich. Im Prinzip hat jedes politik-

wissenschaftliche Modell auch testbare Implikationen für die Zukunft.

Auch empirisch ist es kein weiter Schritt von der retrospektiven Erklärung zur Prognose. Aus einem jeden Regressionsmodell lassen sich ganz natürlich nicht nur Koeffizienten, sondern auch vorhergesagte Werte ableiten. So lässt sich aus jeder quantitativen Anwendung eine Prognose entwickeln. Damit sind Prognosen prinzipiell auf jedes andere politikwissenschaftliche Forschungsfeld anwendbar.

In den Bereichen der Politikwissenschaft, in denen sich Prognosen bereits etabliert haben, bestimmen vor allem methodische Neuerungen, wie bayessche Verfahren, Methoden aus dem Machine Learning sowie auch der Einbezug von Social-Media-Daten die Entwicklung.⁶ In anderen Teilbereichen der Disziplin können aber noch mit deutlich simpleren Methoden neue und relevante Beiträge geleistet werden. Meine Prognose für die Zukunft der Prognosen ist daher, dass wir auch in Zukunft interessante Weiterentwicklungen auf diesem Feld beobachten dürfen und dass Prognosemodelle weiterhin ein fruchtbares Feld für politikwissenschaftliche Forschung darstellen werden.

10 Kommentiertes Literaturverzeichnis

Gschwend, Thomas und Helmut Norpoth (2001): “Wenn am nächsten Sonntag . . .”: Ein Prognosemodell für Bundestagswahlen. In Wahlen Und Wähler: Analysen Aus Anlass Der Bundestagswahl 1998, edited by Hans-Dieter Klingemann and Max Kaase, 473–99. Wiesbaden: Westdeutscher Verlag.

Gschwend und Norpoths “Kanzlermodell” besteht aus einem einfachen Zeitreihenmodell mit drei unabhängigen Variablen. Die Erstveröffentlichung ihres Modells stellt eine zugängliche Darstellung eines Anwendungsfalls struktureller Modelle dar.

Selb, Peter und Simon Munzert (2015): Forecasting the 2013 German Bundestag Election Using Many Polls and Historical Election Results. German Politics 25(1): 73–83.

Selb und Munzert betrachten den Zusammenhang zwischen Umfragewerten und

⁶Während bayesianische Verfahren und Ansätze aus dem Machine Learning in ersten Anwendungen vielversprechende Ergebnisse liefern hat was das Potenzial von Daten aus dem Internet, insbesondere Social Media, angeht bereits Ernüchterung eingesetzt (Huberty 2015).

Ergebnissen vorangegangener Wahlen für die großen Parteien mittels eines Mehrebenenmodells. Sie erhoffen sich durch ihr Modell partei- oder institutsspezifische Verzerrungen zu eliminieren und so zu einer genaueren Prognose zu gelangen als einzelne Umfragen. Ihr Ansatz kann grundsätzlich auch für die Fortschreibung politikwissenschaftlicher Zeitreihen aus mehrere Quellen nutzbar gemacht werden.

Küntzler, Theresa (2017): Using Data Combination of Fundamental Variable-Based Forecasts and Poll-Based Forecasts to Predict the 2013 German Election. German Politics.

Küntzlers synthetisches Modell kombiniert Norpoth und Gschwends “Kanzlermodell” mit einem eigenen Umfrageaggregationsmodell. Um zu einer finalen Prognose zu gelangen, kalkuliert Küntzler ein mit der jeweiligen Unsicherheit gewichtetes Mittel aus den beiden Einzelvorhersagen. Das zu Grunde liegende statistische Verfahren ist grundsätzlich auch auf Anwendungen jenseits der Wahlprognosen anwendbar.

Murr, Andreas (2011): Wisdom of Crowds’? A Decentralised Election Forecasting Model That Uses Citizens’ Local Expectations. Electoral Studies 30(4): 771–83.

Murr zeigt, dass unter Anwendung von Condorcets Jury Theorem sich mittels des “wisdom of the crowd” die Wahlkreisergebnisse der britischen Unterhauswahlen 2010 auf Basis einer Wahlerwartungbefragung relativ gut vorhersagen lassen. Das von Murr zur Anwendung gebrachte Verfahren lässt sich auch auf Bevölkerung- und Expertinnenbefragungen zu anderen Themen anwenden.

Berg, Joyce, Robert Forsythe, Forrest Nelson, und Thomas Rietz (2008): Results from a Dozen Years of Election Futures Markets Research. In Handbook of Experimental Economics Results, 742–51. Elsevier.

Robert Forsythe, Gründer der Iowa Electronic Markets, und Ko-Autoren beschreiben die Funktionsweise der bekannten The Iowa Political Markets und zeigen, dass diese eine Vielzahl an Wahlen relativ genau vorhersagen konnten und dabei mindestens ebenso gut abschneiden wie Wahlumfragen. Im Fokus der Iowa Political Markets stehen Wahlen, es ist aber prinzipiell möglich für jegliche andere politikwissenschaftlich relevante Phänomene Prognosemärkte einzurichten.

Literaturverzeichnis

- Abramowitz, Alan I. 2012. „Forecasting in a Polarized Era: The Time for Change Model and the 2012 Presidential Election“. *PS: Political Science & Politics* 45 (04): 618–19. http://www.journals.cambridge.org/abstract_S104909651200087X.
- Armstrong, J. Scott. 2012. „Illusions in regression analysis“. *International Journal of Forecasting* 28 (3): 689–94. <http://linkinghub.elsevier.com/retrieve/pii/S0169207012000246>.
- Bechtel, Michael M., und Dirk Leuffen. 2010. „Forecasting European Union politics: Real-time forecasts in political time series analysis“. *European Union Politics* 11 (2): 309–27. <http://journals.sagepub.com/doi/10.1177/1465116509360846>.
- Berg, Joyce, Robert Forsythe, Forrest Nelson, und Thomas Rietz. 2008. „Results from a Dozen Years of Election Futures Markets Research“. In *Handbook of Experimental Economics Results*, 1:742–51. Elsevier. doi:10.1016/S1574-0722(07)00080-7.
- Blumenthal, Mark. 2014. „Polls, Forecasts, and Aggregators“. *PS: Political Science & Politics* 47 (02): 297–300. http://journals.cambridge.org/article_S1049096514000055.
- Campbell, James E. 2004. „Introduction - The 2004 presidential election forecasts“. *PS: Political Science & Politics* 37 (4): 733–35.
- Colomer, Josep M. 2007. „What other sciences look like“. *European Political Science* 6 (2): 134–42. <http://link.springer.com/10.1057/palgrave.eps.2210119>.
- Eijk, Cees van der. 2005. „Election forecasting: A sceptical view“. *The British Journal of Politics & International Relations* 7 (2): 210–14.
- Erikson, Robert S., und Christopher Wlezien. 2012. *The Timeline of Presidential Elections: How Campaigns Do (and Do Not) Matter*. Chicago: University of Chicago Press.
- . 2014. „Forecasting US Presidential Elections Using Economic and Noneconomic Fundamentals“. *PS: Political Science & Politics* 47 (02): 313–16. http://journals.cambridge.org/article_S1049096514000092.
- Fair, Ray C. 1978. „The Effect of Economic Events on Votes for President“. *The Review of Economics and Statistics* 60 (2): 159. [http:](http://)

//www.jstor.org/discover/10.2307/1924969?uid=3737864&uid=2129&uid=2&uid=70&uid=4&sid=21102510084707.

Forsythe, R., F. Nelson, G. Neumann, und J. Wright. 1989. „The Iowa presidential stock market: A field experiment“. In *Research in Experimental Economics*, herausgegeben von R. Issac. Bd. 4. Westport: JAI Press.

Foucault, Martial. 2012. „Forecasting the 2012 French legislative election“. *French Politics* 10 (1): 68–83. <https://link.springer.com/article/10.1057/fp.2012.2>.

Gleditsch, Kristian Skrede, und Michael D Ward. 2013. „Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes“. *Journal of Peace Research* 50 (1): 17–31. <http://journals.sagepub.com/doi/10.1177/0022343312449033>.

Graefe, Andreas. 2014. „Accuracy of Vote Expectation Surveys in Forecasting Elections“. *Public Opinion Quarterly* 78 (S1): 204–32. <https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfu008>.

———. 2015a. „German election forecasting: Comparing and combining methods for 2013“. *German Politics* 24 (2): 195–204. <http://www.tandfonline.com/doi/abs/10.1080/09644008.2015.1024240>.

———. 2015b. „Improving forecasts using equally weighted predictors“. *Journal of Business Research* 68 (8): 1792–9. <http://linkinghub.elsevier.com/retrieve/pii/S0148296315001563>.

———. 2017. „The PollyVote’s Long-Term Forecast for the 2017 German Federal Election“. *PS: Political Science & Politics* 50 (03): 693–96. https://www.cambridge.org/core/product/identifier/S1049096517000439/type/journal_article.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones Jr., und Alfred G. Cuzán. 2014. „Combining forecasts: An application to elections“. *International Journal of Forecasting* 30 (1): 43–54. <http://www.sciencedirect.com/science/article/pii/S0169207013000423>.

Gschwend, Thomas, und Helmut Norpoth. 2001. „Wenn am nächsten Sonntag ∴ Ein Prognosemodell für Bundestagswahlen“. In *Wahlen und Wähler: Analysen aus Anlass der Bundestagswahl 1998*, herausgegeben von Hans-Dieter Klingemann und Max Kaase, 473–99. Wiesbaden: Westdeutscher Verlag.

———. 2005. „Prognosemodell auf dem Prüfstand: Die Bundestagswahl 2005“.

Politische Vierteljahresschrift 46 (4): 682–88. <http://www.springerlink.com/index/GW540132632G0MV0.pdf>.

Hibbs, Douglas A. 1982. „President Reagan’s Mandate from the 1980 Elections: A Shift to the Right;“ *American Politics Research* 10 (4): 387–420. <http://apr.sagepub.com/cgi/doi/10.1177/004478082010004001>.

Huberty, Mark. 2015. „Can we vote with our tweet? On the perennial difficulty of election forecasting with social media“. *International Journal of Forecasting* 31 (3): 992–1007. <http://linkinghub.elsevier.com/retrieve/pii/S0169207014001058>.

Kastellec, Jonathan P. 2010. „The statistical analysis of judicial decisions and legal rules with classification trees“. *Journal of Empirical Legal Studies* 7 (2): 202–30.

Kennedy, Ryan, Stefan Wojcik, und David Lazer. 2017. „Improving election prediction internationally“. *Science* 355 (6324): 515–20. <http://science.sciencemag.org/content/355/6324/515.abstract>.

King, Gary, Michael Tomz, und Jason Wittenberg. 2000. „Making the most of statistical analyses: Improving interpretation and presentation“. *American journal of political science*, 347–61.

Klarner, Carl. 2008. „Forecasting the 2008 U.S. House, Senate and Presidential Elections at the District and State Level“. *PS: Political Science and Politics* 41 (4): 723–28.

Krueger, Joel T., und Kenneth Kuttner. 1996. „The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy“. *The Journal of Futures Markets* 16 (8): 865–79.

Küntzler, Theresa. 2017. „Using Data Combination of Fundamental Variable-Based Forecasts and Poll-Based Forecasts to Predict the 2013 German Election“. *German Politics* 0 (0): 1–19. <http://dx.doi.org/10.1080/09644008.2017.1280781>.

Leininger, Arndt. 2015. „Wissenschaftliche Wahlprognosen–Alternative oder Ergänzung zu Umfragen;“ *Zeitschrift für Parlamentsfragen* 46 (4): 675–91. <http://www.nomos-elibrary.de/10.5771/0340-1758-2015-4-675/wissenschaftliche-wahlprognosen-alternative-oder-ergaenzung-zu-umfragen-jahrgang-46-2015-heft-4>.

Lewis-Beck, Michael S. 2005. „Election forecasting: Principles and practice“. *The British Journal of Politics & International Relations* 7 (2): 145–64. <http://www.jstor.org/stable/2645412>.

[//onlinelibrary.wiley.com/doi/10.1111/j.1467-856X.2005.00178.x/full](http://onlinelibrary.wiley.com/doi/10.1111/j.1467-856X.2005.00178.x/full).

Lewis-Beck, Michael S., Richard Nadeau, und Éric Bélanger. 2016. „The British general election: Synthetic forecasts“. *Electoral Studies* 41 (Supplement C): 264–68. <http://www.sciencedirect.com/science/article/pii/S0261379415002176>.

Lewis-Beck, Michael S., und Tom W. Rice. 1992. *Forecasting elections*. Washington, D.C.: CQ Press.

Lewis-Beck, Michael S., und Andrew Skalaban. 1989. „Citizen Forecasting: Can Voters See into the Future?“. *British Journal of Political Science* 19 (1): 146–53. <https://www.cambridge.org/core/journals/british-journal-of-political-science/article/citizen-forecasting-can-voters-see-into-the-future/997F71FA348ED48B613CDAC3138CE7EA>.

Linzer, Drew A. 2013. „Dynamic Bayesian Forecasting of Presidential Elections in the States“. *Journal of the American Statistical Association* 108 (501): 124–34. <http://www.tandfonline.com/doi/abs/10.1080/01621459.2012.737735>.

Martin, Andrew D., Kevin M. Quinn, Theodore W. Ruger, und Pauline T. Kim. 2004. „Competing approaches to predicting supreme court decision making“. *Perspectives on Politics* 2 (4): 761–67.

Mayer, William G. 2014. „What, If Anything, Have We Learned from Presidential Election Forecasting?“. *PS: Political Science & Politics* 47 (02): 329–31. http://www.journals.cambridge.org/abstract_S1049096514000134.

Meirowitz, Adam, und Joshua A. Tucker. 2004. „Learning from Terrorism Markets“. *Perspectives on Politics* 2 (02). http://www.journals.cambridge.org/abstract_S1537592704040204.

Munzert, Simon, Lukas Stötzer, Thomas Gschwend, Marcel Neunhoffer, und Sebastian Sternberg. 2017. „Zweitstimme.org. Ein strukturell-dynamisches Vorhersagemodell für Bundestagswahlen“. *Politische Vierteljahresschrift* 58 (3): 418–41. <https://www.nomos-elibrary.de/index.php?doi=10.5771/0032-3470-2017-3-418>.

Murr, Andreas Erwin. 2011. „Wisdom of crowds? A decentralised election forecasting model that uses citizens’ local expectations“. *Electoral Studies* 30 (4): 771–83. <http://www.sciencedirect.com/science/article/pii/S0261379411000977>.

Norpoth, Helmut. 2016. „Primary Model Predicts Trump Victory“. *PS: Political Science & Politics* 49 (4): 655–58. <https://www.cambridge.org/core/journals/>

ps-political-science-and-politics/article/div-classtitleprimary-model-predicts-trump-victorydiv/
F374BCB3C2A291B21A8A39CD3ECD6BE3.

Norpoth, Helmut, und Thomas Gschwend. 2014. „A Near Miss for the Chancellor Model“. *EUSA: EU Political Economy Bulletin*, Nr. 17: 4–8.

Nurmi, Hannu. 2002. *Voting Procedures under Uncertainty Hannu Nurmi Springer*. Springer. //www.springer.com/de/book/9783540433521.

O’Brien, Sean P. 2010. „Crisis early warning and decision support: Contemporary approaches and thoughts on future research“. *International studies review* 12 (1): 87–104.

Peterson, Jack, Joseph Krug, Micah Zoltu, Austin K. Williams, und Stephanie Alexander. 2015. „Augur: A decentralized oracle and prediction market platform“. *arXiv:1501.01042 [cs]*, Januar. <http://arxiv.org/abs/1501.01042>.

Sanders, David. 1996. „Economic performance, management competence and the outcome of the next general election“. *Political Studies* 44 (2): 203–31.

Schrodt, Philip A. 2013. „Seven deadly sins of contemporary quantitative political analysis“. *Journal of Peace Research*, 287–300. <http://jpr.sagepub.com/content/early/2013/10/08/0022343313499597>.

Selb, Peter, und Simon Munzert. 2015. „Forecasting the 2013 German Bundestag Election Using Many Polls and Historical Election Results“. *German Politics* 25 (1): 73–83. <http://www.tandfonline.com/doi/full/10.1080/09644008.2015.1121454>.

Silver, Nate. 2016. „A User’s Guide To FiveThirtyEight’s 2016 General Election Forecast“. *FiveThirtyEight*. <https://fivethirtyeight.com/features/a-users-guide-to-fivethirteights-2016-general-election-forecast/>.

Sternberg, Sebastian. 2017. „How to Forecast Constitutional Court Decisions? Legal Context and Political Context in a Machine Learning Framework“. *Working Paper*.

Ward, Michael D., Brian D. Greenhill, und Kristin M. Bakke. 2010. „The perils of policy by p-value: Predicting civil conflicts“. *Journal of Peace Research* 47 (4): 363–75. <http://journals.sagepub.com/doi/10.1177/0022343309356491>.

Wolfers, Justin, und Eric Zitzewitz. 2004. „Prediction markets“. *Journal of economic perspectives* 18 (2): 107–26.