

Masterarbeit

Untersuchung von unentschlossenen Wählern –
Ansätze zur adäquaten Berücksichtigung bei
Vorwahlbefragungen und Verwendung für Prognosen
an unterschiedlichen Zeitpunkten

Ludwig-Maximilians-Universität München
Institut für Statistik



Verfasserin : Maren Kellerer
Betreuer : Dominik Kreiß,
Prof. Dr. Thomas Augustin

München, 28. Oktober 2022

Inhaltsverzeichnis

1	Einleitung	1
2	Epistemische Überlegungen zur Einbeziehung mengenwertiger Daten für Wahlprognosen	3
2.1	Berücksichtigung unentschlossener Wähler*innen	3
2.2	Einbeziehen von Stichprobengewichten	6
2.3	Random Forest als Punktschätzer mit Berücksichtigung der Gewichte	7
2.3.1	Homogenitätsannahme	8
2.3.2	Random Forest Algorithmus	9
2.3.3	Einbeziehen der Gewichte im Random Forest	11
2.4	Intervallwertige Schätzer mit Dempster Bounds unter Berücksichtigung der Gewichtung	13
2.4.1	Dempster Bounds ohne ergänzende Annahme	14
2.4.2	Weitere Einschränkungen unter der 80/20 und 50/50 Annahme	15
2.4.3	Mögliche Parteipartitionierungen	16
3	Anwendung bei der Bundestagswahl 2021 zu verschiedenen Zeitpunkten	18
3.1	Datenerhebung unentschlossener Wähler*innen von Civey	18
3.1.1	Analysen zur Stichprobengenerierung von Civey	19
3.1.2	Beschreibung unentschlossener Wähler*innen anhand von mengenwertigen Daten	26
3.2	Anwendung der Modellierungen	29
3.3	Interpretation der Ergebnisse	37
4	Fazit	42
A	Elektronischer Anhang	47

Abbildungsverzeichnis

1	Vergleich der fehlenden Werte in der Befragung einen Monat vor der Wahl – Gesamtdaten versus Stichprobe	20
2	Vergleich der Verteilung der Variable <i>Geschlecht</i> – Gesamtdaten versus Stichprobe	22
3	Vergleich der Verteilung der Variable <i>Alter</i> – Gesamtdaten versus Stichprobe	23
4	Verteilung der Variable <i>Gewicht</i> in der Befragung einen Monat vor der Wahl	24
5	Verteilung der Variable <i>Alter</i> unter Berücksichtigung des Gewichts in der Befragung einen Monat vor der Wahl	25
6	Verteilung der Parteien einen Monat vor der Wahl	27
7	Verteilung der Parteien zwei Wochen vor der Wahl	28
8	Vergleich der Random Forest Schätzungen beider Wellen	30
9	Verteilung einer simulierten Gewichtung	32
10	Vergleich der Dempster Bounds ohne Annahmen in beiden Wellen . .	34
11	Dempster Bounds unter 80/20 Annahme in beiden Wellen	35
12	Vergleich der Dempster Bounds mit Partitionierungen in beiden Wellen	36
13	Vergleich der geschätzten Anteile der Parteien in beiden Wellen . . .	39

Tabellenverzeichnis

1	Verteilung des Geschlechts – Bundestagswahl 2017	21
2	Verteilung des Alters – Bundestagswahl 2017	21
3	Vergleich der Accuracy Werte einen Monat vor der Wahl	33
4	Vergleich der Accuracy Werte zwei Wochen vor der Wahl	33

Zusammenfassung

Die große Zahl unentschlossener Wähler*innen bei Vorwahlbefragungen stellt eine Herausforderung für die Modellierung von Prognosen dar. Unentschlossene Wähler*innen, die zwischen mehreren Parteien unentschieden sind, werden oftmals vernachlässigt, wodurch wertvolle Informationen verloren gehen. Um das zu vermeiden, können die Angaben der unentschlossenen Befragten in Form von mengenwertigen Daten, welche die in Frage kommenden Parteien beinhalten, berücksichtigt werden. Diese Arbeit beschäftigt sich mit der Fragestellung, inwieweit es einen Unterschied macht, Unentschlossene in Prognosen zu berücksichtigen. Außerdem ist von Interesse, wie unterschiedliche Ansätze bei einer adäquaten Berücksichtigung Unentschlossener abschneiden. Dies umschließt sowohl die Anwendung unterschiedlicher Modellierungsansätze unter epistemischer Ungenauigkeit als auch eine Evaluation der Methodik und der Ergebnisse. Für das Ermöglichen einer adäquaten Berücksichtigung unentschlossener Wähler*innen, findet eine deskriptive Aufarbeitung der Stichprobe basierend auf stichprobentheoretischen Grundlagen statt. Durch unterschiedliche Ansätze und Ergebnisse werden punkt- oder intervallbasierte Prognosen erstellt, die an zwei unterschiedlichen Zeitpunkten betrachtet werden. Die Methodik wird anhand einer Umfrage zur Bundestagswahl 2021 angewendet, die das Meinungsforschungsinstitut *Civey* in Zusammenarbeit mit der LMU durchgeführt hat. Anhand der Ergebnisse sind Unterschiede vor allem bei den Parteien AfD und CDU/CSU bzw. SPD, abhängig vom Befragungszeitpunkt, je nach Ein- oder Ausschluss unentschlossener Wähler*innen zu erkennen. Dabei ändern sich die Prognosen insbesondere bei diesen Parteien, wenn unentschlossene Befragte berücksichtigt werden. Außerdem sind in jeder Partei Abweichungen der Prognosen je nach Befragungszeitpunkt zu erkennen.

1 Einleitung

Politische Vorwahlbefragungen sind ein wichtiges Instrument für die Abbildung der Meinung der Wahlbevölkerung und erlangen daher eine große Medienpräsenz. Vor allem die letzte Phase des Wahlkampfes einer Bundestagswahl hat eine große Bedeutung für den tatsächlichen Wahlausgang, da gezielt unentschlossene und taktische Wähler*innen angesprochen werden (Müller, 2021). Politische Vorwahlbefragungen zur Erstellung von Wahlprognosen setzen sich vor jeder Wahl mit dem wiederkehrenden Problem der noch unentschlossenen Wähler*innen auseinander, da es oftmals einen erheblichen Anteil an Befragten gibt, die noch keine klare Wahlpräferenz entwickelt haben. Der Grund dafür liegt in der Abwägungsphase, die bei schwierigen Entscheidungen erforderlich ist. Nach der Theorie von Tversky und Anderen gibt es anfangs eine Menge an Parteien, die bei der bevorstehenden Wahl in Frage kommen. Im Laufe der Zeit wird diese Menge immer weiter eingeschränkt, um letztendlich eine eindeutige Entscheidung treffen zu können (Tversky, 1972).

Unentschlossene Wähler*innen, welche zwischen verschiedensten Optionen abwägen, stellen eine Unsicherheitsquelle dar, die über den üblichen Umfragefehler hinausgeht. Eine systematische Verzerrung entsteht dann, wenn der zugrunde liegende Prozess, ein*e unentschlossene*r Wähler*in zu sein, nicht zufällig ist. Bei der Berücksichtigung unentschlossener Personen werden diese weder von der Befragung ausgeschlossen noch „gezwungen“, eine spezifische Partei anzugeben. Dadurch geht keine Information verloren und der Umfragefehler kann minimiert werden. Die These der Minimierung des Umfragefehlers wird von Oscarsson und Oskarson unterstützt, womit ein Ausschluss der Unentschlossenen zu einer verzerrten Analyse führen würde (Oscarsson und Oskarson, 2019). Bei üblichen Vorwahlumfragen wird dieser Gedankengang allerdings oftmals vernachlässigt und es werden lediglich eindeutige Entscheidungen in die Prognose miteinbezogen. Die zentrale Frage dieser Masterarbeit lautet damit: Verändern sich die Ergebnisse der Wahlprognosen durch die Berücksichtigung unentschlossener Wähler*innen und falls ja, inwiefern verändern sich diese?

Die meisten Vorwahlbefragungen werden von Meinungsforschungsinstituten durchgeführt, um die politische Stimmung des Landes zu erfassen. Dabei ist auffallend, dass die Umfragewerte zwischen den Instituten größere Abweichungen aufweisen, wenn die Bundestagswahl weiter in der Ferne liegt. Dies liegt am unterschiedlichen Umgang bis hin zum Ausschluss unentschiedener Wähler*innen. Je näher die Umfrage am Wahltag liegt, desto ähnlicher sind die Ergebnisse untereinander und zu dem tatsächlichen Wahlergebnis, was auf den geringeren Anteil der Unentschlossenen zurückzuführen ist (Westerhoff, 2022). Diese Arbeit untersucht die Ergebnisse von

Wahlprognosen zu zwei definierten Zeitpunkten, um Aussagen über die Veränderung im Verlauf der Zeit treffen zu können.

Es gibt verschiedene Möglichkeiten unentschlossene Wähler*innen zu operationalisieren. Nach Stefan Merz wird ein Befragter als unentschlossen klassifiziert, wenn dieser bei der Sonntagsfrage, also der Frage nach der Partei, welche bei einer Bundestagswahl gewählt würde, „weiß nicht“ angegeben hat (Merz, 2010). In dieser Arbeit werden die Wähler*innen allerdings direkt nach der Unentschlossenheit gefragt. Das digitale Markt- und Meinungsforschungsunternehmen *Civey* hat in Kooperation mit der LMU München eine zusätzliche Frage nach der Entschlossenheit des*der Befragten eingefügt. Das heißt, es wird zunächst nach der Entschlossenheit gefragt und anschließend eine Menge an möglichen Parteien aufgelistet, welche für die Wahl in Frage kommen. Damit können vor allem die Parteien ausgeschlossen werden, die für die befragte Person definitiv nicht in Frage kommen.

Die daraus entstandenen Daten unentschlossener Wähler*innen können auf zwei verschiedene Arten interpretiert werden, die jeweils mehrere unterschiedliche Modellierungsansätze nach sich ziehen. Zum einen gibt es die sogenannte epistemische Herangehensweise und zum anderen die ontische. Bei einem epistemischen Ansatz (oder auch disjunkten Ansatz) wird die Angabe als Menge an Elementen betrachtet, welche einen unbekanntem wahren Wert enthält. Beim ontischen Ansatz (oder auch konjunktiven Ansatz) hingegen werden Mengen als Einheiten verstanden, welche nicht reduziert werden können und somit als eine präzise Repräsentation der Entscheidung angesehen werden (Couso und Dubois, 2014). Diese Masterarbeit fokussiert sich lediglich auf den epistemischen Ansatz.

In dieser Arbeit werden in Kapitel 2 zunächst unentschlossene Wähler*innen definiert und die Wahrscheinlichkeitsverteilung einer Wahlentscheidung am Wahltag hergeleitet. Nach einer Thematisierung der Stichprobengewichte wird die Theorie und Entstehung der Modelle unter epistemischer Ungenauigkeit diskutiert. Kapitel 3 wendet die zuvor besprochenen Methoden auf die Bundestagswahl 2021 unter Verwendung der Daten von *Civey* an zwei unterschiedlichen Zeitpunkten an. Anschließend werden die Ergebnisse analysiert, die Ansätze verglichen und ein zeitlicher Vergleich herangezogen. Abschließend werden in Kapitel 4 die Herausforderungen erörtert und Anregungen für weitere Forschung gegeben.

2 Epistemische Überlegungen zur Einbeziehung mengenwertiger Daten für Wahlprognosen

Das folgende Kapitel ist in vier Teile gegliedert. Zunächst werden unentschlossene Wähler*innen definiert und die methodische Basis dargelegt. Daraufhin wird die Berücksichtigung der Stichprobengewichte motiviert. Für den Ansatz der epistemischen Ungenauigkeit als Umgang mit mengenwertigen Daten werden hier zwei ausgewählte Möglichkeiten betrachtet. Zum einen wird das zusätzliche partielle Wissen über die jeweiligen Parteipräferenzen genutzt, um punktwertige Schätzungen zu erhalten. Dabei wird die Annahme der strukturellen Ähnlichkeit getroffen und alle vorliegenden Informationen inklusive der Kovariablen einbezogen. Für die Schätzung wird hierbei eine Machine Learning Methodik verwendet. Zum anderen wird die Unsicherheit der mengenwertigen Angaben anhand von Intervallen veranschaulicht, wobei keine oder schwächere Annahmen im Vergleich zum ersten Ansatz getroffen werden. Der intervallwertige Ansatz der Dempster Bounds wird außerdem für die Idee von Koalitionen angewendet.

2.1 Berücksichtigung unentschlossener Wähler*innen

Im Vordergrund dieser Masterarbeit stehen unentschlossene Wähler*innen. Laut Kosmidis und Xezonakis (2010) gibt es zwei unterschiedliche Gründe für die Unentschlossenheit von Wähler*innen. Einerseits kann diese an der Unaufmerksamkeit und der daraus folgenden Unwissenheit gegenüber der Politik liegen, andererseits kann es auch politikinteressierte, unentschlossene Wähler*innen geben. Der zweite Grund ist daher die Folge von entweder gleich schwachen oder starken Argumenten zweier oder mehrerer Parteien, wodurch eine Unentschlossenheit entsteht (Kosmidis und Xezonakis, 2010). Allgemein nimmt die Aufmerksamkeit für die Politik vor der Wahl tendenziell zu und eine Vielzahl von Einflüssen können diesen Personen helfen, eine Entscheidung zu treffen. Unter anderem können Wahlkampagnen oder Instrumente wie der Wahl-O-Mat die Wahlentscheidung auf verschiedenste Arten beeinflussen. So werden beispielsweise während des Wahlkampfes die grundlegenden Parteipositionen aufgeklärt (Wlezien, 2010). Einer Umfrage von Statista zufolge entscheiden sich 26 % der Befragten im Laufe des Wahlkampfes, 19 % in der Woche vor der Wahl und 12 % am Tag der Wahl, welcher Partei sie ihre Stimme geben. Lediglich 34 % haben sich schon zwei Monate vor der Wahl entschieden und 8 % haben nicht vor, zur Wahl zu gehen (Statista, 2009). Einigen Vorwahlumfragen zufolge ist der Prozentsatz der unentschlossenen Wähler*innen in vielen Fällen sogar größer als der Vorsprung einer

Partei zur Anderen. Arcuri zufolge beeinträchtigen nicht beachtete, unentschlossene Befragte die aus den Umfragen abgeleiteten Vorhersagen und tragen zu einer Fehlerkomponente bei (Arcuri u. a., 2008).

Um die Frage zu beantworten, welchen Unterschied es in den zugrundeliegenden Daten macht, Unentschlossene in Vorwahlumfragen einzubeziehen, werden in dieser Arbeit Prognosen mit und ohne diese dargelegt. Dadurch verändert sich die jeweils zugrunde liegende Stichprobe. Im Folgenden werden zunächst Notationen eingeführt und notwendige Definitionen aufgezeigt.

Eine Wahl in einem Mehrparteiensystem kann als eine Wahl der Individuen $i \in \{1, \dots, n\}$ zwischen den Optionen $\{1, \dots, j\} = S$ verstanden werden. Entschiedene Personen wählen dabei ein Element von S und haben somit eine eindeutige Wahl getroffen. Die Position unentschiedener Wähler*innen hingegen, lässt sich durch eine nicht leere Teilmenge $\ell \subset S$, darstellen, die alle Parteien enthält, zwischen denen diese Person abwägt. Die Teilmenge ℓ wird im Folgenden auch Abwägungsmenge genannt und kann als disjunkte Zufallsmenge betrachtet werden, welche einen unbekanntem wahren Wert enthält. Dies beschreibt den Gedankengang der epistemischen Ungenauigkeit, woraufhin Modelle mit mengenwertigen Informationen und festgelegten Annahmen entwickelt werden können (Kreiss und Augustin, 2020).

Der Zustandsraum der Abwägungsmengen besteht aus allen möglichen Kombinationen der ursprünglichen Menge an Parteien und kann als Potenzmenge $P(S)$ der Menge S dargestellt werden. Die Potenzmenge ist eine Menge, welche aus allen möglichen Teilmengen von S besteht. Bei einer Anzahl von beispielsweise sechs Parteien würden $2^6 = 64$ mögliche Teilmengen von S entstehen. Dabei existiert bei unentschiedenen Personen eine Menge ℓ , welche als Realisierung der messbaren Abbildung $\mathcal{Y} : \Omega \rightarrow P(S)$ oder $\Omega \rightarrow 2^S$ aus einem zugrunde liegenden Raum Ω in die Menge aller Kombinationen gesehen werden kann (Kreiss und Augustin, 2021). Dabei gibt es ein wahres, unbekanntes Element $l \in \ell$, das die Wahl der Unentschlossenen am Wahltag darstellt.

Aufgrund fehlender Information über den wahren Wert l müssen weitere Annahmen oder Einschränkungen für die Prognosen unentschlossener Wähler*innen getroffen werden. Letztendlich muss dabei zwischen groben Ergebnissen ohne oder mit schwachen Annahmen und der Rechtfertigung von stärkeren Annahmen bei punktwertigen Ergebnissen entschieden werden (Kreiss und Augustin, 2020).

Die Individuen der Stichprobe werden zum einen durch die sogenannten Abwägungsmengen $\ell \in P(S)$ und zum anderen durch die Kovariablen $X = x$ im Raum \mathcal{X} bestimmt, welche die persönlichen Merkmale der Befragten charakterisieren. Die Abwägungsmenge ist dabei die Menge an Parteien, welche für das Individuum bei der nächsten Bundestagswahl mit der möglicherweise noch unbekanntem Wahlentscheidung am Wahltag $\{Y = l\}$ mit $l \in S$ in Frage kommen. Gegeben der Abwägungsmenge des Teilnehmers $i \in \{1, \dots, n\}$ soll die erwartete Häufigkeit von jedem Element in S mit der latenten Wahrscheinlichkeitsverteilung $P(Y = l)$ für alle $l \in S$ erhalten werden. Diese Verteilung ist eine Multinomialverteilung aller möglichen Parteikombinationen (Zustandsraum) mit $|S| - 1$ Parametern. Es wird angenommen, dass die Beobachtungen Y_i unabhängig und identisch verteilt sind (i.i.d) (Kreiss und Augustin, 2020).

Dabei wird über $(\ell, x) \in (2^S \times \mathcal{X})$ summiert, mit ℓ als ein Element aller möglichen Parteikombinationen und x als ein Element des Raumes \mathcal{X} , welches die charakteristischen Merkmale der Personen beschreibt.

Die gemeinsame Wahrscheinlichkeit lässt sich in bedingte Wahrscheinlichkeiten umformulieren. Daraus ergibt sich folgende Gleichung (Kreiss und Augustin, 2020):

$$\begin{aligned}
 P(Y = l) &= \sum_{(\ell, x) \in (2^S \times \mathcal{X})} P(Y = l, \mathcal{Y} = \ell, X = x) = \\
 &\sum_{(\ell, x) \in (2^S \times \mathcal{X})} \underbrace{P(Y = l | \mathcal{Y} = \ell, X = x)}_{\text{Übergangswahrscheinlichkeiten}} \cdot \underbrace{P(\mathcal{Y} = \ell | X = x)}_{\text{Abwägungsmengen}} \cdot \underbrace{P(X = x)}_{\text{Kovariablen}} \quad (1)
 \end{aligned}$$

Die Wahrscheinlichkeitsverteilung $P(Y = l)$ kann in drei Teile faktorisiert werden:

1. Die Übergangswahrscheinlichkeit, welche die Wahrscheinlichkeit, für eine bestimmte Partei zu stimmen, beschreibt, unter Berücksichtigung der Abwägungsmenge und der Kovariablen.
2. Die Wahrscheinlichkeit der Abwägungsmenge, gegeben den Kovariablen.
3. Die Wahrscheinlichkeit für die Kovariablen selbst.

Die letzten zwei Faktoren der Gleichung, die Wahrscheinlichkeit der Abwägungsmengen und der Kovariablen, können direkt aus den Daten der Vorwahlumfrage geschätzt werden. Der erste Faktor hingegen benötigt weitere Annahmen, da die Wahl l aus den Optionen ℓ bei den Unentschiedenen nicht beobachtbar ist. Für die Entschlossenen beträgt die Übergangswahrscheinlichkeit aufgrund der eindeutigen

Wahl eins, wohingegen bei den Unentschlossenen eine Punkt- oder Intervallschätzung notwendig ist (Kreiss und Augustin, 2021). Für die Prognose wird angenommen, dass sich die Situation in der Zeit vor der Wahl nicht ändert und auf die Zukunft verallgemeinert werden kann. Der Fokus liegt somit auf Prognosen, die als Extrapolation einer Gegenwartsvorhersage in die Zukunft angesehen werden können (Bauer u. a., 2022).

2.2 Einbeziehen von Stichprobengewichten

Eine ideale Befragung zur Wahl würde eine Zufallsstichprobe aller Wahlberechtigten darstellen, bei der jeder Befragte wahrheitsgemäß antwortet. In der Praxis ist dies allerdings nicht der Fall und bei jeder Umfrage entsteht eine Summe von Fehlern, die bei der Erhebung, dem Stichprobenverfahren und der Analyse der erhobenen Daten entstehen können. Der Umfragefehler (Total Survey Error) ist definiert als die Differenz zwischen einem Populationsparameter wie dem Mittelwert und der Schätzung dieses Parameters auf der Grundlage der Stichprobenerhebung (Biemer, 2010). Der Umfragefehler besteht dabei laut Biemer (2010) aus Nicht-Stichprobenfehlern und Stichprobenfehlern. Ein Beispiel für ein Nicht-Stichprobenfehler ist das Ausfallen einer Antwort, dem non-response. Stichprobenfehler hingegen entstehen durch das Ziehen einer Stichprobe aus der Grundgesamtheit, welche in dieser Masterarbeit durch die Wahlberechtigten der Bundestagswahl 2021 in Deutschland beschrieben wird. Bei Internetumfragen wie der von *Civey* ist die Auswahlwahrscheinlichkeit unbekannt, weshalb nicht von einer Zufallsstichprobe gesprochen werden kann und ein Stichprobenfehler entsteht. Die Auswahlwahrscheinlichkeit beschreibt dabei die Wahrscheinlichkeit, mit der eine Beobachtung der Grundgesamtheit in die Stichprobe aufgenommen wird (Couper, 2000).

Die gebräuchlichste Metrik, den Umfragefehler zu quantifizieren, ist der mittlere quadratische Fehler MSE (Mean Squared Error). Schätzungen, die aus Umfragedaten berechnet werden, haben einen entsprechenden MSE, der die Auswirkungen aller Fehlerquellen auf die Schätzung zusammenfasst. Bei den vorliegenden Daten entsprechen die Schätzungen den Anteilen für die jeweiligen Parteien. Allerdings ist es in der Praxis selten möglich, den MSE direkt zu berechnen, da dies in der Regel eine im Wesentlichen fehlerfreie Schätzung des Parameters erfordert. Dennoch ist das Konzept nützlich, um zu verstehen, wie die Auswirkungen von Erhebungsfehlern die Genauigkeit der Schätzung verringern (Biemer, 2010). Der mittlere quadratische Fehler ergibt sich für eine Schätzung $\hat{\theta}$ als Summe aus dessen quadratischem Bias und Varianz:

$$\text{MSE}(\hat{\theta}) = \text{Bias}^2(\hat{\theta}) + \text{Var}(\hat{\theta})$$

Dabei kann jede Fehlerquelle einen zufälligen Fehler, einen systematischen Fehler oder beides verursachen. Zufällige Fehler spiegeln sich in der Varianz der Schätzung wider, während systematische Fehler die Bias-Quadrat-Komponente beschreiben und problematischer sind (Biemer, 2010).

Es gibt umfangreiche Literatur zur nachträglichen Korrektur von Verzerrungen aufgrund von Antwortausfällen oder mangelnder Erfassung. Beispiele dafür sind Poststratifizierungen und Gewichtungsmethoden. Dabei ist die Hauptannahme, dass Stichprobenverzerrungen in den interessierenden Variablen auf Verzerrungen in den zur Gewichtung verwendeten Merkmalen zurückgeführt werden können. Wenn wichtige Merkmale in der Erhebung nicht erfasst wurden, können Verzerrungen nur unzureichend korrigiert werden (Bauer u. a., 2022). Bei der vorliegenden Umfrage soll eine quotierte Stichprobengenerierung der durch die Onlineumfrage erhobenen Daten von *Civey* mit anschließender Poststratifizierung den aufgrund des Auswahlverfahrens entstandenen Bias verringern (Richter, Wolfram und Weber, 2022). Die aus der Poststratifizierung entstandene Gewichtung wird dabei in den folgenden Modellierungen berücksichtigt. Obwohl die Gewichtung ein möglicher Weg ist, diese Probleme zu bewältigen, können nachfolgende Probleme zu weiteren Fehlern führen, wenn die Korrektur unzureichend ist. Die Bewertung des Ausmaßes der Probleme ist allerdings schwierig, da die finale Wahlentscheidung zum Zeitpunkt der Prognose unbekannt ist (Bauer u. a., 2022).

Im Folgenden werden zwei verschiedene Ansätze diskutiert, welche die in Gleichung 1 vorgestellten Wahrscheinlichkeiten schätzen. Dabei gibt es zahlreiche weitere impräzise Modelle für den Umgang mit mengenwertigen Daten, die in dieser Arbeit nicht thematisiert werden. Die beiden betrachteten Ansätze gehen unterschiedlich mit der Wahl zwischen Stärke der Annahme und Breite der Schätzergebnisse um. Außerdem wird die Stichprobengewichtung in beiden Ansätzen unterschiedlich berücksichtigt.

2.3 Random Forest als Punktschätzer mit Berücksichtigung der Gewichte

Um punktwertige Schätzungen zu ermöglichen, müssen zusätzliche Annahmen getroffen werden. Im Folgenden wird zunächst die Homogenitätsannahme begründet. Anschließend wird der zugrunde liegende Algorithmus für die Schätzung der Prognosen vorgestellt und die Stichprobengewichtung in den Prozess einbezogen.

2.3.1 Homogenitätsannahme

Eine mögliche Annahme für eine punktbewertete Schätzung ist, dass sich unentschiedene Personen unter Berücksichtigung der Kovariablen nicht strukturell von entschiedenen Personen unterscheiden. Daher wird von Kreiss und Augustin (2020) die Homogenitätsannahme vorgeschlagen. Da keine Informationen über die Wahl der Unentschlossenen im Datensatz vorliegen, wird in diesem Ansatz auf die Daten der entschiedenen Personen zusammen mit den vorhandenen Kovariablen zurückgegriffen. Es wird angenommen, dass sich Unentschiedene durchschnittlich wie Entschiedene in Bezug auf die Wahlentscheidung verhalten, abhängig von den Kovariablen. Die Annahme der strukturellen Ähnlichkeit ist zum einen intuitiv und zum anderen umstritten, da eine vollständige Homogenität in der Praxis vermutlich nicht gegeben ist. Dennoch ist dies eine Möglichkeit, alle vorliegenden Informationen einzubeziehen, und bietet einen ersten Ansatz für eine Punktschätzung.

Bei diesem Vorgehen gilt es, die Einschränkung in Form der Abwägungsmengen einzelner unentschlossener Personen zu berücksichtigen. Das heißt, bei der Schätzung werden die angegebenen Parteipräferenzen unentschlossener Wähler*innen einbezogen.

Mit Hilfe der Entschiedenen kann die Wahrscheinlichkeitsverteilung $P(Y_i = l | X_i = x_i, I_d = 1)$ aus den Daten geschätzt werden. I_d beschreibt hierbei die Indikatorfunktion der Entschlossenheit, wobei $I_d = 0$ entschiedene Personen definiert und $I_d = 1$ Unentschiedene. Die in Gleichung 1 beschriebene Abwägungsmenge ist dabei eine Einschränkung der möglichen Ergebnisse. Vorhergesagt wird also die Neigung zu einer bestimmten Partei aus dieser Abwägungsmenge unter Berücksichtigung der Entchiedenen und der Kovariablen. Um alle nicht in ℓ enthaltene Optionen auszuschließen, müssen die Prognosen der Parteiwahl der Unentschlossenen so skaliert werden, dass sie mit der Multinomialverteilung übereinstimmen. Um das zu erreichen, werden für alle $l \in \ell$ die vorhergesagten Tendenzen durch die Summe aller in der Abwägungsmenge enthaltenen Parteioptionen a geteilt (Kreiss und Augustin, 2020):

$$\underbrace{\hat{P}(Y = l | \mathcal{Y} = \ell, X = x)}_{\substack{\text{Übergangs-} \\ \text{wahrscheinlichkeiten}}} = \frac{\hat{P}(Y = l | X = x, I_d = 1)}{\sum_{a \in \ell} \hat{P}(Y = a | X = x, I_d = 1)} \quad (2)$$

Das heißt, bei unentschlossenen Personen wird die geschätzte Wahrscheinlichkeit einer bestimmten Wahlentscheidung gegeben den Kovariablen durch die Summe der geschätzten Wahrscheinlichkeiten der in Frage kommenden Parteien aus der Abwägungsmenge geteilt. Dadurch bleibt die Summe der Wahrscheinlichkeiten aller

Parteien aus der Abwägungsmenge für ein Individuum eins.

Die bedingten Verteilungen für jedes Individuum aus Gleichung 2 können durch unterschiedlichste Methoden geschätzt werden, wobei sich in dieser Arbeit auf die Random Forest Methodik im Bereich des Machine Learnings bezogen wird. Ein weiterer möglicher Ansatz ist das lineare Logit-Modell, welches weitere Arbeiten umsetzen, siehe z.B. Kreiss und Augustin (2020).

2.3.2 Random Forest Algorithmus

Das Ziel ist es, mittels einer Machine Learning Methodik einen Algorithmus auszugeben, der die Wahrscheinlichkeit für die Wahl einer bestimmten Partei schätzt. Dabei wird der Algorithmus zunächst auf einen Datensatz trainiert und anschließend auf neue Daten angewendet (Bhavsar und Ganatra, 2012). In dieser Arbeit werden die Daten der entschlossenen Personen zum Trainieren des Algorithmus verwendet, um Wahrscheinlichkeiten der Entscheidung Unentschlossener für die jeweiligen Parteien zu bestimmen. Um zunächst Aussagen über die Performance des Modells treffen zu können, wird das Modell anhand eines definierten Trainingsdatensatzes trainiert und anschließend an einem ausgewählten Testdatensatz getestet. Dafür werden hier die Daten der entschlossenen Befragten herangezogen und in einen Trainings- (80% der Daten) und einen Testdatensatz (20% der Daten) aufgeteilt. Die Güte des trainierten Modells wird mittels der Accuracy beschrieben, welche den Anteil der richtigen Prognosen im Testdatensatz beschreibt. Für eine genauere Evaluierung der Performance wird eine Kreuzvalidierung durchgeführt, wobei durch die mehrmalige Aufteilung in Teilmengen (Folds) der Test- und Trainingsdatensätze fundiertere Aussagen getroffen werden können.

Bei einer k -Fold Kreuzvalidierung wird der gesamte Datensatz zufällig in k nahezu gleich große Teilmengen partitioniert. Jede dieser Mengen stellt genau einmal den Testdatensatz dar, während die restlichen $k - 1$ Folds als Trainingsdatensatz fungieren. Das Modell wird basierend auf dem Trainingsdatensatz gefittet und anhand des Testdatensatzes getestet. Insgesamt werden k Accuracy Werte bestimmt, wovon ein gemittelter Wert berechnet wird und die Modellgüte schätzt (Refaeilzadeh, Tang und Liu, 2009).

Die Aufteilung der einzelnen Teilmengen erfolgt üblicherweise zufällig. Im vorliegenden Fall liegen jedoch unausgewogene Daten bei der Zielvariablen der Parteipräferenz vor. Im Falle von Parteidaten ist es üblich, dass bestimmte Parteien bei einer Wahl deutlich mehr Stimmen erhalten als andere, wodurch die Daten unausgewogen

sind. Dadurch entsteht das Problem, dass das Trainingsmodell mehr Informationen über die Mehrheitsklassen hat und nicht genug über die Minderheitsklassen lernen kann. Um dies zu vermeiden, wird eine stratifizierte Kreuzvalidierung durchgeführt. Durch die Stratifizierung wird die Verteilung der Zielvariablen aus den Daten in den einzelnen Folds beibehalten und eine genauere Beurteilung der Performance ermöglicht (Zeng und Martinez, 2000). Nachdem Schlussfolgerungen über die Generalisierung des Modells gezogen wurden, wird das trainierte Modell für die Vorhersage der Wahlanteile bei unentschlossenen Wähler*innen angewendet. Die Methodik des Random Forests, die sowohl an jedem der Folds für die Performancebeurteilung als auch für die Prognose der unentschlossenen Personen durchgeführt wird, ist im folgenden beschrieben.

Die Machine Learning Methodik des Random Forests hat sich als erfolgreiche Klassifizierungsmethodik durchgesetzt. Dabei werden mehrere einzelne Entscheidungsbäume kombiniert und die Prognose anhand einer Mittelwertbildung zusammengefasst (Breiman, 2001). Jeder einzelne Baum wächst dabei in drei Phasen:

1. Es wird eine Bootstrap-Stichprobe aus dem Trainingsdatensatz der Größe B gezogen.
2. Der Baum wächst, indem folgende Schritte rekursiv für jeden Knoten wiederholt werden, bis alle Beobachtungen zugeordnet sind oder ein bestimmtes Stoppkriterium erreicht wird:
 - (a) Es wird zufällig eine Anzahl $mtry$ aus den p vorhandenen Kovariablen gezogen.
 - (b) Es werden die optimale Variable und Variablenausprägung unter $mtry$ bestimmt (default: Giny importance).
 - (c) Der Knoten wird in zwei Tochterknoten anhand des gewählten Splitkriteriums geteilt (default: Giny impurity).

Es entstehen insgesamt R Bäume, die anhand eines Majority-Votings den Random Forest bilden. Das heißt, um eine ungesehene Beobachtung i zu klassifizieren, gibt jeder Baum des Waldes eine Entscheidung in Form einer Partei aus. Es wird anschließend die Klasse, im vorliegenden Datensatz die Partei, als Prognose bestimmt, welcher die Beobachtung i am häufigsten zugeordnet wurde (Müntefering, 2016).

Eine genauere Beschreibung der Klassifizierungsmethodik ist in Breiman u. a. (2017) zu finden. Bei den einzelnen Entscheidungsbäumen wird die Auswahl der jeweiligen

Kovariablen an jedem Knoten in Schritt 2.a zufällig getroffen, um den Split zu bestimmen. Dadurch wird die Korrelation zwischen den einzelnen Bäumen verringert. Üblicherweise wird bei einem Klassifikationsproblem $mtry = \lfloor \sqrt{p} \rfloor$ mit p Kovariablen gewählt. Laut Probst, Wright und Boulesteix (2019) stellt der Defaultwert allerdings nicht in jeder Datensituation die beste Wahl dar. Niedrigere Werte von $mtry$ führen zwar einerseits zu weniger korrelierten Bäumen, andererseits schneiden diese im Durchschnitt schlechter ab, da möglicherweise unwichtige Variablen ausgewählt werden. Daher sollte ein optimaler Kompromiss zwischen geringer Korrelation und Stärke des Baumes gefunden werden (Probst, Wright und Boulesteix, 2019). In dieser Arbeit wird $mtry$ als Tuningparameter behandelt, das heißt, es werden Modelle mit allen möglichen $mtry$ Parametern erstellt und die daraus entstandenen Accuracy Werte miteinander verglichen. Dabei wird die beschriebene Kreuzvalidierung mit $mtry \in \{1, \dots, p\}$ insgesamt p mal durchgeführt, was der Anzahl der Kovariablen entspricht.

Nachdem die Performancebeurteilung und das Parametertuning stattgefunden haben, wird das Modell für die Prognosen der unentschlossenen Wähler*innen herangezogen. Dabei fließen die Parteiwahl und die Kovariablen aller entschlossenen Personen in das Modell ein, um die Wahl der Unentschlossenen unter Berücksichtigung ihrer Abwägungsmenge vorherzusagen. Für die Prognose unentschlossener Wähler*innen wird dabei der aus der Kreuzvalidierung resultierende, optimale $mtry$ Wert verwendet.

Zusammenfassend bedeutet dies für die Prognose, dass die Daten der Entschlossenen direkt als Trainingsdaten in den Lernprozess einfließen, da sowohl die Parteiwahl Y als auch die Kovariablen X bekannt sind. Dadurch entsteht ein trainiertes Modell. Daraufhin wird anhand des trainierten Modells eine Prognose unter Berücksichtigung der Kovariablen und Abwägungsmenge der unentschlossenen Wähler*innen erstellt. Vorausgesetzt wird dabei, dass die bedingten Verteilungen von Y in Abhängigkeit der Kovariablen für entschiedene und unentschiedene Wähler*innen gleich sind: $P(Y = l|X = x, I_d = 1) = P(Y = l|X = x, I_d = 0)$. Dafür wird die Annahme der strukturellen Ähnlichkeit getroffen (Kreiss, Nalenz und Augustin, 2020).

2.3.3 Einbeziehen der Gewichte im Random Forest

Die Datenbasis der einzelnen Entscheidungsbäume aus Schritt 1. bildet wiederholt gezogene Stichproben aus den Originaldaten. Das zugrunde liegende Verfahren wird Bagging (Bootstrap Aggregation) genannt und wurde von Breiman (1996) eingeführt. Es kann also sein, dass bestimmte Beobachtungen durch das Ziehen mit Zurücklegen

wiederholt oder gar nicht vorkommen (Breiman, 1996). Da in der Stichprobe eine Gewichtung vorhanden ist, wird hier allerdings nicht zufällig aus den Originaldaten gezogen, sondern unter Berücksichtigung dieser Gewichte. Das heißt, Beobachtungen mit hohen Gewichten haben eine höhere Wahrscheinlichkeit, für die Entstehung des Baumes gezogen zu werden, als Beobachtungen mit niedrigen Gewichten. Auf diese Weise ist die Wahrscheinlichkeit, dass eine Beobachtung ausgewählt wird, proportional zu ihrem Stichprobengewicht. Die Sampling Methodik heißt dabei Probability Proportional to Size Sampling (PPS Sampling) (Sasikala, Bharathidasan und Venkateswaran, 2015).

Die PPS Sampling Methodik ist ein Stichprobenverfahren, bei dem jedes Element der endlichen Grundgesamtheit $i \in \{1, \dots, n\}$ eine bestimmte, unabhängige Wahrscheinlichkeit π_i hat, bei einer Ziehung in die Stichprobe aufgenommen zu werden. Die Wahrscheinlichkeit π_i ist dabei proportional zu dem vergebenen Gewicht w_i mit $w_i > 0$ (Skinner, 2014).

Bei diesem Verfahren wird in das Subsampling eingegriffen, wodurch Vor- und Nachteile entstehen können. Der Vorteil ist, dass die angegebenen Gewichte der Stichprobe berücksichtigt werden und unterrepräsentierte Beobachtungen in der Stichprobe häufiger in dem Random Forest Prozess auftreten als bei einer zufälligen Ziehung. Han, Williamson und Fong (2021) erwähnen allerdings das Problem einer stärkeren Gewichtung der Mehrheitsklasse bei unausgewogenen Daten. Das kann zu einer noch unausgewogeneren Bootstrap-Stichprobe führen und die Prognoseperformance der Minderheitsklasse negativ beeinflussen (Han, Williamson und Fong, 2021). Bei Betrachtung der Gewichtung innerhalb der Parteien fällt hier allerdings nicht auf, dass die großen Parteien eine höhere Gewichtung vorliegen haben als die kleineren Parteien. Der Grund ist, dass zwar unausgewogene Daten in der Zielvariablen vorliegen (siehe Kapitel 3.1.1), aber nur wenige in den Kovariablen, aus welchen sich die Gewichtung ergibt. Die Frage ist dennoch, ob sich die Berücksichtigung der Gewichte im Vergleich zu einer zufälligen Bootstrap-Stichprobenziehung auf die Performance auswirkt.

Da die Gewichte sowohl in die stratifizierte Kreuzvalidierung als auch in das Modell für die Prognosen der Unentschlossenen einfließen, ist von Interesse, wie diese die Performance beeinflussen. Die Frage ist also, ob sich die Performance durch den Eingriff in das Subsampling ändert und wenn ja, ob dies speziell auf die vorliegenden Gewichte zurückzuführen ist. Um dabei einen Allgemeinheitsgrad für die vorliegenden Daten bestimmen zu können, wird eine Simulation erstellt. Hierbei wird untersucht, inwiefern sich Bootstrap-Ziehungen unter Berücksichtigung der Gewichte auf die

Performance des Random Forests auswirkt. Die Simulation beschreibt eine zufällige Ziehung und Zuweisung einer GewichtungsvARIABLEN zu den Beobachtungen. Dabei wird eine Zufallsgenerierung der Exponentialverteilung mit Rate 1 herangezogen. Der Mittelwert ergibt sich bei einer Exponentialverteilung durch $1/\text{Rate}$ und die Varianz durch $1/\text{Rate}^2$. Für eine allgemeinere Aussage werden 100 Modelle gefittet, bei welchen jeweils die Accuracy berechnet wird. Dabei wird jedes Mal ein zufälliger Vektor als GewichtungsvARIABLEN aus der Exponentialverteilung gezogen und den Beobachtungen aus dem Trainingsdatensatz zugewiesen. Außerdem wird der Datensatz wiederholt in einen Trainings- und Testdatensatz aufgeteilt, wobei auch hier eine Stratifizierung durchgeführt wird, um die Verteilung der Parteien beizubehalten. Von den insgesamt 100 entstehenden Accuracy Werten wird der Mittelwert und die Varianz als Streuungsmaß berechnet. Hierbei ist von Interesse, ob sich die Accuracy Werte über die 100 Modelle ändern oder konstant bleiben. Die Simulation dient dabei lediglich als Mittel zur Interpretation der Performance beim Eingreifen in das Subsampling eines Random Forests, bezogen auf die vorliegenden Daten und Gewichte. Die Ergebnisse der Performancebeurteilung werden in Kapitel 3 vorgestellt.

Die Random Forest Methodik ergibt den ersten Schritt der Prognose. Die sich daraus ergebenden probabilistischen Schätzungen für jedes Individuum werden dann unter Verwendung von Gleichung 3 normiert, um die jeweiligen Abwägungsmengen zu berücksichtigen. Dadurch ergibt sich die Wahrscheinlichkeit, dass eine unentschlossene Person eine bestimmte Partei aus ihrer Auswahl wählt. Die punktuellen Schätzungen der jeweiligen Parteien ergeben sich anschließend durch die Berechnung des Durchschnitts über die einzelnen individuellen Schätzungen. Die Schätzung der Gesamtverteilung aller Befragten erschließt sich aus den konkreten Angaben der Entschlossenen und den punktuellen Schätzungen der Unentschlossenen.

2.4 Intervallwertige Schätzer mit Dempster Bounds unter Berücksichtigung der Gewichtung

Ein weiterer Ansatz für die Prognose der Wahlentscheidung unentschlossener Wähler*innen sind Dempster Bounds (Dempster, 2008). Dabei können intervallwertige Ergebnisse dargestellt werden, welche auf keinen oder weniger starken Annahmen beruhen als der vorherige Ansatz. Genau genommen wird kein Intervall geschätzt, sondern eine obere und untere Grenze, woraus sich ein Intervall ergibt. Dabei werden lediglich die in den Daten verfügbaren Informationen verwendet. Im Folgenden werden drei unterschiedliche Anwendungen der Dempster Bounds vorgestellt.

2.4.1 Dempster Bounds ohne ergänzende Annahme

Zunächst werden die Dempster Bounds ohne weitere Einschränkungen betrachtet. Obwohl dieser Ansatz zu einer groben Vorhersage führt, spiegelt sich damit die gesamte Unsicherheit resultierend aus der Unentschiedenheit wider. Durch die fehlende Information der Unentschlossenen bezüglich der Parteiwahl reichen die Dempster Grenzen jeder einzelnen Partei vom schlechtesten Fall bis hin zum besten Fall. Der schlechteste Fall beschreibt damit die Entscheidung gegen die betrachtete Partei bei jeder Beobachtung, die diese Partei in der Abwägungsmenge enthält, und der beste Fall die Entscheidung für diese Partei. Die untere Grenze kann, ausgenommen vom Umfragefehler, als das garantierte Minimum an Stimmen und die obere Grenze als das mögliche Potenzial an Stimmen angesehen werden. Durch das Widerspiegeln der gesamten Unsicherheit der unentschiedenen Bevölkerung sind die Grenzen tendenziell breit (Kreiss und Augustin, 2020).

Da keine Information vorliegt, welcher Wert in ℓ die wahre Wahlentscheidung ist, wird für jede Menge ℓ , die mehr als ein Element enthält, die Unentschiedenheit durch eine Übergangswahrscheinlichkeit zwischen 0 und 1 widerspiegelt. Dabei bedeutet 0, dass keine unentschlossene Person die betrachtete Partei wählt, und 1, dass alle diese wählen, vorausgesetzt diese Partei befindet sich in den jeweiligen Abwägungsmengen. Kombiniert man diese Reichweite mit den Angaben der entschiedenen Personen, ergeben sich intervallwertige Prognosen. Die Dempster Bounds beschreiben einen Bereich für den Anteil der Individuen, die die Partei Y wählen, welche den wahren Anteil ohne Umfragefehler enthält. Der Bereich ergibt sich aus den unteren und oberen Grenzen des Möglichen (Kreiss und Augustin, 2021). Dabei werden die Gewichte der einzelnen Beobachtungen $w_i, i \in \{1, \dots, n\}$ mit $w_i > 0$ berücksichtigt, indem diese mit den jeweiligen Übergangswahrscheinlichkeiten von 0 und 1 multipliziert werden. Für alle Parteien als Wahlentscheidung am Wahltag Y können die Grenzen wie folgt definiert werden:

$$p_{\text{oben}}(Y = y) = \frac{1}{n} \sum_{i=1}^n (P(\{Y\} \subseteq \ell_i) \cdot w_i) \quad (3)$$

$$p_{\text{unten}}(Y = y) = \frac{1}{n} \sum_{i=1}^n (P(\{Y\} = \ell_i) \cdot w_i) \quad (4)$$

Bei diesem Ansatz wird die Übergangswahrscheinlichkeit aus Gleichung 1 unabhängig von den Kovariablen auf die Extremwerte gesetzt.

Die obere Grenze berücksichtigt die Fälle, bei denen die betrachtete Partei

Y eine Teilmenge der Abwägungsmenge ist. Die untere Grenze hingegen umfasst nur die Beobachtungen, bei denen die betrachtete Partei gleich der Partei aus der Abwägungsmenge ist. Das heißt, die untere Grenze beinhaltet lediglich entschlossene Personen, da die Abwägungsmenge nur in diesen Fällen aus einem Element besteht. Die obere Grenze umfasst hingegen alle unentschlossenen Personen, die die betreffende Partei für die nächste Bundestagswahl in Betracht ziehen. Für den jeweiligen Anteil der Ober- und Untergrenze werden die Wahrscheinlichkeiten aller Wähler*innen $i \in \{1, \dots, n\}$ aufsummiert und durch die Anzahl der Teilnehmer n geteilt.

In allen vorgestellten Versionen der Dempster Grenzen werden die vorhandenen Gewichte der Stichprobe ebenfalls berücksichtigt. Die Gewichte w_i der einzelnen Personen werden mit den möglichen Wahrscheinlichkeiten von 0 oder 1 multipliziert und bilden somit die untere oder obere Schranke. Da die Summe der Gewichte gleich der Anzahl der Beobachtungen entspricht, ändert sich der Umfang der Stichprobe dadurch nicht. Allerdings werden unterrepräsentierten Beobachtungen höher und überrepräsentierte Beobachtungen niedriger gewichtet. In Bezug auf die Wahlentscheidung der Partei FDP würde dies beispielhaft bedeuten, dass die Wahrscheinlichkeit $P(\{Y\} \subseteq \ell_i)$ bei der unteren Grenze gleich 0 ist, wenn die entschiedene Person i eine andere Partei als die FDP angegeben hat, und gleich 1, wenn die Abwägungsmenge aus der FDP besteht. Bei der oberen Grenze ist die Wahrscheinlichkeit für Person i gleich 1, wenn die FDP ein Element der Betrachtungsmenge ist, und 0, wenn sie es nicht ist. Daraufhin werden die entstandenen Wahrscheinlichkeiten in beiden Grenzen mit dem Gewicht der Person i multipliziert.

2.4.2 Weitere Einschränkungen unter der 80/20 und 50/50 Annahme

Die im vorherigen Abschnitt betrachteten Grenzen spiegeln die gesamte Unsicherheit der unentschlossenen Wähler*innen wider. Da es jedoch recht unwahrscheinlich ist, dass eines dieser Extremszenarios für eine der Parteien eintritt und sich alle Personen für oder gegen die betrachtete Partei aus der Abwägungsmenge entscheiden, können diese Grenzen weiter eingeschränkt werden. Eine Einschränkung ist den prozentualen Anteil, wie viele unentschlossene Personen sich minimal oder maximal für die Partei entscheiden, zu bestimmen. Ein realistischer Ansatz bietet hierbei die 80/20 Annahme, die besagt, dass insgesamt nicht weniger als 20% und nicht mehr als 80% der Unentschlossenen für die entsprechende Partei stimmen. Bei einem 80/20 Ansatz bedeutet das beispielsweise für die SPD, dass mindestens 20% und höchstens 80 % der Befragten, die zwischen der SPD und mindestens einer weiteren Partei abwägen, am Ende die SPD wählen (Kreiss und Augustin, 2021). Eine weitere

Möglichkeit ist, das Intervall zu einem Punkt mit der 50/50 Annahme zu verringern. Das heißt die Hälfte der unentschlossenen Wähler*innen entscheiden sich für die betrachtete Partei und die andere Hälfte dagegen. In diesem Fall entspricht die untere Grenze gleich der oberen Grenze. Der Ansatz, die Hälfte der unentschlossenen Befragten einer Grenze zuzuordnen, erscheint zwar intuitiv, spiegelt jedoch einen „naiven“ Ansatz wider, da bei dieser Methodik keine weitere Information einbezogen wird.

Die Definition der eingeschränkten Grenzen wird im Folgenden beschrieben. Dabei steht $s \in [0, 1]$ für den Anteil der Einschränkung der Unentschlossenen bei der oberen Grenze. In vorliegenden Fall beträgt dieser 0,8 oder 0,5.

$$p_{\text{oben}}(Y = y) = \frac{1}{n} \sum_{i=1}^n (P(\{Y\} = \ell_i | I_d = 0) \cdot w_i + P(\{Y\} \subseteq \ell_i | I_d = 1) \cdot w_i \cdot s) \quad (5)$$

$$p_{\text{unten}}(Y = y) = \frac{1}{n} \sum_{i=1}^n (P(\{Y\} = \ell_i | I_d = 0) \cdot w_i + P(\{Y\} \subseteq \ell_i | I_d = 1) \cdot w_i \cdot (1 - s)) \quad (6)$$

Es wird anhand der Indikatorfunktion bei jeder Beobachtung differenziert, ob es sich um eine entschlossene oder unentschlossene Person handelt. Entschlossene Personen zählen wie bei Gleichung 3 und 4 sowohl in die untere als auch in die obere Grenze. Bei den unentschlossenen Wähler*innen werden allerdings lediglich 80% der oberen Grenze und 20% der unteren Grenze zugewiesen. Dies gilt analog zu Dempster Bounds mit anderen Einschränkungen, wie z.B. der 50/50-Annahme. Die Gewichte werden, wie in dem vorherigen Abschnitt beschrieben, mit den Übergangswahrscheinlichkeiten multipliziert.

2.4.3 Mögliche Parteipartitionierungen

Ein präzentes Thema bei einer Bundestagswahl sind mögliche Parteikoalitionen. Nach Bär beachten einige Wähler*innen bei ihrer Wahlentscheidung, welche Koalitionen zustande kommen könnten (Bär, 2017). Dabei ist von Interesse, welche Parteien sich vereinen könnten, um 50% der Stimmen zu erreichen und eine neue Regierung zu bilden. Im folgenden Abschnitt werden die Prognosen von bestimmten Parteikombinationen betrachtet. Dabei wird die Koalition als eine Partitionierung des politischen Raumes gesehen, welche durch die Kombination von Parteien entsteht. Hierbei verändert bzw. erweitert sich der Zustandsraum. Dadurch entsteht der Vorteil, dass ein Teil der Unsicherheit durch die Mengenangaben der unentschiedenen Befragten

aufgelöst werden kann. Ist eine Person beispielsweise zwischen der FDP und der AfD indifferent, wird diese vermutlich die Stimme der Koalition {FDP, AfD} geben. Für diese Person wird die ursprünglich unpräzise, partielle Angabe präzise (Kreiss und Augustin, 2021).

Die Eigenschaft der präzisen Angaben aus einer anfangs unpräzisen Angabe entsteht lediglich bei den Personen, die die ausgewählten Partitionierungen als ihre Abwägungsmenge angegeben haben. Falls eine Person zwischen den Parteien einer ausgewählten Kombination unentschieden ist, wird diese wie entschiedene Personen bei den vorherigen Dempster Bounds sowohl zu der unteren als auch der oberen Grenze hinzugefügt. Hat eine unentschlossene Person mehr Parteien angegeben, als die betrachtete Partitionierung enthält, wird diese zwar zur oberen Grenze, aber nicht zur unteren Grenze, gezählt. Entschiedene Personen sind Teil aller unteren bzw. oberen Grenzen, wenn die angegebene Partei einem Element der Koalition entspricht. Auch hier werden die Gewichte der Beobachtungen analog zu den vorherigen Versionen der Dempster Bounds berücksichtigt.

Ein kritischer Aspekt der Annahme ist dabei, dass entschlossene Personen, die beispielsweise die FDP wählen würden, eine Koalition zwischen dieser und der AfD vielleicht nicht in Betracht ziehen würden, dennoch aber in beide Grenzen fallen. Der Fokus liegt hier allerdings auf den unentschlossenen Personen und auf der Breite der entstandenen Intervalle.

3 Anwendung bei der Bundestagswahl 2021 zu verschiedenen Zeitpunkten

Für das Nachvollziehen der Datenerhebung unentschlossener Wähler*innen zur Bundestagswahl 2021 wird zunächst diskutiert, wie diese von *Civey* berücksichtigt wurden. Danach wird die Stichprobengenerierung analysiert, welche von dem digitalen Markt- und Meinungsforschungsunternehmen durchgeführt wurde und eine deskriptive Analyse in Bezug auf die Gewichtung herangezogen. Anschließend werden die vorgestellten Möglichkeiten zur Modellierung der Wahlprognosen zur Bundestagswahl 2021 anhand der zur Verfügung gestellten Daten von *Civey* angewendet. Die Ergebnisse werden abschließend interpretiert und im Hinblick auf die verschiedenen Zeitpunkte und Modellierungsarten verglichen.

3.1 Datenerhebung unentschlossener Wähler*innen von Civey

Das Meinungsforschungsinstitut *Civey* liefert die Datengrundlage der folgenden Anwendungen in Form einer Online-Vorwahlbefragung. Diese ist in Kooperation mit der LMU entstanden und somit an die Absicht der Erhebung von unentschlossenen Wähler*innen angepasst. Dabei wurde eine sogenannte Trendstudie durchgeführt, das heißt, dieselbe Studie wurde zu mehreren Zeitpunkten mit jeweils unterschiedlichen Stichproben durchgeführt. Dadurch ist es möglich, Veränderungen im Laufe der Zeit zu verstehen. Die Befragungswellen weisen ähnliche Anteile an den erhobenen Variablen auf, sind daher vergleichbar und lassen Aussagen über die Veränderungen je nach Zeitpunkt zu. Da unterschiedliche Datensätze zugrunde liegen, ist es allerdings nicht möglich, Aussagen auf Individualebene zu treffen (Caruana u. a., 2015).

Aus der *Civey*-Umfrage entstehen Daten aus insgesamt drei verschiedenen Wellen, die jeweils eine eigenständige Stichprobe für einen bestimmten Zeitpunkt darstellen. Die erste Umfrage wurde zwei Monate vor der Bundestagswahl 2021 durchgeführt, die zweite einen Monat davor und die dritte zwei Wochen vor der Wahl. In allen Umfragen werden die Teilnehmer zunächst gefragt, ob sie sich ihrer zukünftigen Wahlentscheidung sicher sind. Diejenigen, die sich bereits sicher sind, werden nach der Einzelpartei gefragt. Befragte, die nicht sicher sind, können zwischen einer Auswahl an Parteien diejenigen angeben, die sie in Betracht ziehen. In beiden Fällen stehen folgende Optionen zur Auswahl: {SPD, CDU/CSU, Grüne, FDP, Linke, AfD, Eine andere Partei, Sonstige, Nichtwähler}. Für jeden unentschlossenen Teilnehmer wird eine Menge an Parteien ermittelt, zwischen welchen die Person noch unentschlossen ist. Das heißt, alle Befragten können die Wahlpräferenz so genau angeben, wie diese aktuell in der Lage sind, dies zu tun. Des Weiteren wurden zwölf soziodemographi-

sche Variablen der Befragten erhoben: *Geschlecht*, *Altersgruppe*, *Bevölkerungsdichte*, *Kaufkraft*, *berufliche Stellung*, *Berufsbildung*, *Beschäftigungsstatus*, *Bildungsabschluss*, *Familienstand*, *Kinder im Haushalt*, *Ost/West* und *Religion*. Alle erhobenen Kovariablen sind diskret und weisen zwei bis fünf unterschiedliche Ausprägungen auf.

Letztendlich sind für jede Beobachtung eine bestimmte Partei oder eine Menge an möglichen Parteien, die eine noch unbekannte Entscheidung der Wahl beinhaltet, und die aufgelisteten Kovariablen im Datensatz vorhanden.

3.1.1 Analysen zur Stichprobengenerierung von Civey

Die entstandenen eigenständigen Datenwellen enthalten jeweils etwa 25.000 Beobachtungen. Das Meinungsforschungsinstitut *Civey* hat pro Welle eine Stichprobe generiert, die jeweils 5.000 Beobachtungen enthält. Diese Generierung basiert auf einer quotierten Stichprobe, die der Verteilung von ausgewählten soziodemographischen Variablen, wie beispielsweise Alter oder Geschlecht, der Grundgesamtheit entspricht. In diesem Fall wird die Verteilung der Variablen aller Wahlbeteiligten der vorherigen Bundestagswahl 2017 herangezogen. Die verbleibenden Diskrepanzen nach der Quotenstichprobenziehung zwischen der Stichprobe und der Grundgesamtheit in Bezug auf die bekannten Variablen werden durch eine auf Raking basierende Gewichtung so gut wie möglich entfernt. Gruppen, die in der Stichprobe unterrepräsentiert sind erhalten somit ein höheres Gewicht, während Gruppen, die überrepräsentiert sind, ein geringeres Gewicht erhalten. Eine Annahme ist dabei die Unabhängigkeit von den Antworten der Teilnehmer und der Auswahl in die Stichprobe (Richter, Wolfram und Weber, 2022). Diese Herangehensweise beschreibt die in Kapitel 2.2 dargestellte Möglichkeit, einen Teil der entstandenen Verzerrungen zu korrigieren. Um die Stichprobengenerierung und anschließende Poststratifizierung von *Civey* nachvollziehen zu können, wird im folgenden Abschnitt das Vorgehen am Beispiel der Befragung einen Monat vor der Wahl analysiert.

Von den über 25.000 Beobachtungen des gesamten Datensatzes der zweiten Welle sind etwa 22.000 vollständig. Bei 13 % fehlt mindestens eine der zwölf relevanten Variablen, sowohl im Gesamtdatensatz als auch in der generierten Stichprobe. Abbildung 1 veranschaulicht den Anteil der fehlenden Angaben pro Variable für beide Datengrundlagen.

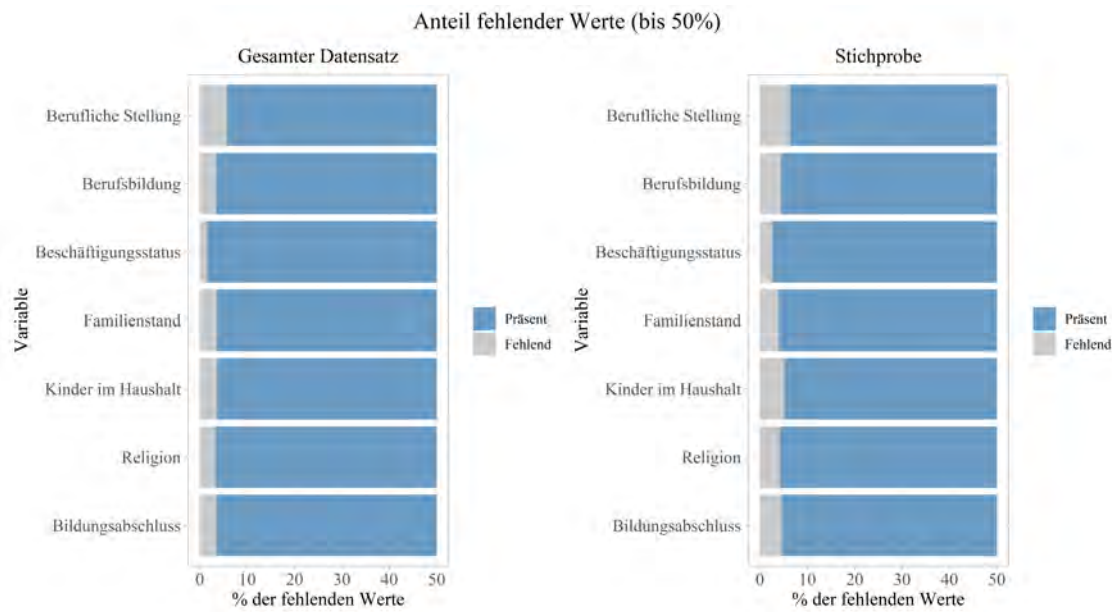


Abbildung 1: Vergleich der fehlenden Werte in der Befragung einen Monat vor der Wahl (bis zu 50%) – Gesamter Datensatz versus Stichprobe

Es ist in beiden Fällen zu erkennen, dass keine Variable mehr als 7% fehlende Werte aufweist, wobei die Variable *Berufliche Stellung* die meisten darunter kennzeichnet. Für die nicht aufgelisteten Variablen liegen keine fehlenden Werte vor. Beim Vergleich beider Plots ist auffallend, dass der Anteil fehlender Werte in der Stichprobe proportional etwas größer ist als im gesamten Datensatz.

Das Fehlen einiger Antworten ist ein Problem bei fast jeder Analyse, die Daten aus freiwilligen Erhebungen verwendet. Falls sich Personen, die die Frage beantwortet haben, systematisch von den Personen unterscheiden, die nicht geantwortet haben (Missing not at random), kommt es zu einer Verzerrung der Ergebnisse. Insbesondere sind dabei die Gründe für die Nichtbeantwortung der Fragen oftmals unbekannt, wodurch das Verringern der Verzerrungen schwer wird (Rubin, 2004). In dieser Arbeit werden lediglich die vollständig verfügbaren Beobachtungen betrachtet, wodurch die Nachteile des Informationsverlustes und der möglichen Verzerrung bei Nichtbeachtung der Nonresponse in Kauf genommen werden. Der Grund für diesen Umgang mit den fehlenden Werten liegt in der gleichen Handhabung von Referenzanalysen und in der Wahrung der Einfachheit der Analyse. Ebenfalls möglich wäre eine multiple Imputation, um die erzeugten Unsicherheiten zu berücksichtigen (Rubin, 2004).

Da die Verteilungen der Kovariablen *Alter* und *Geschlecht* der letzten Bundestagswahl öffentlich einsehbar sind, basieren die folgenden Vergleiche des gesamten Datensatzes und der gezogenen Stichprobe auf diesen Variablen. Allerdings sollte

beachtet werden, dass weitere zehn Variablen im Datensatz vorhanden sind. Des weiteren sind die genauen Quellen für die Berücksichtigung der Verteilungen bei der Quotenstichprobe nicht bekannt, weshalb hier exemplarisch die Bundestagswahl 2017 herangezogen wird. Im Folgenden sind Übersichtstabellen der Verteilung dieser Variablen zur Bundestagswahl 2017 veranschaulicht (Bundeswahlleiter, 2017):

Geschlecht	Anteil BTW 2017
Männlich	48,3 %
Weiblich	51,7 %

Tabelle 1: Verteilung des Geschlechts – Bundestagswahl 2017

Alter	Anteil BTW 2017
18 - 29	15,4 %
30 - 39	13,9 %
40 - 49	14,7 %
50 - 59	20,0 %
60 - 69	15,4 %
70 +	20,7 %

Tabelle 2: Verteilung des Alters – Bundestagswahl 2017

Anhand der Verteilung des Geschlechts wird deutlich, dass etwa gleich viele Frauen wie Männer an der Wahl teilgenommen haben (Tabelle 1). In Tabelle 2 ist zu erkennen, dass die drei jüngsten Altersgruppen zwischen 13 und 15 % darstellen und die älteste Altersgruppe über 20 % aufweist. Dabei sollte beachtet werden, dass sich die Einteilung der Altersstufen von der Einteilung, die im Datensatz vorhanden ist, unterscheidet.

Angesichts des Vorgehens der Quotenstichprobenziehung werden unter anderem die abgebildeten Anteile (Tabelle 1 und Tabelle 2) bei der Ziehung aus den gesamten Beobachtungen berücksichtigt. Das Ziel ist dabei, die Anteile der vorherigen Bundestagswahl bezüglich *Geschlecht* und *Alter* in der Stichprobe zu erhalten. Im Folgenden werden die Anteile dieser Variablen des gesamten Datensatzes und der daraus resultierenden Stichprobe verglichen. Abbildung 2 veranschaulicht die relative Häufigkeit der Variable *Geschlecht* in beiden Datensätzen unter Einbeziehung der Entschlossenheit.

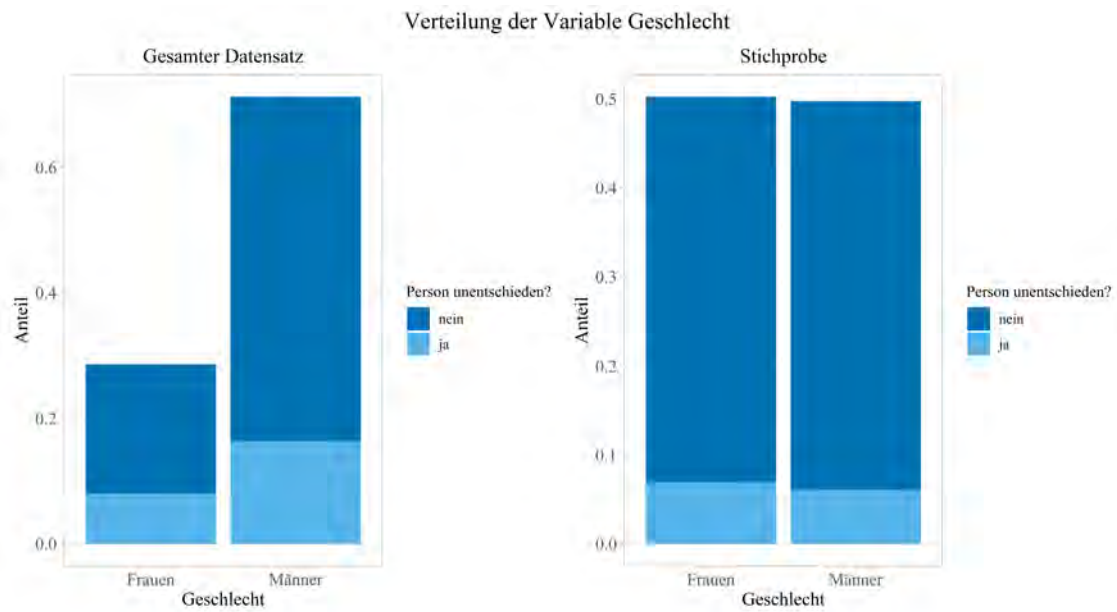


Abbildung 2: Vergleich der Verteilung der Variable *Geschlecht* in der zweiten Welle – Gesamtdaten versus Stichprobe

Der gesamte Datensatz weist knapp 30 % Frauen und über 70 % Männer auf. Die daraus entstandene Stichprobe kennzeichnet jedoch bei beiden Geschlechtern etwa 50 %. Somit wird die tatsächliche Verteilung bei der Bundestagswahl 2017 (Tabelle 1) in der Stichprobe widerspiegelt. Der Anteil unentschlossener Personen ist dabei bei beiden Geschlechtern annähernd ausgeglichen.

Abbildung 3 beschreibt die relative Häufigkeit der Variable *Alter* in jeweils beiden Datensätzen unter Berücksichtigung der Entschlossenheit.

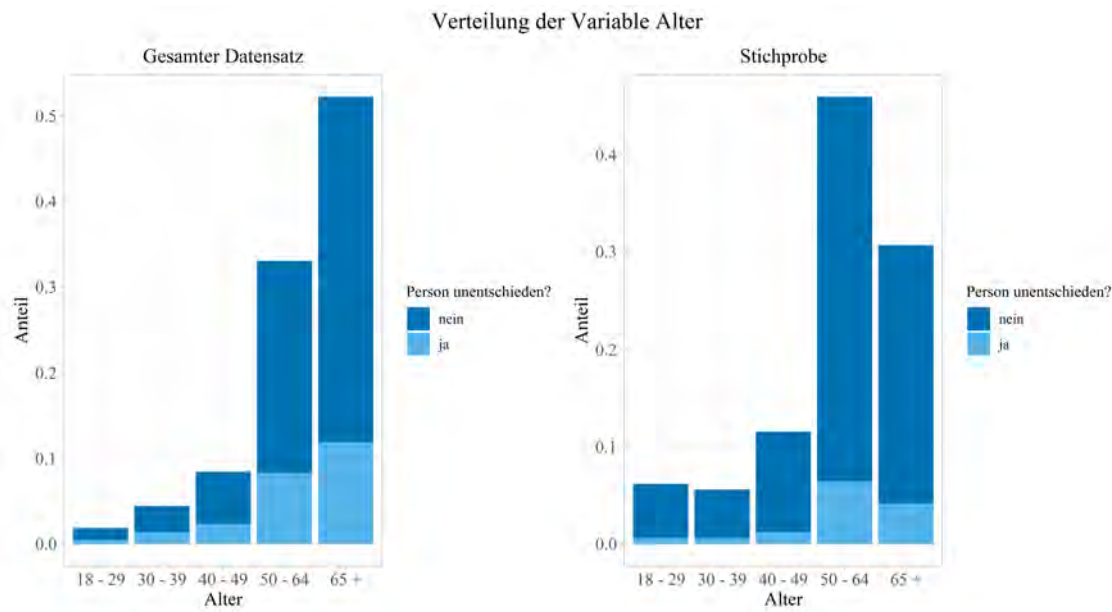


Abbildung 3: Vergleich der Verteilung der Variable *Alter* einen Monat vor der Wahl – Gesamtdaten versus Stichprobe

Bei Betrachtung der Verteilung des Alters im gesamten Datensatz fällt auf, dass eine rechtssteile Verteilung vorliegt. Das heißt, je höher die Altersgruppe, desto mehr Personen haben bei der Umfrage teilgenommen. Über 50 % der Teilnehmer*innen im gesamten Datensatz weisen ein Alter über 65 auf. Die tatsächliche Verteilung bei der vorherigen Bundestagswahl (Tabelle 2) beschreibt allerdings approximativ einen jeweils ähnlichen Anteil in den drei niedrigsten Altersgruppen und den beiden höchsten Altersgruppen. Nach der Stichprobenziehung ergeben sich im Vergleich zum gesamten Datensatz zwar höhere Anteile in den niedrigeren Altersgruppen, allerdings betragen diese nur 6-12 % und nicht rund 14 %. Im Gegensatz dazu liegen über 45 % statt etwa 25-30 % der Beobachtungen im Alter von 50 bis 64 vor. Bei der Stichprobe wird außerdem erkennbar, dass im Vergleich zum gesamten Datensatz in den drei niedrigsten Altersgruppen proportional weniger unentschlossene Personen vorhanden sind.

Aufgrund der bleibenden Diskrepanz zwischen dem Gesamtdatensatz und der Stichprobe in Bezug auf beispielsweise das Alter wird die bereits erwähnte Poststratifizierung in Form einer Gewichtungsvariablen herangezogen. Die Verteilung der Variable *Gewicht* der Stichprobe ist in Abbildung 4 veranschaulicht:

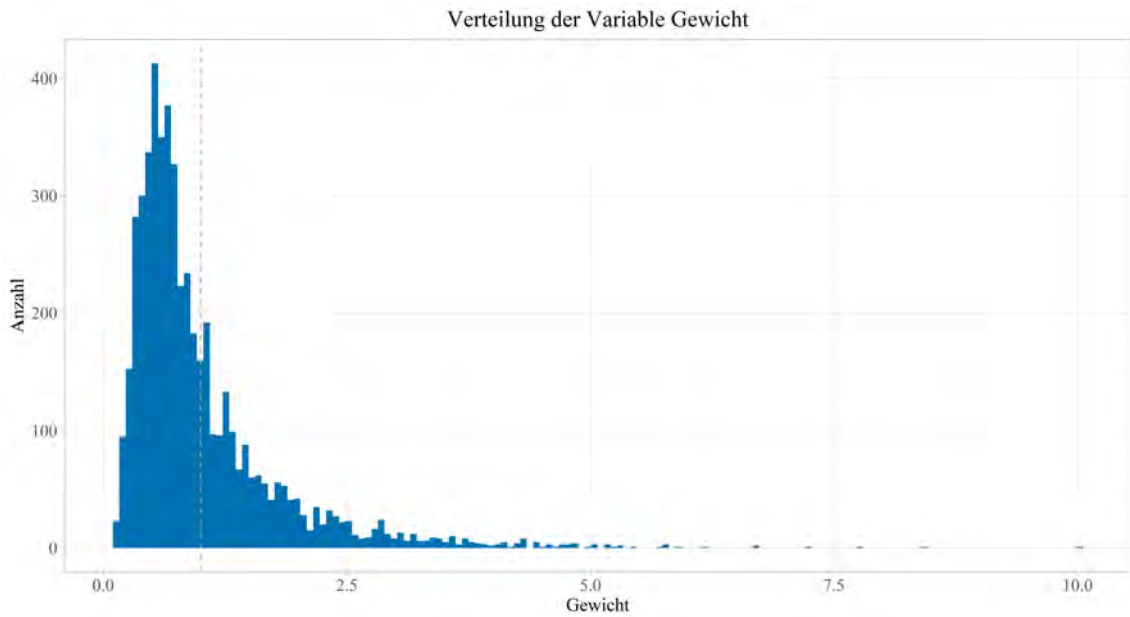


Abbildung 4: Verteilung der Variable *Gewicht* in der Stichprobe der Befragung einen Monat vor der Wahl

Die Variable *Gewicht* reicht vom Minimum 0,1 bis hin zum Maximum 10,0. Das arithmetische Mittel beträgt hierbei genau eins, was bedeutet, dass die Summe der Gewichte der Anzahl an Beobachtungen entspricht. Es wird dabei erkennbar, dass zwei Drittel der Beobachtungen eine Gewichtung kleiner eins zugewiesen bekommen und somit nach unten gewichtet werden. Etwa 10 % der Beobachtungen weisen ein Gewicht über zwei auf und nur wenige davon sind über vier (ca. 1 %). Ein Grund für diese Verteilung könnte unter anderem das hohe Durchschnittsalter der Befragten sein.

Für eine Veranschaulichung der Gewichtung in Bezug auf das Alter ist in Abbildung 5 auf der linken Seite die Verteilung des Alters der Stichprobe mit Einbeziehung des Gewichts abgebildet. Die Gewichtung ist dabei in drei Intervalle $(0; 1]$, $(1; 3]$ und $(3; 9]$ eingeteilt, die durch eine farbliche Abstufung kenntlich gemacht werden. Die rechte Abbildung veranschaulicht die Verteilung des Alters, wobei statt dem „Count“ das Gewicht für die Berechnung der Anteile verwendet wird. Da die Summe des Gewichts der Anzahl an Beobachtungen entspricht, ändert sich die Anzahl dadurch nicht.

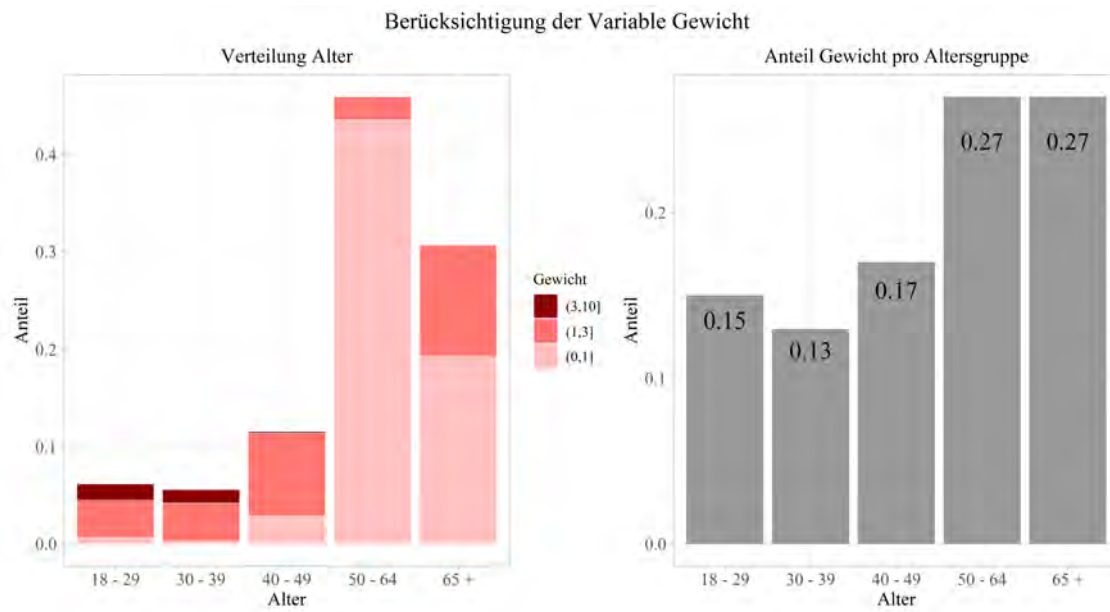


Abbildung 5: Verteilung der Variable *Alter* mit Berücksichtigung des Gewichts in der Befragung einen Monat vor der Wahl

Um die bestehende Diskrepanz der Stichprobe und der tatsächlichen Verteilung des Alters auszugleichen, werden vor allem Beobachtungen in jüngeren Altersgruppen, welche unterrepräsentiert sind, hoch gewichtet. Die meisten Beobachtungen, die ein Alter zwischen 50 und 64 und zum Teil auch über 65 aufweisen, werden hingegen nach unten gewichtet, um die Überrepräsentation auszugleichen. Bei Betrachtung der rechten Abbildung fällt auf, dass die Anteile der Beobachtung pro Altersgruppe unter Berücksichtigung des Gewichts Anteile aufweisen, die nahe der tatsächlichen Verteilung bei der Bundestagswahl 2017 liegen (vgl. Tabelle 1 und 2).

Insgesamt kann festgestellt werden, dass, wie bereits in Kapitel 2.2 erwähnt, keine freiwillige Umfrage frei von Erhebungsfehlern ist. Um die entstandene Verzerrung in Bezug auf das Auswahlverfahren in der Stichprobe zu verringern, wird die von *Civey* generierte Stichprobe verwendet und die berechnete Gewichtung in der Modellierung berücksichtigt. Die folgenden Anwendungen der Modelle aus Kapitel 2 beziehen sich lediglich auf die zweite und dritte Welle, da in der ersten Welle keine Gewichtung vorhanden ist.

Auffallend ist allerdings, dass der Gesamtdatensatz im Vergleich zur resultierenden Stichprobe proportional mehr Unentschlossene enthält. Im gesamten Datensatz sind etwa ein Viertel der Personen unentschlossen, während die daraus entstehende Stichprobe lediglich 13% unentschlossener Wähler*innen aufweist. Das kann zu einer weiteren Verzerrung in der Stichprobe führen. Das heißt, es können weniger Aussagen über Prognosen von beispielsweise jungen unentschlossenen Wähler*innen getroffen

werden. Auch bei Berücksichtigung der Gewichtung anstatt der Anzahl analog zum rechten Diagramm in Abbildung 5 liegt der Anteil der unentschlossenen Befragten bei etwa 12%. Da die Unentschlossenheit im Vordergrund dieser Arbeit ist, kann dieser Teil der Stichprobenziehung als kritisch betrachtet werden.

3.1.2 Beschreibung unentschlossener Wähler*innen anhand von mengenwertigen Daten

In diesem Abschnitt wird der Fokus auf die endogene Variable der Parteipräferenz gelegt und die Eigenschaften dieser Variable untersucht. Zunächst wird dafür die Datenaufbereitung vorgestellt. Wie im vorherigen Kapitel erwähnt, werden lediglich vollständige Angaben in Betracht gezogen. Das bedeutet, in der zweiten Welle werden 12% und in der dritten Welle 18% der Beobachtungen ausgeschlossen. Da bei der Modellierung lediglich Parteien einbezogen werden, die üblicherweise einen Wahlanteil von über 5% aufweisen und somit im Bundestag vertreten sind, werden die Parteiangaben *Eine Andere Partei*, *Sonstige* und *Nichtwähler* als Antwortmöglichkeit sowohl bei den Entschlossenen als auch bei den Unentschlossenen nicht berücksichtigt. In beiden Wellen haben etwa 9% der Befragten eine dieser Möglichkeiten angegeben, wovon in etwa 1,5% nicht beabsichtigen, zur Wahl zu gehen. Personen, die angegeben haben, unentschlossen zu sein und sich dennoch bei der Parteiauswahl für nur eine Partei entschieden haben, werden als entschlossen eingestuft. Das betrifft in etwa ein Viertel der vermeintlich unentschiedenen Personen in beiden Wellen. Ein Grund für den hohen Anteil ist unter anderem der Ausschluss der Angaben *Eine Andere Partei* und *Sonstige*, wodurch bei einer Unentschlossenheit zwischen einer spezifischen Partei und einer Sonstigen die Person als entschlossen kategorisiert wird. Für eine übersichtliche Menge an möglichen Parteikombinationen der Unentschiedenen werden lediglich die Kombinationen berücksichtigt, welche mindestens 20 Ausprägungen aufweisen. Dadurch werden in beiden Wellen circa 5% der unentschlossenen Beobachtungen ausgeschlossen. Letztendlich sind in der zweiten Welle 3964 Beobachtungen und in der dritten Welle 3740 Beobachtungen vorhanden. In jeweils beiden Befragungen sind nach der Datenaufbereitung in etwa 93% der Befragten bezüglich der Wahl entschieden und circa 7% unentschieden. Vor der Datenaufbereitung lag der Anteil der Unentschlossenen in der früheren Befragung bei 13% und in der späteren Befragung bei 14%. Der beschriebene Ausschluss einiger Unentschlossenen ist für die folgende Modellierung notwendig, um eine übersichtliche Darstellung zu ermöglichen. Da einige Beobachtungen aus dem Datensatz ausgeschlossen sind, wird die Variable *Gewicht* neu skaliert, um einen Mittelwert von eins beizubehalten. Somit ist im aufbereiteten Datensatz die Summe des Gewichts wiederum gleich der Anzahl an

Beobachtungen.

Die Angaben der entschlossenen Personen bezüglich der bevorstehenden Wahl ist in Abbildung 6 und 7 auf der linken Seite dargestellt. Die Mengenangabe der unentschlossenen Personen ist dabei auf der rechten Seite veranschaulicht (Abbildung 6 für die zweite Welle und Abbildung 7 für die dritte Welle). Bei den folgenden beiden Darstellungen wird die Gewichtung bereits analog zum rechten Diagramm in Abbildung 5 berücksichtigt, um einen Vergleich der folgenden Prognosen zu ermöglichen.

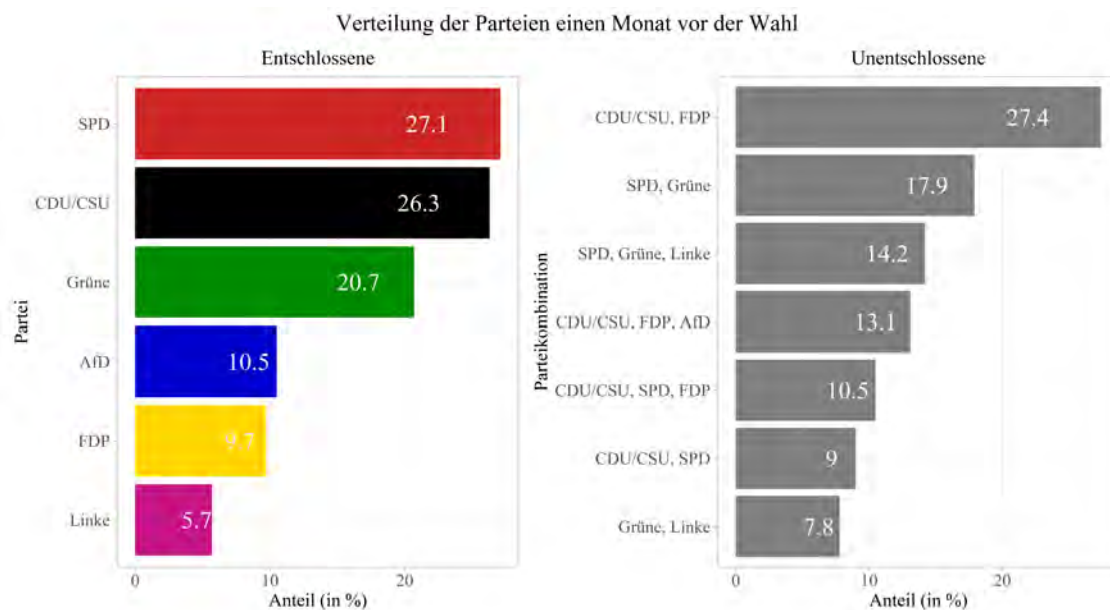


Abbildung 6: Verteilung der Parteien aufgeteilt nach der Entschlossenheit einen Monat vor der Wahl

Es ist zu erkennen, dass unter den Entschlossenen der früheren Befragung 27,1 % die SPD als bevorzugte Partei angegeben haben, dicht gefolgt von der CDU/CSU mit 26,3 %. Die Linke würden lediglich 5,7 % der Befragten bei einer bevorstehenden Wahl wählen. Unter den unentschlossenen Befragten ist die Kombination {CDU/CSU, FDP} am häufigsten (27,4 %) und die Kombination {Grüne, Linke} am seltensten (7,8 %) vertreten.

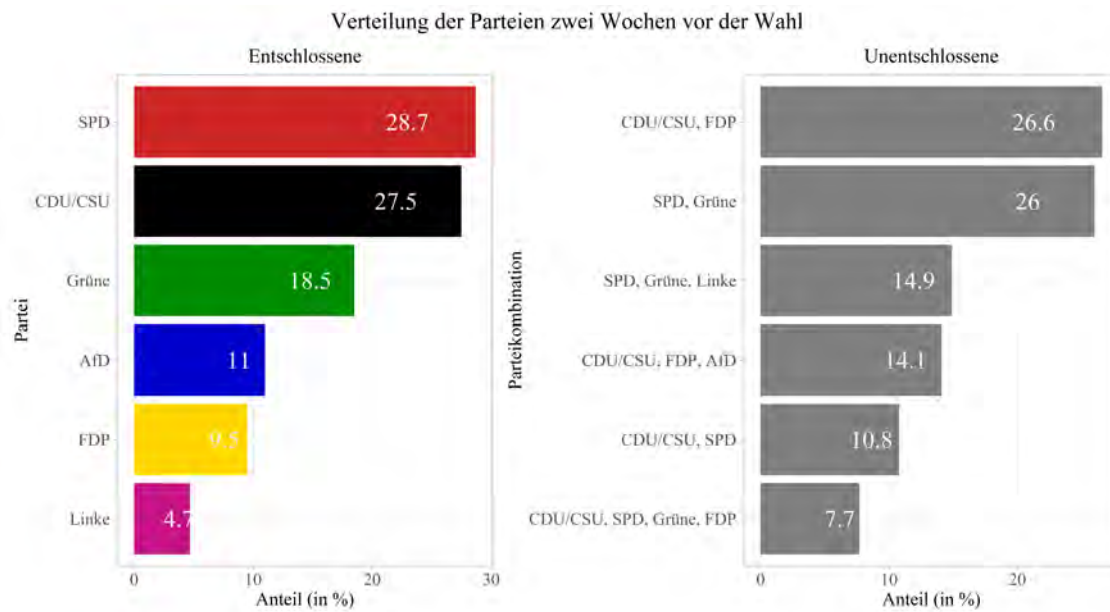


Abbildung 7: Verteilung der Parteien aufgeteilt nach der Entschlossenheit zwei Wochen vor der Wahl

Bei Betrachtung der späteren Befragung unterscheiden sich die Präferenzen der Entschlossenen leicht von der vorherigen Welle. Während die SPD, CDU/CSU und AfD häufiger angegeben wurden, weisen die Grünen, FDP und Linke kleinere Anteile als in der vorherigen Befragung auf. Der größte Unterschied ist dabei, bei der Partei Grüne zu erkennen, die in der dritten Welle einen um 2,2 % geringeren Anteil aufweist. Bei den unentschlossenen Befragten wurde ebenfalls die Kombination CDU/CSU, FDP} mit 26,6 % am häufigsten genannt. Die Parteikombination {SPD, Grüne} weist dabei mit 26,0 % allerdings einen ähnlich großen Anteil auf. Die Reihenfolge der ersten vier Parteimengen ist in beiden Wellen identisch. Dennoch gibt es unterschiedliche Kombinationen ja nach Befragungszeitpunkt, welche mindestens 20 mal angegeben wurden. In beiden Wellen scheinen fast alle unentschiedenen Personen zwischen zwei oder drei Parteien zu schwanken. Auffallend ist außerdem, dass in fast jeder Parteikombination entweder die SPD oder die CDU/CSU ein Element der angegebenen Menge ist. Das heißt, die meisten unentschlossenen Personen schwanken zwischen einer der zwei größten Parteien und mindestens einer weiteren Partei.

Es werden außerdem die bereits erwähnten unausgewogenen Daten in der Zielvariable erkennbar (imbalanced data). Das heißt, die Angaben der Parteipräferenz ist nicht gleich verteilt, sondern es gibt Mehrheitsklassen (SPD, CDU/CSU) und Minderheitsklassen (FDP, Linke). Durch die bereits beschriebene stratifizierte Kreuzvalidierung in Kapitel 2 wird dies in der Random Forest Methodik berücksichtigt.

3.2 Anwendung der Modellierungen

Im folgenden Abschnitt werden die Ansätze, die in Kapitel 2 vorgestellt wurden, an den vorliegenden Daten von *Civey* angewendet, interpretiert und verglichen. Dafür wird zunächst der Ansatz des Random Forests unter der Homogenitätsannahme verfolgt, um eine Punktschätzung zu erhalten. Anschließend wird die Unsicherheit anhand Dempster Bounds in Form von Intervallen veranschaulicht. Die anschließende Interpretation bezieht sich auf die Unterschiede je nach Ein- und Ausschluss unentschlossener Wähler*innen, die Stichprobengewichtung und den zeitlichen Vergleich zwischen den beiden erhobenen Wellen.

Die folgenden Modellierungen werden in dem statistischen Softwareprogramm R (Version 4.2.1) durchgeführt. Das `ranger` package in R bietet eine schnelle Implementierung von Random Forest Modellen mit der Möglichkeit, Gewichte für das Subsampling bei der Entstehung der einzelnen Bäume (Argument `case.weights`) zu übergeben (Wright und Ziegler, 2015). Dabei wird zunächst die stratifizierte Kreuzvalidierung mit fünf Folds durchgeführt. Das bedeutet, der Datensatz der Entschlossenen wird in fünf Teilmengen aufgeteilt, wovon jeder dieser Folds einmal den Testdatensatz darstellt, während die restlichen Beobachtungen den Trainingsdatensatz bilden. Durch die Stratifizierung sind in jedem Fold die Anteile der Parteien nahezu gleich, welche die Anteile der Entschlossenen widerspiegeln (Abbildung 6 und 7). Insgesamt wird fünf mal ein Random Forest modelliert, wobei die Performancebeurteilung anhand des gemittelten Wertes der fünf entstehenden Accuracys erfolgt. Die Accuracy beschreibt den Anteil der richtigen Parteiprognozen im Testdatensatz. Da der Parameter `mtry` als Tuningsparameter behandelt wird, wird die stratifizierte Kreuzvalidierung wiederum 12 mal durchgeführt, da dies der Anzahl an Kovariablen im Modell entspricht. Jede Kreuzvalidierung wird mit einem veränderten `mtry` Parameter durchgeführt, sodass letztendlich 12 gemittelte Accuracy Werte berechnet werden. Anhand des größten Accuracy Wertes kann der optimale Parameter für `mtry` bestimmt werden.

Die Durchführung der Performance Beurteilung einschließlich der *Civey* Gewichtung hat bei der zweiten Welle die höchste Accuracy von 0,30 bei `mtry = 3` und bei der dritten Welle eine Accuracy von 0,33 bei `mtry = 2` ergeben. Das heißt, beispielhaft an der früheren Befragung, dass 30% der Beobachtungen innerhalb der entschlossenen Beobachtungen richtig klassifiziert wurden. Die niedrigen Accuracy Werte lassen sich unter anderem durch den definierten Zustandsraum erklären. Dieser ist als Potenzmenge definiert, wodurch eine Vorhersage schwieriger ist als im binären Fall. Außerdem bedeuten die Werte, dass die berücksichtigten Kovariablen wenig

prädiktiv sind. Das heißt, der Einfluss der sozioökonomischen Variablen auf das Wahlverhalten ist eher gering. Schoen und Zett veranschaulichen zum Beispiel den Einfluss der Variablen *Ost/West* und *Religion* auf das Wahlverhalten im Laufe der Jahre. Dabei ist zu erkennen, dass dieser deutlich abgenommen hat. Beispielsweise hat sich die Wahrscheinlichkeit, mit einer katholischen Konfession die CDU/CSU bei der Bundestagswahl zu wählen, innerhalb von 50 Jahren halbiert (Schoen und Zettl, 2012).

Für die Prognose der unentschlossenen Personen werden alle Beobachtungen der entschiedenen Befragten als Trainingsdatensatz verwendet, um möglichst viele Informationen zu erhalten. Außerdem werden die optimierten *mtry* Parameter bei der Prognose der jeweiligen Welle übernommen. Sowohl bei der stratifizierten Kreuzvalidierung für die Performance-Evaluierung als auch für die Prognose der Unentschlossenen, werden die angegebenen Gewichte bei der Entstehung der Bootstrap-Stichproben anhand der PPS sampling Methodik berücksichtigt. Die Abwägungsmenge der einzelnen unentschlossenen Personen wird dabei, wie in Gleichung 2 beschrieben, normiert, damit lediglich die präferierten Parteien selektiert werden.

Abbildung 8 beschreibt die Prognose beider Wellen für die Bundestagswahl, welche die Ergebnisse des Random Forests für die Unentschlossenen und die Angabe der Entschlossenen umfasst. Bei den Angaben der Entschlossenen werden die Gewichte analog zu den Abbildungen 6 und 7 berücksichtigt.

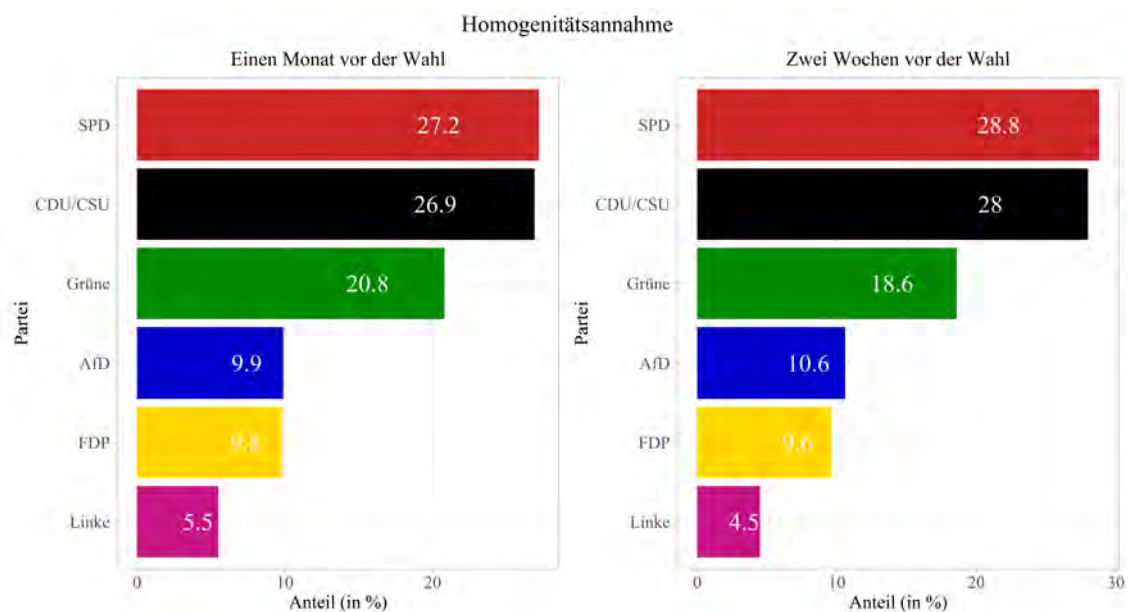


Abbildung 8: Schätzungen des Random Forest Modells unter Homogenitätsannahme - Vergleich der Befragung einen Monat (links) und zwei Wochen (rechts) vor der Wahl

Beim Vergleich der Ergebnisse mit den Angaben ohne unentschiedene Wähler*innen (Abbildung 6 und 7) fallen kleine Abweichungen auf. Während die Parteien SPD, CDU/CSU und FDP unter Einbeziehung der Unentschlossenen in beiden Wellen stärker vertreten sind, weisen die restlichen Parteien einen kleineren Anteil auf. Die größten Differenzen sind dabei in den Parteien AfD (-0,6 bzw. -0,4 %) und CDU/CSU (+0,6 bzw. +0,5 %) zu beobachten. Das heißt, laut der Prognose des Random Forest Modells entscheiden sich unentschlossene Personen tendenziell gegen die AfD und für die CDU/CSU.

Der zeitliche Vergleich ergibt bei der späteren Befragung höhere Schätzwerte bei den Parteien SPD, CDU/CSU und AfD. Die restlichen Parteien weisen niedrigere Werte auf. Die gleiche Erkenntnis ist bereits bei den Abbildungen der Entschlossenen zu erkennen. Die größte Differenz ist analog zu den Angaben der Entschlossenen bei den Grünen zu erkennen, wobei die spätere Befragung einen um 2,2 % geringeren Anteil aufweist. Bei der Interpretation der Prognosen sollte man dennoch vorsichtig sein, da beispielsweise Nonresponse-Strukturen nicht bereinigt sind und die getroffene Annahme umstritten ist.

Für die Beurteilung, inwiefern sich die Gewichtung auf die Performance auswirkt, wurde die in Kapitel 2 dargelegte Simulation der Gewichtungvariable angewendet. Dabei wurden 10.000 Beobachtungen aus der Exponentialverteilung generiert und die Aufteilung in Test- und Trainingsdaten mit dem anschließendem Fitten des Modells insgesamt 100 Mal durchgeführt. Bei jeder der 100 Durchführungen werden zufällig n Beobachtungen aus den 10.000 Beobachtungen gezogen und als Gewichtung in das Modell übergeben. Mit der Simulation soll festgestellt werden, ob sich die Accuracy durch den Eingriff in das Subsampling ändert und ob eine mögliche Änderung auf die vorliegende Gewichtung von *Civey* zurückzuführen ist. Abbildung 9 zeigt zunächst eine mögliche simulierte Gewichtungvariable mit n gezogenen Beobachtungen auf. Hierbei wurden insgesamt $n = 3958$ Beobachtungen generiert, was der Anzahl an Beobachtungen der früheren Befragung entspricht.

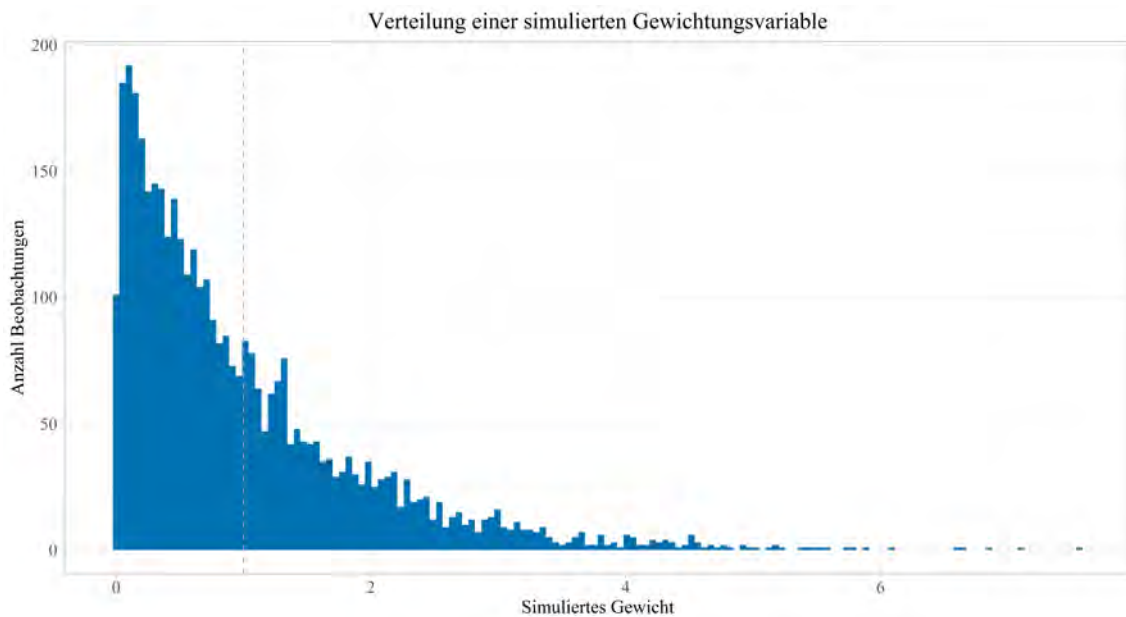


Abbildung 9: Verteilung einer simulierten Gewichtung mit Anzahl Beobachtungen der Befragung einen Monat vor der Wahl

Der Mittelwert ergibt sich bei einer Exponentialverteilung durch $1/\text{Rate}$, wodurch der Mittelwert von eins bei der Wahl von $\text{Rate} = 1$ der *Civey* Gewichtung erhalten bleibt. Aufgrund der Eigenschaften einer Exponentialverteilung, dass es den Wert 0 nicht gibt und die meisten Werte zwischen 0 und 1 liegen, kann die Verteilung der Gewichtungsvariablen von *Civey* nachgeahmt werden. Die Summe der Gewichte als Anzahl der Beobachtungen bleiben dabei in etwa erhalten. Ähnlich zu der zur Verfügung gestellten Gewichtung von *Civey* liegen bei der Simulation zwei Drittel der Gewichte unter eins und die Werte reichen von 0,001 bis 8,6. Es sollen lediglich Aussagen über das Eingreifen des Subsamplings beim Random Forest getroffen werden, weshalb kleine Abweichungen der Gewichtung in Kauf genommen werden können. Zur Verallgemeinerung dieser Ziehung und das Fitten des Modells wird die Simulation 100 Mal durchgeführt. Die Aufteilung in den Test- und Trainingsdatensatz erfolgt bei jedem Mal ebenfalls stratifiziert, wodurch die Verteilung der Parteien der Entschlossenen erhalten bleibt.

Um eine Aussage bezüglich der Veränderung der Performance je nach Umgang des Gewichts zu formulieren, wird die Modellierung sowohl mit als auch ohne Gewichtung durchgeführt. Bei der Gewichtung wird wiederum unterschieden, ob diese von *Civey* gegeben oder simuliert ist. Hierbei sollte darauf geachtet werden, dass sich die Interpretation lediglich auf die vorliegenden Daten und auf das Eingreifen in das Subsampling und die damit einhergehende Performance bezieht. Die Tabellen 3 und 4 beschreiben die Accuracy Werte des gefitteten Random Forest Modells. Die Accuracy

der Simulation ist dabei der gemittelte Wert der 100 berechneten Werte. Um eine Reproduzierbarkeit zu ermöglichen, wird ein sogenannter `seed` in R festgelegt.

Ohne	Civey	Simulation
0.310	0.303	0.303

Tabelle 3: Vergleich der Accuracy Werte einen Monat vor der Wahl

Ohne	Civey	Simulation
0.321	0.327	0.324

Tabelle 4: Vergleich der Accuracy Werte zwei Wochen vor der Wahl

Beim Modell der simulierten Gewichtung hat sich eine Varianz von kleiner 0,001 in beiden Wellen ergeben, wodurch keine große Streuung in den Accuracy Werten erkennbar ist. Die Abweichungen zwischen den Accuracy Werten sind in beiden Wellen kleiner 0,01, wobei die Accuracy ohne Gewichtung in der früheren Befragung und die Accuracy mit der *Civey* Gewichtung in der späteren Befragung am höchsten ist. Aus den geringen Differenzen und keiner klaren Tendenz der beiden Befragungswellen kann geschlossen werden, dass das Eingreifen in das Subsampling bei der Entstehung des Random Forest keine großen Auswirkungen auf die Performance des Modells hat. Dies gilt in Bezug auf die vorliegenden Daten allgemein und ist nicht auf die *Civey* Gewichtung zurück zu führen. Somit ist dies ein sinnvoller Weg, eine Gewichtung in Random Forest Modellen zu berücksichtigen.

Um die komplette Unsicherheit unentschlossener Wähler*innen darzustellen, werden die Dempster Bounds aus Gleichung 3 und 4 herangezogen. Die geschätzten Grenzen vom Minimum und Maximum der Stimmen beider Wellen ohne weitere Annahmen und mit Berücksichtigung des Gewichts sind in Abbildung 10 eingezeichnet.

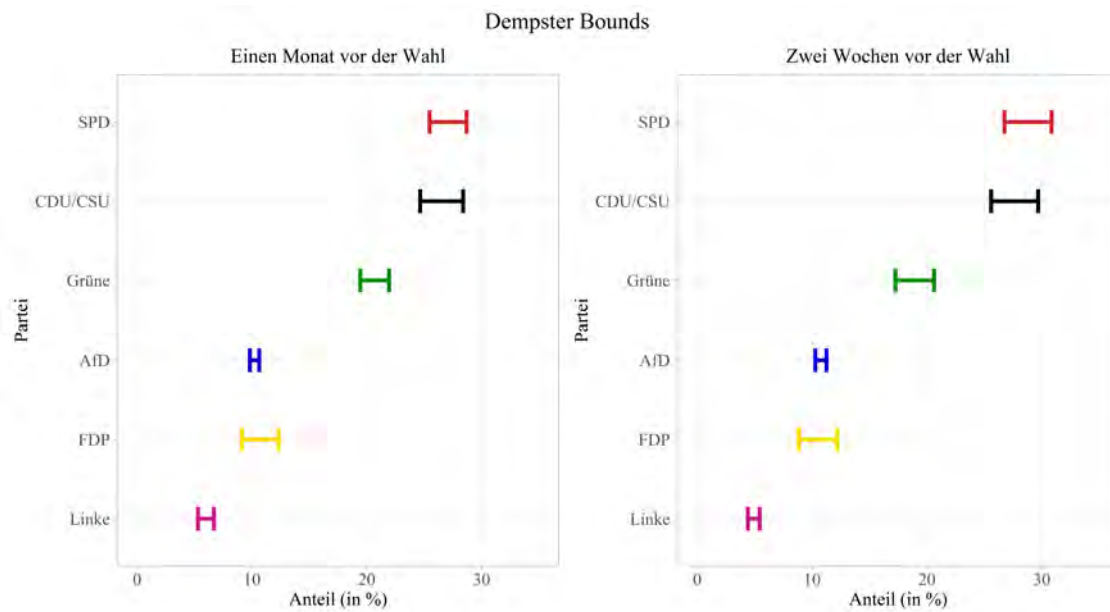


Abbildung 10: Vergleich der Dempster Bounds mit intervallwertigen Schätzungen ohne weitere Annahmen in beiden Wellen

Für die intervallbewerteten Dempster-Grenzen werden die oberen und unteren Wahrscheinlichkeiten mit zwei senkrechten Linien dargestellt, welche zu einem Intervall verbunden werden. Der tatsächliche Anteil an der Wahl liegt zwischen diesen beiden Grenzen. Beispielhaft an der SPD bei der früheren Befragung bedeutet dies, dass abgesehen vom Umfragefehler mindestens 25,5% und maximal 28,7% der Unentschlossenen die SPD wählen werden. Es ist zu erkennen, dass die Parteien SPD, CDU/CSU, Grüne und FDP bei beiden Befragungen die breitesten Intervalle aufweisen. Das heißt, dass hier die größte Unsicherheit der Unentschlossenen auftritt. Auffallend ist die FDP, die im Verhältnis zur Größe der Partei ein breites Intervall aufzeigt. Das deutet darauf hin, dass viele Personen zwischen dieser und anderen Parteien abwägen. Im Fall von CDU/CSU und SPD war bereits in den Abbildungen 6 und 7 zu erkennen, dass die meisten Unentschlossenen eine der beiden Parteien in Betracht ziehen. Dies spiegelt sich auch hier in den breiten Intervallen wieder. Auffallend ist außerdem, dass die Partei AfD das kleinste Intervall aufweist. Das deutet daraufhin, dass die Unsicherheit der Unentschlossenen in dieser Partei am kleinsten ist. Bezüglich des Vergleichs beider Wellen lassen sich in der Breite der Intervalle Unterschiede in den Parteien SPD und Grüne erkennen. In der späteren Befragung scheint bei diesen Parteien die Unsicherheit größer zu sein. Außerdem sind die Intervalle je nach Befragungszeitraum etwas versetzt, was sich durch die Angaben der entschlossenen Wähler*innen ergibt. Während die SPD und CDU/CSU in der dritten Welle sowohl bei der unteren als auch oberen Grenze höhere Anteile

aufweisen, tritt bei den Parteien Grüne, FDP und Linke der gegenteilige Fall ein. Somit verschiebt sich bei der späteren Befragung je nach Partei der minimale und maximale Anteil an Stimmen der Unentschlossenen.

Insgesamt hat sich die Methodik als einfach anwendbar erwiesen und führt zu plausiblen Wahlergebnissen. Die geschätzten Grenzwerte zeigen jedoch Extremfälle, die zu groben Ergebnissen führen.

Für eine Eingrenzung der Streuung der Dempster Bounds werden weitere Einschränkungen herangezogen. Abbildung 11 illustriert sowohl die aus Abbildung 10 erzeugten Intervalle in einer transparenten Darstellung als auch die mit dem 80/20 Ansatz entstandenen Intervalle aus Gleichung 5 und 6. Anhand der Überlappung wird deutlich, inwiefern diese Annahme die Grenzen aus dem vorherigen Abschnitt einengt. Der abgebildete schwarze Punkt beschreibt die Prognose, die durch die 50/50 Annahme entsteht.

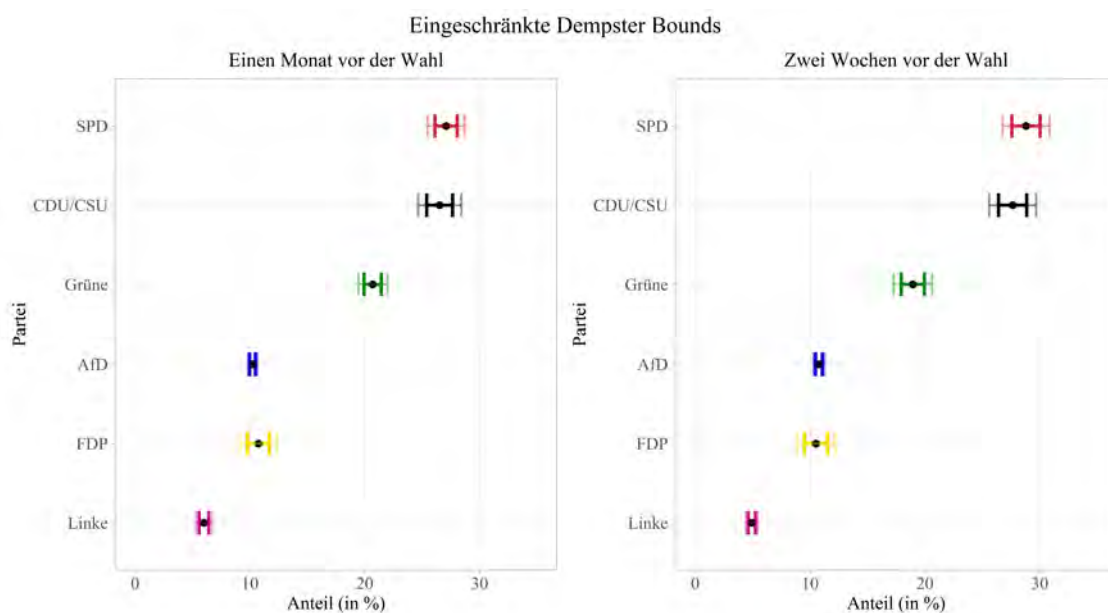


Abbildung 11: Vergleich der Dempster Bounds mit intervallwertigen Schätzungen ohne, mit 80/20 und mit 50/50 Annahme in beiden Wellen

Diese eher schwache Annahme verengt die Grenzen bereits und führt zu den in Abbildung 11 dargestellten Prognosen. Gerade bei breiteren Intervallen wie bei SPD, CDU/CSU, Grüne und FDP wird dies deutlich. Die neu erzeugten Grenzen der AfD und Linke weisen im Vergleich zu den Dempster Grenzen ohne Annahmen nur kleine Änderungen auf. Beim Vergleich der Schätzungen des Random Forest Modells und der Dempster Bounds fällt auf, dass die Schätzungen aller Parteien innerhalb

der Intervalle unter der 80/20 Annahme liegen. Dabei weisen die Prognosen der SPD, CDU/CSU und Grüne Werte auf, die nahe der Schätzung unter der 50/50 Annahme sind. Die Parteien FDP, AfD und Linke orientieren sich an der unteren Grenze der Dempster Bounds mit der 80/20 Annahme. Die 80/20 Annahme kann als Kompromiss zwischen der Homogenitätsannahme und den Dempster Bounds gesehen werden, da die Intervalle kleiner sind, ohne zu strenge Annahmen treffen zu müssen.

Da eine große Anzahl von Parteipartitionierungen möglich sind, werden im Folgenden lediglich sechs Kombinationen aufgezeigt, die sich aus den Angaben der Unentschiedenen ergeben. Es werden die Parteikombinationen der Unentschiedenen betrachtet, die sowohl in der früheren als auch in der späteren Befragung mindestens 20 Mal vorgekommen sind, um Aussagen über die Veränderung in der Zeit treffen zu können. Das heißt, es werden Kombinationen vom Zustandsraum ohne politischen Zusammenhang gewählt. Abbildung 12 veranschaulicht die Prognosen der sechs Parteikombinationen für die Ausführung des Koalitionsgedankens aus Abschnitt 2.4.3. Die graue vertikale Linie markiert dabei den Schwellwert von 50 %, also dem Anteil an Stimmen, der für eine mögliche Koalitionsbildung notwendig ist.

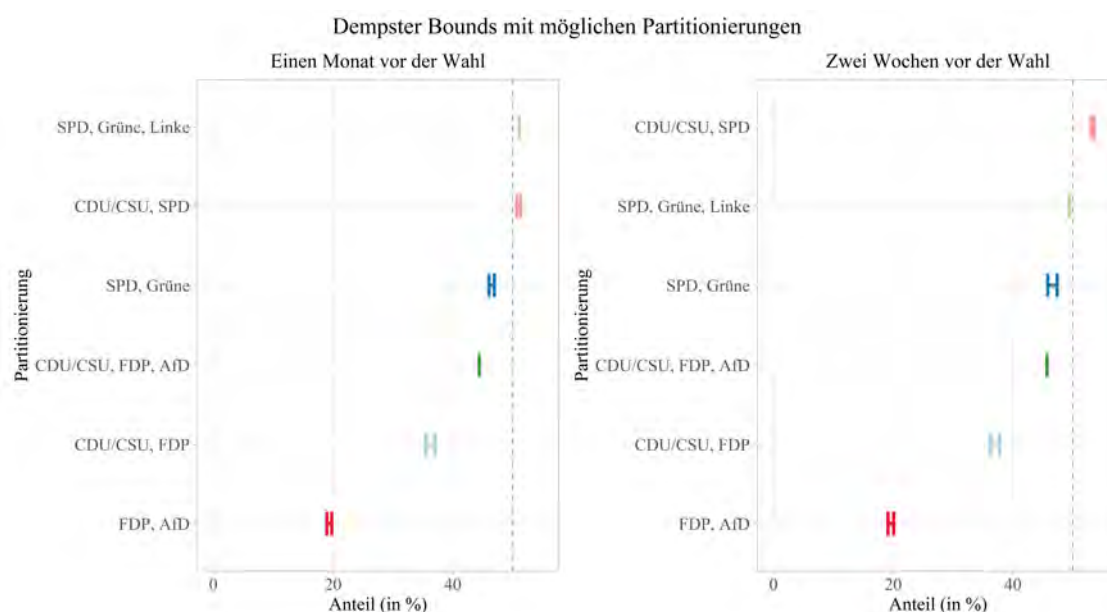


Abbildung 12: Vergleich der Dempster Bounds mit der Partitionierung des politischen Raumes – Darstellung möglicher Koalitionen in beiden Wellen

Die Grenzen zeigen wie bei den Dempster Bounds ohne Annahmen das Minimum und das Potential der Partitionierungen, abgesehen vom Umfragefehler. Im Unterschied zu Abbildung 10 liegen die Grenzen hier näher zusammen, da die Unsicherheit

mit dem neuen Zustandsraum reduziert wird. Teilweise entspricht die untere Grenze gleich der oberen Grenze, wie im Falle der Kombinationen {SPD, Grüne, Linke} und {CDU/CSU, FDP, AfD}. Das liegt daran, dass keine unentschlossene Person eine Abwägungsmenge mit diesen Parteien und mindestens einer weiteren aufweist und somit keine Person in der oberen Grenze, aber nicht in der unteren Grenze vertreten ist. In Bezug auf die Interpretation liegt der Fokus vor allem auf dem Erreichen von mindestens 50 % der Stimmen, welche für die Bildung einer Koalition notwendig sind. Die Ergebnisse zeigen auf, dass bei der früheren Befragung zwei Koalitionen sowohl mit der unteren als auch mit der oberen Grenze über diesen Schwellwert kommen würden. Das betrifft die Parteikombinationen {SPD, Grüne, Linke} und {CDU/CSU, SPD}. Bei der späteren Befragung ist nach den Ergebnissen der oberen und unteren Grenze ohne Berücksichtigung des Umfragefehlers lediglich die große Koalition dazu in der Lage, eine neue Regierung zu bilden. Der hohe Anteil der großen Koalition in beiden Wellen stammt zum einen von den Ergebnissen der Entschlossenen, welche eine der beiden Parteien angegeben haben. Zum anderen liegt dies an der Tatsache, dass die Kombination der beiden Parteien oftmals Elemente der angegebenen Abwägungsmengen der Unentschlossenen sind (vgl. Abbildung 6 und 7). Am weitesten entfernt von den erforderlichen 50 % ist die Kombination {FDP, AfD}, da eine Koalition von zwei kleinen Parteien unwahrscheinlich ist. Auffallend ist, dass die tatsächliche Regierung, die 2021 zwischen der SPD, den Grünen und der FDP entstanden ist, kein Teil der Dempster Bounds ist, da nur wenige Wähler*innen unentschlossen zwischen diesen sind. Dempster Bounds in Form von Koalitionen bieten eine erste Idee, durch die Änderung des Zustandsraumes aus anfänglich impräzisen Informationen präzise Angaben zu erhalten.

3.3 Interpretation der Ergebnisse

Die zentrale Frage, die in dieser Arbeit untersucht werden soll, ist, inwiefern die Ergebnisse abweichen, wenn unentschlossene Wähler*innen in die Modellierung einbezogen werden. Wie in Kapitel 2.2 erwähnt, entsteht durch das Ziehen einer Stichprobe aus der Grundgesamtheit eine Verzerrung. Sowohl die Einbeziehung als auch der Ausschluss unentschlossener Personen führt dabei zu einer Verzerrung. Beim Einschluss liegen einerseits mehr Informationen vor, andererseits müssen für punktwertige Schätzungen weitere Annahmen getroffen werden, die wiederum nicht überprüfbar sind. Durch den Ausschluss werden jedoch wertvolle Informationen vernachlässigt und es kommt zu Verzerrungen, insbesondere wenn ein systematischer Unterschied zwischen entschlossenen und unentschlossenen Personen besteht. Um eine Annäherung der Antwort zu erhalten, in welchem Fall eine geringere Verzerrung

vorliegt, wird die sogenannte Sonntagsfrage herangezogen, welche ebenfalls in den Daten von *Civey* vorhanden ist.

Im Folgenden wird die in den vorherigen Abschnitten betrachtete Möglichkeit zur punktwertigen Prognose der Wahlergebnisse unter Einschluss der Unentschlossenen sowohl mit den Angaben der Entschlossenen als auch mit den Angaben von allen Befragten bei der Sonntagsfrage verglichen. Diese lautet: „Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?“. Laut Fenwick wird die Unentschlossenheit reduziert, wenn die Teilnehmer gefragt werden, für welche Partei sie stimmen würden, wenn die Wahl am Tag der Umfrage stattfinden würde, anstatt der Frage nach der Wahl im Oktober (Fenwick u. a., 1982). Somit werden die unentschlossenen Befragten zu einer Entscheidung „gezwungen“, welche sich auf Grund der eigentlichen Unentschlossenheit bis zur tatsächlichen Wahl noch ändern kann. Dennoch dient diese Angabe als Referenz für den Vergleich des Umgangs mit unentschlossenen Personen. Die folgenden Aussagen beziehen sich auf die Ergebnisse der betrachteten Stichprobe und den von *Civey* gegebenen Gewichten.

Abbildung 13 beschreibt die Angaben der Entschlossenen, die Prognosen der Unentschlossenen zusammen mit den Entschlossenen unter der Homogenitätsannahme und die Angabe von allen Befragten bei der Sonntagsfrage. Dabei wird bei den Beobachtungen der Entschlossenen und bei allen Befragten bei der Sonntagsfrage die Gewichtung einbezogen, indem statt dem „Count“ das jeweilige Gewicht in der Prognose analog zu Abbildung 6 und 7 berücksichtigt wird. Die linke Seite zeigt die Ergebnisse der Befragung einen Monat vor der Wahl und die rechte Seite die Ergebnisse zwei Wochen vor der Wahl.

Vergleich der punktwertigen Prognosen

Einen Monat und zwei Wochen vor der Wahl

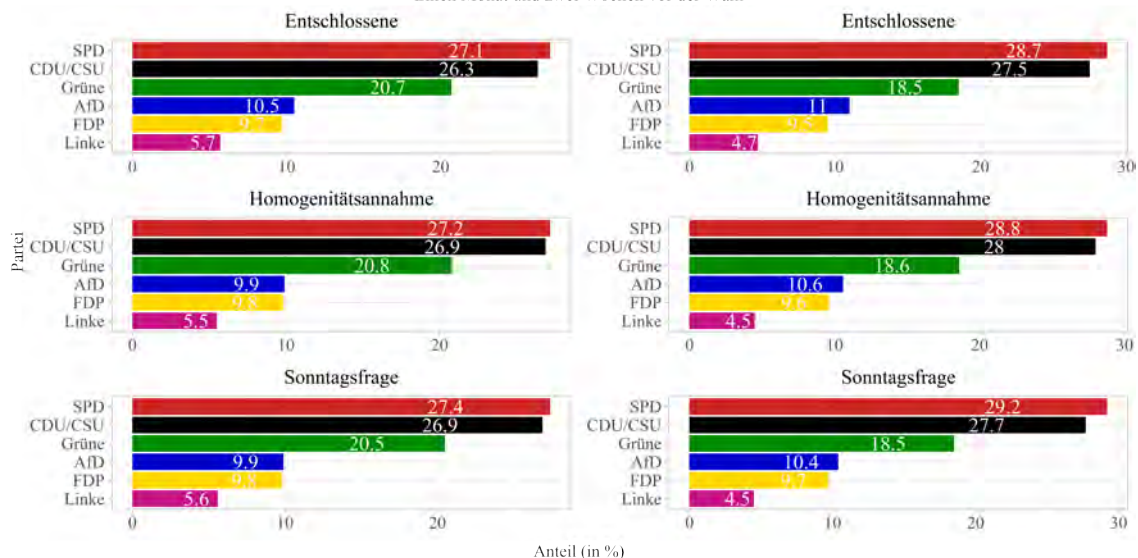


Abbildung 13: Vergleich der Modelle mit unterschiedlichem Umgang der unentschlossenen Wählerinnen in der Befragung einen Monat vor der Wahl (links) und zwei Wochen vor der Wahl (rechts)

Anhand der Abbildung ist zu erkennen, dass die Ergebnisse der Entschlossenen und die der Sonntagsfrage leichte Abweichungen aufzeigen. Während die Anteile der Parteien Grüne, AfD und Linke ohne unentschlossene Wähler*innen in beiden Wellen größer ausfallen, weisen die restlichen Parteien geringere Werte als bei der Angabe der Sonntagsfrage auf. Daraus lässt sich schließen, dass unentschiedene Personen tendenziell die SPD, CDU/CSU und FDP präferieren, wenn sie zu dem Zeitpunkt der Unentschlossenheit eine Partei wählen müssten. Die größten Unterschiede zwischen den Angaben der Entschlossenen und allen Befragten bei der Sonntagsfrage sind anhand der AfD (-0,6 %) in beiden Wellen, der CDU/CSU (+0,6 %) in der früheren Befragung und der SPD (+0,5 %) in der späteren Befragung zu erkennen. Das heißt, Unentschlossene würden sich, wenn sie zum Zeitpunkt der Unentschlossenheit eine Entscheidung treffen müssten, tendenziell gegen die AfD in beiden Wellen und für die CDU/CSU oder SPD, je nach Befragungszeitpunkt, entscheiden. Die Unentschlossenheit könnte somit mit einer bestimmten Parteipräferenz oder Ablehnung zusammenhängen. Betrachtet man zusätzlich die eingeschränkten Dempster-Grenzen mit der 80/20 Annahme, so stellt man analog zu den Prognosen der Homogenitätsannahme fest, dass die Angaben der Sonntagsfrage bei den Parteien AfD, FDP und Linke Ergebnisse nahe der unteren Grenze aufweisen. Die Ergebnisse der Sonntagsfrage für die Parteien CDU/CSU, SPD und Grüne sind dabei ähnlich zum mittleren Wert (50/50 Annahme) der Dempster Bounds. Dies ist sowohl bei der früheren

als auch bei der späteren Befragung beobachtbar. Insgesamt liegen alle Werte der Sonntagsfrage innerhalb der Grenzen unter der 80/20 Annahme, wodurch die Wahl von $s = 0,8$ für die Darstellung der Dempster Bounds sinnvoll erscheint. Bei der früheren Befragung bei den Parteien CDU/CSU und Grüne bzw. bei der späteren Befragung bei den Parteien SPD und Grüne weist der „naive“ Ansatz durch die 50/50 Annahme näher an der Sonntagsfrage liegende Werte auf als der Ansatz mit einer Vernachlässigung unentschlossener Wähler*innen.

Bei Betrachtung der Ergebnisse einschließlich der Unentschlossenen unter der Homogenitätsannahme fällt auf, dass vor allem die Prognosen der CDU/CSU, AfD und FDP in der früheren Welle und der Linken in der späteren Welle sehr ähnliche Werte zur Sonntagsfrage aufweisen. Die Prognosen der Partei Grüne bei der früheren Befragung und CDU/CSU, Grüne und AfD bei der späteren Befragung werden anhand der Homogenitätsannahme etwas überschätzt. Die Parteien SPD und Linke (bzw. SPD und FDP) werden hingegen unterschätzt. Die größten absoluten Differenzen weisen bei der früheren Welle die Grünen (0,3%) und bei der späteren Welle die SPD (0,4%) auf. Bei fast allen Parteien ist eine Annäherung an den Prognosen der Sonntagsfrage durch die Berücksichtigung der Unentschlossenen zu erkennen. Die einzigen Parteien, die nach den Prognosen unter Homogenitätsannahme Werte ergeben, die weiter von der Sonntagsfrage entfernt sind als die Angaben der Entschlossenen, sind in der früheren Befragung die Grünen und in der späteren Befragung die CDU/CSU. Allerdings ist in beiden Fällen lediglich eine Differenz von 0,1% zu vermerken. In der Stichprobe der früheren Welle, vor allem bei den Parteien CDU/CSU und AfD, ist eine deutliche Annäherung der Sonntagsfrage unter Berücksichtigung der unentschlossenen Wähler*innen zu erkennen. Analog gilt dies für die Partei AfD in der späteren Welle. Das heißt, bei den vorliegenden Stichproben macht es insbesondere für die Parteien AfD bei beiden Wellen und CDU/CSU bei der Befragung einen Monat vor der Wahl einen Unterschied, ob unentschlossene Wähler*innen bei der Prognose berücksichtigt werden. Würde man diese nicht berücksichtigen, dann gäbe es eine Überrepräsentation der AfD in beiden Wellen und eine Unterrepräsentation der CDU/CSU in der früheren Welle.

Obwohl eine vollständige Homogenität in der Praxis nicht gegeben ist, ergeben die meisten daraus resultierenden Prognosen Werte, die nahe den Angaben der Sonntagsfrage liegen. Die Anwendung der Machine Learning Methodik ist dabei eine gute Möglichkeit, alle Informationen des Datensatzes zu berücksichtigen und die Stichprobengewichtung einfließen zu lassen. Vor allem in Kombination mit den intervallwertigen Ergebnissen der Dempster Bounds unter der 80/20 Annahme können fundierte Aussagen über die Prognosen der unentschlossenen Wähler*innen getroffen

werden.

Bezüglich der unterschiedlichen Zeitpunkte kann festgestellt werden, dass die Popularität der Parteien CDU/CDU, SPD und AfD in einer Umfrage kurz vor der Bundestagswahl zunimmt, während die der Grünen, FDP und Linken abnimmt. Dies ist in allen aufgezeigten Modellierungen zu erkennen. Auch eine Koalition, welche die CDU/CSU und SPD beinhaltet, ist bei der späteren Befragung wahrscheinlicher als bei der früheren Umfrage.

4 Fazit

Diese Arbeit befasst sich mit der Fragestellung, inwieweit Unterschiede entstehen, wenn unentschiedene Wähler*innen für Prognosen im Mehrparteiensystem berücksichtigt werden. Dabei ist außerdem von Interesse, wie unterschiedliche Ansätze bei einer adäquaten Berücksichtigung unentschlossener Personen abschneiden. Hierbei liegt der Schwerpunkt sowohl auf der Evaluierung verschiedenster Methodiken bei einer angemessenen Berücksichtigung unentschlossener Befragten, als auch auf dem Vergleich der daraus resultierenden Ergebnissen. Die deskriptive Aufarbeitung basierend auf stichprobentheoretischen Grundlagen ermöglicht dabei eine geeignete Berücksichtigung unentschlossener Wähler*innen in Form von mengenwertigen Daten. Für die Prognose unentschlossener Wähler*innen muss zwischen groben Ergebnissen ohne oder mit schwachen Annahmen und der Rechtfertigung von stärkeren Annahmen bei punktwertigen Ergebnissen entschieden werden. Für punktwertige Prognosen wurde die Homogenitätsannahme mit einer Random Forest Schätzung herangezogen und für intervallwertige Prognosen wurden Dempster Bounds mit keinen und weiteren Einschränkungen dargestellt.

Die von *Civey* erstellte Umfrage zur Bundestagswahl 2017 enthält Abwägungsmengen und soziodemografische Kovariablen für jede Beobachtung. Dabei weist jede Beobachtung eine Gewichtung auf, die aufgrund der bleibenden Differenz der gezogenen quotierten Stichprobe aus der Grundgesamtheit entstanden ist. Das Ziel ist hierbei, Verzerrungen des Umfragefehlers zu reduzieren, weshalb die bereitgestellte Gewichtung von *Civey* in allen Modellierungsmöglichkeiten berücksichtigt wird. Die Anwendung des Random Forests bietet die Möglichkeit, Gewichte in die Bootstrap-Stichprobengenerierung der einzelnen Bäume einfließen zu lassen. Im vorliegenden Datensatz konnte festgestellt werden, dass die Berücksichtigung der Gewichtungsvariable bei der Entstehung des Waldes keinen negativen Einfluss auf die Accuracy hat. Die Simulation hat dabei ergeben, dass dies nicht auf die von *Civey* gegebene Gewichtung zurückzuführen ist, sondern allgemein gilt. Bei der Bildung der Dempster Bounds fließt die Gewichtung direkt ein.

Die Ergebnisse der Wahlprognosen zeigen Ähnlichkeiten zwischen den meisten Schätzungen des konventionellen Ansatzes, in Form der von den Entschlossenen bereitgestellten Informationen, und des epistemischen Ansatzes auf. Dennoch lassen sich auch einige Unterschiede feststellen. Unentschlossene Wähler*innen scheinen sich in Bezug auf die erhobenen Angaben der Sonntagsfrage tendenziell je nach Befragungswelle für die größeren Parteien CDU/CSU bzw. SPD und gegen die AfD zu entscheiden. Das heißt, die Unentschlossenheit könnte mit einer Parteipräferenz zusammenhängen, wodurch der Ausschluss unentschlossener Wähler*innen zu einer

Unterrepräsentation der Partei CDU/CSU bzw. SPD und einer Überrepräsentation der AfD führt. Die Frage, inwiefern unentschlossene Wähler*innen berücksichtigt werden sollten, lässt sich partiell beantworten. Bei den vorliegenden Stichproben, insbesondere bei den Parteien AfD bei beiden Befragungen und CDU/CSU bei der früheren Befragung, ist ein Unterschied der Prognosen anhand der Homogenitätsannahme bei Einschluss unentschlossener Wähler*innen zu erkennen. Bei der Betrachtung der Dempster Bounds mit einer 80/20 Einschränkung liegen die Prognosen des Random Forest Modells und die Angaben von allen Befragten bei der Sonntagsfrage innerhalb der berechneten Grenzen. Daher erscheint die 80/20 Annahme eine sinnvolle Wahl zu sein.

Zusammenfassend lässt sich sagen, dass bei einer adäquaten Berücksichtigung unentschlossener Wähler*innen viele Möglichkeiten entstehen, Prognosen bezüglich der Wahlentscheidung zu erstellen. Je nach Breite der Ergebnisse und getroffenen Annahmen können punkt- oder intervallbasierte Prognosen erstellt werden. In der Arbeit wurden dabei zwei Möglichkeiten vorgestellt, welche die Stichprobengewichtung berücksichtigen können: die Homogenitätsannahme und die Dempster Bounds. Diese beruhen auf verschiedenen starken Annahmen und ergeben demnach unterschiedlich grobe Prognosen. Neben den vorgestellten Ansätzen, sind weitere impräzise Modelle zur Schätzung von Wahlprognosen möglich. Die Einbeziehung weiterer Informationsquellen, wie zum Beispiel Expertenwissen oder Ergebnisse vorheriger Wahlen, könnten die Genauigkeit der Prognosen und die zugrunde liegenden Annahmen verbessern. Da die Unsicherheitsstruktur von Wahlmodellen sehr komplex ist, bietet dies ein breites Feld für weitere Forschung, unter anderem im Bereich des Machine Learnings.

Literatur

- Arcuri, Luciano u. a. (2008). „Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters“. In: *Political Psychology* 29(3), S. 369–387.
- Bär, Katja (2017). *Wie Koalitionssignale die Wahl beeinflussen*. URL: <https://idw-online.de/de/news676324> (besucht am 26.10.2022).
- Bauer, Alexander u. a. (2022). „Mundus vult decipi, ergo decipiatur: Visual Communication of Uncertainty in Election Polls“. In: *PS: Political Science & Politics* 55(1), S. 102–108.
- Bhavsar, Hetal und Amit Ganatra (2012). „A comparative study of training algorithms for supervised machine learning“. In: *International Journal of Soft Computing and Engineering (IJSCE)* 2(4), S. 2231–2307.
- Biemer, Paul P (2010). „Total survey error: Design, implementation, and evaluation“. In: *Public opinion quarterly* 74(5), S. 817–848.
- Breiman, Leo (1996). „Bagging predictors“. In: *Machine learning* 24(2), S. 123–140.
- Breiman, Leo (2001). „Random forests“. In: *Machine learning* 45(1), S. 5–32.
- Breiman, Leo u. a. (2017). *Classification and regression trees*. Routledge.
- Bundeswahlleiter, Der (2017). *Bundestagswahl 2017: 61,5 Millionen Wahlberechtigte*. URL: https://www.bundeswahlleiter.de/info/presse/mitteilungen/bundestagswahl-2017/01%5C_17%5C_wahlberechtigte.html (besucht am 26.10.2022).
- Caruana, Edward Joseph u. a. (2015). „Longitudinal studies“. In: *Journal of thoracic disease* 7(11).
- Couper, Mick P (2000). „Web surveys: A review of issues and approaches“. In: *The Public Opinion Quarterly* 64(4), S. 464–494.
- Couso, Inés und Didier Dubois (2014). „Statistical reasoning with set-valued information: Ontic vs. epistemic views“. In: *International Journal of Approximate Reasoning* 55(7), S. 1502–1518.
- Dempster, Arthur P (2008). „Upper and lower probabilities induced by a multivalued mapping“. In: *Classic works of the Dempster-Shafer theory of belief functions*. Springer, S. 57–72.
- Fenwick, Ian u. a. (1982). „Classifying undecided voters in pre-election polls“. In: *Public Opinion Quarterly* 46(3), S. 383–391.
- Han, Sunwoo, Brian D Williamson und Youyi Fong (2021). „Improving random forest predictions in small datasets from two-phase sampling designs“. In: *BMC medical informatics and decision making* 21(1), S. 1–9.

-
- Kosmidis, Spyros und Georgios Xezonakis (2010). „The undecided voters and the economy: Campaign heterogeneity in the 2005 British general election“. In: *Electoral Studies* 29(4), S. 604–616.
- Kreiss, Dominik und Thomas Augustin (2020). „Undecided voters as set-valued information—towards forecasts under epistemic imprecision“. In: *International Conference on Scalable Uncertainty Management*. Springer, S. 242–250.
- Kreiss, Dominik und Thomas Augustin (2021). „Towards a Paradigmatic Shift in Pre-election Polling Adequately Including Still Undecided Voters—Some Ideas Based on Set-Valued Data for the 2021 German Federal Election“. In: *arXiv*.
- Kreiss, Dominik, Malte Nalenz und Thomas Augustin (2020). „Undecided voters as set-valued information-machine learning approaches under complex uncertainty“. In: *ECML/PKDD 2020 Workshop on Uncertainty in Machine Learning*.
- Merz, Stefan (2010). *Das Kreuz mit dem Kreuz: Unentschlossene Wähler in Deutschland*. Nomos Verlagsgesellschaft mbH & Co. KG.
- Müller, Claudia (2021). *ARD-Vorwahlumfragen: SPD auch in MV und Berlin vorn*. URL: <https://www.tagesschau.de/inland/deutschlandtrend/vorwahlumfragen-mecklenburg-vorpommern-berlin-101.html> (besucht am 26.10.2022).
- Müntefering, Eva-Maria (2016). „Boosted Random Forest“. Magisterarb.
- Oscarsson, Henrik und Maria Oskarson (2019). „Sequential vote choice: Applying a consideration set model of heterogeneous decision processes“. In: *Electoral Studies* 57, S. 275–283.
- Probst, Philipp, Marvin N Wright und Anne-Laure Boulesteix (2019). „Hyperparameters and tuning strategies for random forest“. In: *Wiley Interdisciplinary Reviews: data mining and knowledge discovery* 9(3).
- Refaeilzadeh, Payam, Lei Tang und Huan Liu (2009). „Cross-validation.“ In: *Encyclopedia of database systems* 5, S. 532–538.
- Richter, Gerrit, Tobias Wolfram und Charlotte Weber (2022). „Die Statistische Methodik von Civey; Eine Einordnung im Kontext gegenwärtiger Debatten über das Für und Wider internetbasierter nicht-probabilistischer Stichprobenziehung“. In: *Statistische Methodenlehre* 24(1), S. 1–24.
- Rubin, Donald B (2004). *Multiple imputation for nonresponse in surveys*. Bd. 81. John Wiley & Sons.
- Sasikala, S, S Bharathidason und C Jothi Venkateswaran (2015). „Improving classification accuracy based on random forest model through weighted sampling for noisy data with linear decision boundary“. In: *Indian Journal of Science and Technology* 8(S8), S. 614–619.
- Schoen, Harald und Christian Zettl (2012). „Sozialstruktur und Wählerverhalten Kapitel 6“. In: *Wählerverhalten in der Demokratie: eine Einführung* 3533, S. 149.

-
- Skinner, Chris J (2014). „Probability proportional to size (PPS) sampling“. In: *Wiley StatsRef: Statistics Reference Online*, S. 1–5.
- Statista (2009). *Wann werden Sie voraussichtlich entscheiden, welche Partei Sie bei der kommenden Bundestagswahl im September wählen werden?* URL: <https://de.statista.com/statistik/daten/studie/13497/umfrage/zeitpunkt-der-wahlentscheidung-fuer-bundestagswahl/> (besucht am 26.10.2022).
- Tversky, Amos (1972). „Elimination by aspects: A theory of choice.“ In: *Psychological review* 79(4), S. 281–299.
- Westerhoff, Nikolas (2022). *Deutschland vor der Wahl*. URL: <https://www.deutschlandfunk.de/deutschland-vor-der-wahl-118.html> (besucht am 26.10.2022).
- Wlezien, Christopher (2010). „Election campaigns“. In: *Comparing democracies* 3, S. 98–117.
- Wright, Marvin N und Andreas Ziegler (2015). „ranger: A fast implementation of random forests for high dimensional data in C++ and R“. In: *arXiv*.
- Zeng, Xinchuan und Tony R Martinez (2000). „Distribution-balanced stratified cross-validation for accuracy estimation“. In: *Journal of Experimental & Theoretical Artificial Intelligence* 12(1), S. 1–12.

A Elektronischer Anhang

Die Analyse wurde mit dem Softwareprogramm R (Version 4.2.1) durchgeführt. Der dazugehörige Code kann unter <https://github.com/MarenKellerer/Masterarbeit> abgerufen werden.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Ort, Datum

Unterschrift