

Zur Multiplen Imputation bei Item Non-response am Beispiel eines Analysemodells zum Abstimmungsverhalten bei Bürgerentscheiden

Multiple Imputation Concerning Item Nonresponse Using the Example of an Analysis Model for Direct Democratic Voting Behaviour

Martin Messingschlager und Marcel Preisig

Zusammenfassung

In diesem Artikel werden Fragestellungen unterschiedlicher Ebenen behandelt. Zum einen befasst sich der Aufsatz mit dem politischen Verhalten bei direktdemokratischen Abstimmungen auf kommunaler Ebene. Konkret wird ein multivariates Logitmodell geschätzt, das themenspezifische Variablen als wichtigste Determinanten eines in Lichtenfels (Oberfranken) stattgefundenen Bürgerentscheids identifiziert. Die zweite Ebene, die zugleich den Kern der Arbeit darstellt, behandelt einen Vergleich von fall-reduzierendem Vorgehen gegenüber einer Imputation als Korrekturmethode auf Basis des realwissenschaftlichen Modells der ersten Fragestellungsebene. Die Verknüpfung der realwissenschaftlichen mit einer methodischen Fragestellung führt dabei vor Augen, wie wichtig ein sensibler Umgang mit fehlenden Daten – auch in den Sozialwissenschaften – ist. Der erhobene Daten-

Abstract

In this article, we deal with problems of two different natures. On the first level, the paper handles with direct democratic voting behavior on local level, concretely via logistic regression models which identify issue oriented variables as the most important determinants. The second level, which is the actual core of this article, discusses two approaches for dealing with missing values: case-reducing versus imputation of the missing data. The linkage between a real scientific and a methodological question shows the importance of a sensitive handling of missing data – not least in social sciences. Our sample is used for the comparison of Completes-Cases-Analyses with Multiple Imputation. By means of a procedure which combines the advantages of a simulation with those of real data we create a valid base for our comparison. The analytical and graphical results show clearly



satz dient als Ausgangspunkt für einen Vergleich der Complete-Cases-Methode und der Multiplen Imputation als Ergänzungsverfahren bei Item Nonresponse. Mit Hilfe eines Verfahrens, das die Vorteile von Simulationen mit denen realer Daten kombiniert, wird eine aussagekräftige Grundlage für den Methodenvergleich geschaffen. An den Ergebnissen, die sowohl grafisch als auch analytisch aufbereitet werden, wird deutlich, dass sich die Multiple Imputation in der vorliegenden Stichprobe zur Korrektur von Item Nonreponse im Vergleich zur Complete-Cases-Analyse empfiehlt.

that it's appropriate to use Multiple Imputation instead of using a Complete-Cases-Analysis.

1 Einleitung

Fehlende Werte stellen ein auch in den Sozialwissenschaften hinlänglich bekanntes Problem dar: So werden Datennutzer in nahezu allen Surveys damit konfrontiert, dass Befragte die Beantwortung einzelner Fragen, oder gar die gesamte Teilnahme, verweigern oder aus Unwissenheit keine Antworten geben.¹ Neben einer Verringerung der Fallzahl und dem damit verbundenen Informationsverlust stellt sich bei einem solchen Nonresponse ebenso die Frage, ob der Ausfall lediglich zufälliger oder eher systematischer Natur ist. Letzterer wiederum kann zu verzerrten Schätzparametern führen, was die Qualität der Datenbasis zusätzlich verringert.

Der vorliegende Beitrag nimmt sich dieses Problems an und dient einem Methodenvergleich zum Umgang mit Item Nonresponse. Dabei geht es in erster Linie darum, anhand eines konkreten Beispiels aus der Sozialwissenschaft zu zeigen, dass es z.B. in Form einer Multiplen Imputation (MI) eine relativ einfach anwendbare Alternative zur Complete-Cases-Analyse (CC) gibt.² Für den Methodenvergleich werden zunächst Analysemodelle auf Grundlage von multipel imputierten Datensätzen spezifiziert. Um die Sensitivität der berechneten Regressionsschätzer

- 1 Spricht man im Falle teilweiser Beantwortung von Item Nonresponse, wird beim sogenannten Unit Nonresponse hingegen kein einziges der interessierenden Merkmale erhoben (zur Übersicht vgl. Groves/Fowler/Couper/Leplowski/Singer/Tourangeau 2004).
- 2 Obwohl in der Literatur hinlänglich die Vorteile von MI gegenüber CC diskutiert worden sind, findet sich beispielsweise unter den offiziell registrierten Veröffentlichungen des ALLBUS nahezu keine andere Vorgehensweise als CC bei Item Nonresponse (vgl. Messingschlagler 2012). Ein Grund hierfür könnte sein, dass entsprechende Veröffentlichungen aus der Statistik evtl. nur unzureichend in den Sozialwissenschaften rezipiert werden.

zu testen, wird anschließend eine simulative Grundgesamtheit auf Basis der empirischen Stichprobe erzeugt, aus welcher wiederum 1000 Zufallsstichproben gezogen werden, die im Einzelnen imputiert und analysiert werden. Die gleiche Vorgehensweise wird ebenso für CC angewandt, also ohne eine Ergänzung der fehlenden Werte.³ Die vorgenommenen Vergleiche weisen auf eine weitaus höhere Stabilität (in Form einer geringeren Varianz der Schätzparameter und deren p-Werte) einer MI gegenüber der Verwendung von CC hin. Um dieses Ergebnis nochmals zu evaluieren, werden daraufhin informationsreichere Imputationsmodelle spezifiziert und zur Überprüfung der Stabilität herangezogen.

Als Datenbasis dient eine Erhebung, die im Vorfeld eines Bürgerentscheids im oberfränkischen Lichtenfels durchgeführt wurde. Hierbei wurde über die Frage abgestimmt, ob sich die Bürger für oder gegen die Einstellung eines Bauleitverfahrens zum Bau eines Fachmarktzentrums (FMZ) im Lichtenfelser Mainau-Gebiet aussprechen.

Der Abstimmung vorausgegangen war ein Beschluss des Stadtrates, der sich am 13. September 2010 mit 21:10 Stimmen für die Einleitung des Bauleitverfahrens entschieden hatte. Daraufhin gründete sich die Bürgerinitiative „Rettet das Herz der Lichtenfelser Innenstadt“, welche den Bau ablehnte und ein entsprechendes Bürgerbegehren initiierte. Der Initiative gelang es, 2.680 Unterschriften zu sammeln, was bei einer benötigten Anzahl von lediglich 1.326 Befürwortern zu einem erfolgreichen Begehren führte. Da sich der Stadtrat jedoch weiterhin für das Projekt aussprach, fand der entsprechende Bürgerentscheid am 20. März 2011 statt. Bei diesem entschieden sich fast 70 Prozent der Bürger (exakt: 5.602) für das Projekt, ca. 30 Prozent dagegen (exakt: 2.496) (vgl. Stadt Lichtenfels 2011).

Auf Grundlage theoretischer Überlegungen wird im Folgenden der Frage nachgegangen, welche Determinanten das Abstimmungsverhalten beim Bürgerentscheid beeinflusst haben könnten. Entsprechende Hypothesen werden im folgenden Abschnitt abgeleitet. Dem theoretischen Analyserahmen soll jedoch nur verhältnismäßig wenig Platz eingeräumt werden; im Fokus stehen die statistischen und methodischen Überlegungen zum systematischen Datenausfall und der Korrektur des Nonresponse – vor allem mit Blick auf die nur sehr geringe Datenmenge von 220 Beobachtungen.

3 Hierbei werden lediglich diejenigen Fälle für die Regressionsanalyse verwendet, welche für alle Merkmale des Modells beobachtete Werte aufweisen.

2 Theoretische Überlegungen

Bürgerbegehren und -entscheide (also durch Bürger initiierte Abstimmungen auf kommunaler Ebene) erfreuen sich seit geraumer Zeit einer immer größer werdenden Beliebtheit (vgl. Weixner 2006: 18; Mehr Demokratie e.V.). Umso erstaunlicher erscheint es da, dass sich die empirische politische Verhaltensforschung diesem Thema bislang nur marginal widmet. Stattdessen wird vornehmlich die Frage diskutiert, wie direktdemokratische Verfahren aus normativer und institutioneller Sichtweise zu werten seien.⁴

Auf Länderebene hingegen nimmt die Anzahl empirischer Arbeiten zu: Im Fokus zahlreicher Analysen stand hierbei u.a. die Volksabstimmung zum Nicht-raucherschutzgesetz in Bayern (vgl. Schoen 2011; 2012; Schoen/Glantz/Teusch 2011a; 2011b). Hierbei erwies sich eine modifizierte Version des klassischen Michigan-Modells zur Erklärung von Wahlverhalten (vgl. Campbell/Gutin/Miller 1954; Campbell/Gutin/Miller/Stokes 1960) als adäquater Theorierahmen zur Erklärung des Abstimmungsverhaltens: Laut klassischem Modell ließe sich die Wahlentscheidung mit Hilfe eines Kausalitätstrichters erklären, an dessen dünnem Ende letztlich die Wahlentscheidung stehe, die von politischen Einstellungen und Erfahrungen determiniert sei. Als effektvolle Determinanten werden hierbei die affektive Bindung an eine Partei (Parteiidentifikation) sowie die Bewertungen der Spitzenkandidaten (Kandidatenorientierung) und der parteipolitischen Policies (Sachfragenorientierung) spezifiziert, wobei der Parteiidentifikation (PID) ein besonderer Effekt zugebilligt wird: Diese stehe als langfristig ausgebildete Komponente kausal vor den beiden weiteren Einstellungen (zur weiteren Übersicht vgl. etwa Schoen/Weins 2005).

In einer modifizierten Variante des Michigan-Modells wird bei Schoen (2011; 2012) sowie Schoen/Glantz/Teusch (2011a; 2011b) nun der Sachfragenorientierung der vermeintlich stärkste Effekt zugeschrieben: Schließlich stand beim Volksentscheid in Bayern, anders als bei Wahlen, nicht die Entscheidung für oder gegen eine Partei, sondern die Meinung gegenüber einer Sachfrage im Mittelpunkt. Gleich-

4 Zur Frage, ob Bürger das benötigte Wissen zur Abgabe einer fundierten Meinung bei Volksentscheiden aufweisen, vgl. Oberreuter 2002; Decker 2005; Mendelsohn/Cutler 2000. Zum Argument, dass Entscheide selbst zu erhöhtem Wissen bei Bürgern führen, vgl. qualitative Arbeiten von Kirchgässner/Feld/Savioz 1999; Frey/Kirchgässner 1993. Für quantitative Arbeiten hierzu vgl. Smith 2002; Benz/Stutzer 2004; Christin/Hug/Scarini 2002. Zur Meinungsänderung durch Kampagneneffekte vgl. de Vreese/Semetko 2004; Joslyn/Haider-Merkel 2000; Kriesi 2002. Zum Effekt auf die Meinungsbildung, der durch Konsens auf der Elitenebene auftritt, vgl. Clarke/Kornberg 1996, Clarke/Kornberg/Stewart 2004; de Vreese 2006; Kriesi 2006; Selb/Kriesi/Hänggli/Marr 2009.

ches trifft ebenso auf das hier vorliegende Bürgerbegehren und den -entscheid zu, weshalb das modifizierte Michigan-Modell auch in dieser Arbeit das theoretische Fundament bildet.

Aus diesem Grund kann angenommen werden, dass sich referendumsspezifische Merkmale auch im Falle Lichtenfels vergleichsweise stark auf das Abstimmungsverhalten auswirken. Des Weiteren erscheint es aufgrund des vorausgegangenen Bürgerbegehrens plausibel, dass sich Bürger, die für das Bürgerbegehren unterschrieben haben, auch beim Entscheid eher für den Baustopp aussprechen. Wie sich die weiteren Einstellungen bzw. sozioökonomischen Merkmale auf das Abstimmungsverhalten auswirken, wird mit Hilfe empirischer Daten analysiert.

3 Zum verwendeten Datensatz

Als Datengrundlage dient eine CATI-Zufallsstichprobe von insgesamt $n=220$ Personen, die im Zeitraum vom 17. bis 24. Februar 2011, und somit etwa vier Wochen vor dem eigentlichen Bürgerentscheid, stattfand. Als Grundgesamtheit dient die wahlberechtigte Bevölkerung in Lichtenfels.

3.1 Operationalisierung

Da die Determinanten des Abstimmungsverhaltens beim Entscheid im vordergründigen Interesse stehen, entspricht die Abstimmungsentscheidung dem abhängigen Merkmal der späteren Analysemodelle. Hierzu wurden die Befragten zunächst um eine Stellungnahme hinsichtlich ihrer Teilnahmebereitschaft am Entscheid gebeten. Antworteten sie hierauf, dass sie wahrscheinlich bzw. sehr wahrscheinlich teilnehmen werden, wurden sie entsprechend nach ihrer Abstimmungsentscheidung gefragt. Gaben sie hingegen an, nur vielleicht oder wahrscheinlich nicht teilnehmen zu wollen, wurden die Befragten um eine hypothetische Entscheidung gebeten. Sowohl für diejenigen Befragten mit einer mehr oder minder sicheren als auch für solche mit einer eher unwahrscheinlichen Teilnahme wird im Falle einer (hypothetischen) Ablehnung des Baustopps (was ein Ja beim Entscheid bedeutete) beim abhängigen Merkmal eine 1 zugeordnet, anderenfalls eine 0. Weiß nicht- und Nicht-Angaben werden als fehlende Werte definiert; gleiches gilt für Befragte, die mit Sicherheit nicht am Entscheid teilnehmen werden.

Für die Imputations- und Analysemodelle werden auch alle weiteren Variablen⁵ dichotomisiert (siehe Anhang: Operationalisierung),⁶ wobei stets sowohl das Nichtwissen als auch fehlende Angaben als Missings definiert werden.⁷

Nach Zuweisung der fehlenden Werte bleiben nur noch vier Variablen übrig, die vollständig beobachtet werden, was einen vergleichsweise hohen Item Nonresponse bedeutet – ein mit Blick auf die Fallzahl des Analysemodells und den damit verbundenen großen Standardfehlern massives Problem. Ebenso ist die Frage zu stellen, ob dieser Ausfall unsystematisch ist, oder ob sich die jeweils Antwortenden von den Nicht-Antwortenden in den zu untersuchenden Merkmalen systematisch unterscheiden. Wäre dies der Fall, unterlägen die CC-Schätzer einer Verzerrung. Entsprechend widmet sich der folgende Abschnitt einer Deskription des Datenausfalls.

3.2 Deskription des Datenausfalls und Multiple Imputation

Tabelle 1 im Anhang beinhaltet die jeweiligen Anteile an fehlenden Werten innerhalb jeder Variable des Datensatzes.

Hiernach unterliegt der Datenausfall einer relativ großen Variation: Beträgt der Ausfall für eine ganze Reihe von Merkmalen nicht einmal 5%, liegt der Nonresponse in immerhin drei Variablen bei über 20% – darunter auch die abhängige Variable *Abstimmungsentscheidung* mit einem Anteil von 25,45%. Würden

- 5 Ein besonderes Augenmerk sei dabei auf die Erhebung der Parteineigung geworfen: Um diese abzubilden, wurde die prospektive Stadtratswahlentscheidung erhoben. Somit muss für die Messung des Effekts der Parteibindung eine Einschränkung gemacht werden: Der „klassische“ Indikator zur Messung der PID (siehe zum genauen Wortlaut etwa Schoen/Weins 2005: 209) wird in diesem Datensatz nicht verwendet, jedoch scheint die Operationalisierung der PID durch eine Verwendung der prospektiven Wahlentscheidung sinnvoll: Diese nämlich misst immerhin annäherungsweise die dauerhafte und affektive Bindung an eine politische Partei.
- 6 Die Dichotomisierung erfolgte vor allem wegen der Anzahl der zu schätzenden Parameter und den damit verbundenen Freiheitsgraden des späteren multivariaten Analysemodells. So weisen die meisten erhobenen Merkmale ein ordinales Skalenniveau auf. Um diese – ohne Informationsverlust – statistisch adäquat in die Regression einfließen zu lassen, wäre es notwendig, bei einer z.B. fünfstufigen Ordinalskala vier Dummyvariablen zu bilden. Aufgrund der geringen Fallzahl von 220 würde dies zu einer unverhältnismäßig hohen Anzahl an Modellparametern, wenigen Freiheitsgraden und großen Konfidenzintervallen führen. Da die dichotomen Variablen die empirische Wirklichkeit homomorph abbilden, scheint die vorgenommene Codierung also durchaus praktikabel – sowohl aus theoretisch inhaltlicher, methodischer als auch statistischer Sichtweise.
- 7 Eine Ausnahme hierbei bilden die Wissensfragen: Gaben die Befragten bei der Frage zum Bau des Casinos, des Media Marktes oder der Verkaufsfläche an, die Antwort nicht zu wissen, wird ihnen eine nicht-korrekte Antwort und damit eine 0 zugeordnet.

alle Variablen in ein multivariates Analysemodell ohne Imputation aufgenommen, ergäbe sich ein kumulierter Datenausfall von über 60%, wodurch nur noch 82 der ursprünglich 220 Fälle verwendet werden könnten.

Wie bereits angedeutet, ergibt sich durch den Datenausfall, neben der geringen Fallzahl, ein weiteres Problem: Es ist zu hinterfragen, ob die CC-Schätzer systematischen Verzerrungen unterliegen. Die Antwort hierauf können deskriptive und analytische Vergleiche liefern. Diese erfolgen beispielhaft anhand der Merkmale *Abstimmungsentscheidung* und *Teilnahmewahrscheinlichkeit* am Bürgerentscheid. Wie Tabelle 1 zu entnehmen ist, beträgt der Ausfall innerhalb der späteren abhängigen Variable *Abstimmungsentscheidung* 25,45%, der im Merkmal *Teilnahmewahrscheinlichkeit* dagegen nur 3,18%.

Ob der Ausfall in der Variable *Abstimmungsentscheidung* zufällig ist, lässt sich nun durch den Vergleich zweier Verteilungen des Merkmals *Teilnahmewahrscheinlichkeit* bestimmen: Die erste setzt sich aus *allen* Befragten zusammen, die eine Angabe zur Teilnahmewahrscheinlichkeit am Bürgerentscheid machten. Von diesen 213 Befragten gaben ca. 84% (exakt: 179 Befragte) an, am Bürgerentscheid teilnehmen zu wollen. Die zweite Verteilung für die Teilnahmewahrscheinlichkeit ergibt sich nun lediglich aus denjenigen Befragten, die auf die Frage zur Abstimmungsentscheidung eine Antwort gaben – und damit keine Missings darstellen. Der entsprechende Teildatensatz umfasst eine Fallzahl von nur noch 164 Befragten, wobei von diesen nun fast 93% angaben, an der Abstimmung teilnehmen zu wollen (siehe hierzu die Verteilungen in Abbildung 1).

Somit scheinen diejenigen Befragten, die keine Angabe zur Abstimmungsentscheidung machen, tendenziell seltener am Entscheid teilnehmen zu wollen. Sobald nämlich diese Stichprobeneinheiten aus der Verteilung zur Teilnahmewahrscheinlichkeit ausgeschlossen werden, erhöht sich der Anteil derjenigen, die beim Entscheid eine Stimme abgeben wollen. Diese Erkenntnis erscheint auch theoretisch plausibel: Befragte, die in der Erhebung (und damit vier Wochen vor dem eigentlichen Bürgerentscheid) dazu in der Lage sind, eine Entscheidung bezüglich der Abstimmung zu treffen, könnten am Bürgerentscheid tendenziell eher teilnehmen wollen. Gleichermaßen könnten – im Umkehrschluss – Personen, die sich eher gegen eine Teilnahme am Bürgerentscheid aussprechen, weniger dazu bereit sein, eine Abstimmungsentscheidung abzugeben.

Dieser Verdacht des systematischen Datenausfalls in der Variable zur Abstimmungsentscheidung wird durch einen Chi-Quadrat-Test bestätigt: Hierbei kann die Nullhypothese, wonach die ursprüngliche Verteilung der Teilnahmewahrscheinlichkeit und diejenige, welche sich lediglich aus dem Teildatensatz ergibt, aus einer identischen Grundgesamtheit stammen, bei einer Irrtumswahrscheinlichkeit

von 1% abgelehnt werden, womit ein rein zufälliger Datenausfall auszuschließen ist.

Auch für alle weiteren Merkmale kann mindestens eine Variable beobachtet werden, durch deren Verteilungsvergleich ein rein zufälliger Ausfallmechanismus ausgeschlossen werden kann.

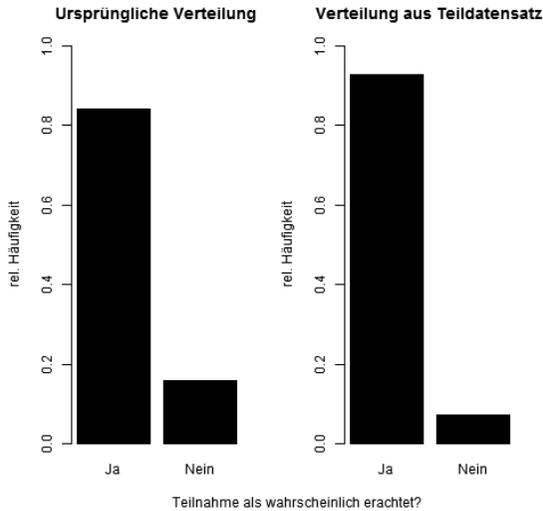
Aus diesem Grund wurden die fehlenden Werte des vorliegenden Datensatzes mit Hilfe einer MI⁸ ergänzt.⁹ Dabei entspricht die Spezifikation des Imputationsmodells derjenigen des Analysemodells (Tabelle 2¹⁰ im Anhang).¹¹

Hiernach scheint sich die Annahme, wonach besonders referendumsspezifische Merkmale das Abstimmungsverhalten determinieren, zu bestätigen. Die Bewertung des vorhandenen Angebots sowie des Standortes als auch die Frage nach der Gefährdung von Arbeitsplätzen können die Entscheidung für oder gegen den Bau des FMZ signifikant erklären. Auch das Leisten einer Unterschrift wirkt sich auf das Abstimmungsverhalten aus, genau wie das referendumsspezifische Wissen und ein hohes Alter.¹²

Um die Schätzung der MI besser evaluieren zu können, werden im Folgenden die Vor- bzw. Nachteile von MI- und CC-Schätzungen miteinander verglichen.

- 8 Eine exaktere Darstellung der Verfahrensannahmen erfolgt im nächsten Abschnitt, um die Vorteile gegenüber der Verwendung von CC-Schätzungen besser verdeutlichen zu können.
- 9 Hierzu wird für die Imputation des vorliegenden Datensatzes das *R-Paket Multivariate Imputation by Chained Equations* (kurz: *mice*) (vgl. van Buuren/Groothuis-Oudshoorn 2011) verwendet. Zur Ergänzung der fehlenden Werte wird ein Gibbs-Sampler (vgl. German/Geman 1984) mit 20 Iterationen verwendet, um insgesamt $m=10$ komplettierte Datensätze zu erzeugen.
- 10 Es sei darauf hingewiesen, dass allen Befragten, die hinsichtlich ihrer Teilnahmewahrscheinlichkeit angaben, mit Sicherheit nicht an der Abstimmung teilnehmen zu wollen (im erhobenen Datensatz gaben dies 16 Personen an), nach der Imputation erneut ein fehlender Wert für die Abstimmungsentscheidung zugewiesen wird, da diese Befragten explizit keine (hypothetische) Stimme abgeben. Dies trifft ebenso auf diejenigen Fälle zu, für welche innerhalb der Imputation in mindestens einem der 10 ergänzten Datensätze eine sichere Nichtteilnahme für die Teilnahmewahrscheinlichkeit imputiert wird. Im Falle der imputierten Datensätze, die als Datenbasis für Tabelle 2 dienen, ist dies für 21 Personen zutreffend, weshalb diese Merkmalsträger, trotz Imputation, für die Analyse nicht zur Verfügung stehen.
- 11 Tabelle 2 beinhaltet somit Regressionsparameter, die unter Verwendung des ursprünglichen und multipel imputierten Datensatzes geschätzt wurden. Diese Datenbasis wurde also nicht aus der später simulierten synthetischen Grundgesamtheit gezogen.
- 12 Da die statistischen und methodischen Verfahren im Vordergrund dieser Arbeit stehen, wird auf eine weitere politikwissenschaftliche Interpretation der Koeffizienten verzichtet.

Abbildung 1 Vergleich der ursprünglichen Verteilung des Merkmals *Subjektive Wichtigkeit* mit Verteilung aus Teilgesamtheit, die sich aus Antwortenden des Merkmals *Abstimmungsentscheidung* ergibt



4 Umgang mit fehlenden Werten

Im Allgemeinen läuft der Umgang mit fehlenden Werten auf zwei Strategien hinaus: Zum einen kann bei der Erhebung bereits von Anfang an auf eine Vermeidung fehlender Werte gezielt werden, zum anderen lassen sich diese aber auch nach ihrer Entstehung korrigieren (Post-Survey Adjustments); hierbei wird versucht, die durch die Missings verursachte Verzerrung zu reduzieren (vgl. de Leeuw/Hox/Dillman 2008: 10). Im vorliegenden Fall bleibt nur die letztgenannte Möglichkeit. Für die Frage, welches Verfahren im vorliegenden Fall bezüglich seiner Stabilität der Parameterschätzungen und p-Werte das zu Empfehlende ist, wird in dieser Arbeit der Blick auf zwei mögliche Vorgehensweisen geworfen, wobei eine von beiden – trotz häufiger praktischer Anwendung – nicht zur Familie der Korrekturmethode zählt und den Ausfallmechanismus damit weitestgehend ignoriert: die Eliminierung fehlender Werte und Verwendung der lediglich auf alle Fragen antwortenden Merkmalsträger. Wie bereits zu Beginn erwähnt, stellt gerade in den Sozialwissenschaften die CC-Methode das in der Regel angewandte Vorgehen dar. Auch wenn

in diesem Beitrag CC-Methoden mit MI verglichen wird, gäbe es natürlich noch viele andere Alternativen über MI hinaus.¹³

4.1 Multiple Imputation

Ohne auf die formale Theorie der MI (vgl. Rubin 1987; 1996) näher einzugehen, muss zumindest einer Annahme zur Modellierung eines adäquaten Imputationsmodells genauer nachgegangen werden: Der Ausfallmechanismus sollte – unter der Annahme, dass die unten folgenden Inferenzen lediglich auf die interessierenden Parameter des Analysemodells und nicht auf die Parameter des Antwortmechanismus abzielen – Completely At Random oder zumindest At Random sein. In diesem Falle würde ein sogenannter ignorierbarer Ausfallmechanismus vorliegen, weshalb dieser im Imputationsmodell nicht explizit modelliert werden müsste und die Modellspezifikation vereinfacht würde. Hierzu eine kurze Erläuterung (vgl. Rubin 1976; Little/Rubin 1987; 2002):

Der Ausfallmechanismus heißt „Missing Completely At Random“ (MCAR), wenn der Datenausfall rein zufällig ist. Ein Beispiel hierfür ist der Verlust von Fragebögen auf dem postalischen Weg vom Erhebungs- zum Analyseinstitut. Dieser Fall liegt jedoch praktisch nie vor, was auch für den vorliegenden Datensatz gilt (siehe obige Ausfalldeskription in Abschnitt 3.2).

Ist der Ausfall innerhalb eines Merkmals von den beobachteten Ausprägungen der anderen im Datensatz vorhandenen Variablen abhängig, nicht aber von der fehlenden Ausprägung des Merkmals selbst, heißt der Mechanismus „Missing At Random“ (MAR). Beispielsweise soll das Gewicht erhoben werden, wobei Frauen ihr Gewicht seltener angeben als Männer. Somit wäre der Ausfall in der Variable Gewicht vom Geschlecht, nicht jedoch zusätzlich vom Gewicht selbst abhängig. Mit anderen Worten wäre der Nonresponse vom Gewicht innerhalb des Merkmals Geschlecht wiederum zufällig.

Sollte der Ausfall einer Variable jedoch von der Ausprägung selbst abhängen, so ist der Ausfallmechanismus „Missing Not At Random“ (MNAR). In diesem Fall verkompliziert sich die Spezifikation des Imputationsmodells erheblich, da der Ausfallmechanismus nicht ignorierbar und damit explizit modelliert werden muss, auch wenn die spätere Inferenz lediglich auf die Parameter des Analysemodells abzielt. Auf das obige Beispiel angewandt hieße der Ausfall MNAR, wenn die Variable Gewicht gerade aufgrund des Gewichts einer Person nicht erhoben würde. Dies

13 Eine Zusammenfassung und Systematisierung von Korrekturmethode findet sich beispielsweise bei Rässler 2000, Göthlich 2007, Spieß 2008.

läge z.B. vor, wenn Befragte aufgrund eines sehr niedrigen oder hohen Gewichts die Angabe verweigerten.¹⁴

Ob die Missings im vorliegenden Datensatz MAR oder gar MNAR sind, lässt sich jedoch durch Tests nicht bestimmen: Zu messen, ob ein Befragter die Antwort auf eine Frage gerade wegen der von ihm generierten Antwort verweigert hat, würde voraussetzen, dass die nicht gegebene Antwort beobachtbar wäre.

Deswegen ist sowohl mit Blick auf die *Weiß nicht*-Kategorie als auch auf die Antwortverweigerung zu hinterfragen, welcher kognitive Prozess den Item Nonresponse verursacht haben könnte. Hierzu sei auf das Entscheidungsmodell für Item Nonresponse nach Beatty und Herrmann (2002) verwiesen. Demnach lässt sich die Response-Entscheidung eines Befragten in verschiedene Abschnitte zerlegen: Im ersten Schritt bedarf es einer verständlichen Interpretation der Fragestellung seitens des Befragten. Ist dieser nicht dazu in der Lage, den Sinn des Items zu verstehen, sei es ihm entsprechend auch nicht möglich, eine Angabe zu machen. Hat der Teilnehmer die Fragestellung hingegen verstanden, muss er darüber entscheiden, ob er den Aufwand betreibt, Eindrücke bzw. Wissen zum abgefragten Thema zu sammeln und zu verarbeiten. Entscheidet sich der Befragte dagegen, hängt der Nonresponse offensichtlich mit nicht ausreichend vorhandenen kognitiven Fähigkeiten oder einer fehlenden Motivation zusammen. Grund für den Nonresponse wäre somit nicht die generierte Antwort selbst, was auf einen MAR-Mechanismus hindeuten ließe. MNAR läge hingegen vor, wenn der Befragte zwar eine Antwort generieren wollte und könnte, diese aber aufgrund ihres Inhalts geheim hielte – beispielsweise aus sozialer Erwünschtheit. In diesem Fall wäre die Antwort selbst der Grund des Nonresponse.¹⁵

Auch weil die Messung nach einem MAR- oder MNAR-Mechanismus nicht möglich ist, wird für diese Arbeit die Annahme getroffen, dass die fehlenden Angaben bzw. *Weiß nicht*-Antworten einem MAR-Mechanismus unterliegen.

- 14 Diesem Ausfallmechanismus folgen, wie bereits oben dargestellt, auch die Missings in der Variable *Abstimmungsentscheidung*, die aufgrund der sicheren Nicht-Teilnahme am Bürgerentscheid definiert wurden. Der Ausfall in *Abstimmungsentscheidung* ist zwar von der Ausprägung des Merkmals *Abstimmungswahrscheinlichkeit* abhängig, jedoch innerhalb dieser nicht zufällig: Wenn ein Befragter *sicher nicht teilnehmen* wird, wird er mit Sicherheit auch keine Abstimmungsentscheidung treffen. Somit ist dieser Ausfall weder MCAR noch MAR.
- 15 Zu unterscheiden ist diese Form der Antwortverweigerung von der Geheimhaltung aufgrund fehlender Präzision: Hiermit gemeint ist der Eindruck des Befragten, dass die von ihm generierte Antwort qualitativ unzureichend begründet sei, weshalb er lieber keine Antwort abgibt. Auch in diesem Fall wäre der Antwortmechanismus MAR, da der Nonresponse nicht mit der Antwort selbst, sondern der unzureichenden Sicherheit des Befragten bei der Antwortvergabe zusammenhinge.

Dafür spricht erstens, dass der verwendete Fragebogen keinerlei Items bzw. Antwortkategorien enthält, die aufgrund ihrer Sensibilität – d.h. mit Blick auf soziale Erwünschtheit – einen vordergründigen MNAR-Ausfall vermuten ließen. Hinzu kommt, dass ein Großteil der kognitiven Prozesse auf dem Weg zur Item Non-response-Entscheidung auf einen MAR-Ausfall schließen lässt. Aufgrund dieser Annahmen kann die Modellierung des Imputationsmodells erheblich vereinfacht und durchgeführt werden.

4.2 Complete Cases

Das zweite Vorgehen stellt – wie schon erwähnt – keine Korrektur dar. Complete Cases eliminiert vielmehr alle Merkmalsträger, die auch nur hinsichtlich eines der ins Analysemodell aufgenommenen Merkmale fehlende Werte aufweisen. Dies führt zu einer Reduktion der für die Analyse verwendeten Fälle. Je nach Ausfallmuster kann selbst ein geringer Ausfall einzelner Variablen in einem multivariaten Modell zu einem hohen Gesamtausfall führen. Besonders in kleinen Stichproben wiegt jeder zusätzlich fehlende Merkmalsträger schwer, so dass sich schon aus ökonomischen Gründen eine CC-Analyse verbieten sollte. Die einzigen Ausfallmechanismen, unter denen CC unverzerrt schätzen kann, sind MCAR und ein besonderer Fall von MAR.¹⁶ Wie in Kapitel 3.2. gezeigt wurde, liegt innerhalb des verwendeten Datensatzes jedoch kein rein zufälliger Ausfall vor, wodurch die CC-Analyse in den nun folgenden Methodenvergleich mit zwei Vorbelastungen geht: einer unerfüllten Annahme zum Ausfallmechanismus und einer verfahrensimmanenten Reduktion der ohnehin kleinen Stichprobe.

5 Methodenvergleich und Simulation

Für einen validen Methodenvergleich ist es unabdingbar, eine Basis zu finden, die für beide miteinander verglichenen Vorgehensweisen zum Umgang mit fehlenden Werten die gleiche Ausgangslage bietet. Doch ein Methodenvergleich sollte noch mehr können: Veröffentlichungen zum Vergleich von Korrekturmetho-

16 Wenn lediglich die abhängige Variable eines Analysemodells fehlende Werte aufweist und die Wahrscheinlichkeit des Fehlens nur von den unabhängigen Variablen des Modells abhängt, dann führt die CC zu unverzerrten Schätzern im interessierenden Modell. Dass ein solcher MAR-Ausfall im hier verwendeten Datensatz vorliegt, erscheint jedoch aufgrund der ebenso in den erklärenden Variablen des Analysemodells auftretenden Missings als äußerst unplausibel.

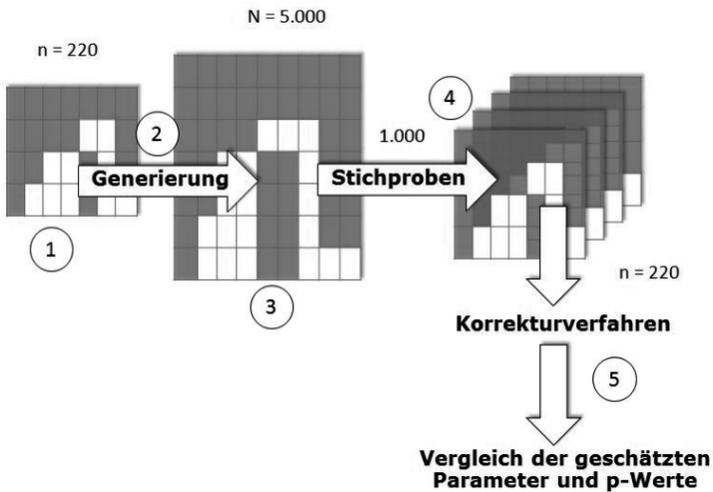
beinhalten entweder die bloße Durchführung der Methode anhand eines ausgewählten empirischen Datensatzes (vgl. Longford 2000) oder eine modellbasierte Simulationsstudie (vgl. Rässler 2000). Beide Optionen bringen jedoch Nachteile mit sich: Die einfache Durchführung mit einem realen Datensatz berücksichtigt zwar die Herausforderungen realer Datengegebenheiten, sie hat allerdings eine eingeschränkte Aussagekraft über den ausgewählten Datensatz hinaus. Simulationen hingegen müssen immer mit dem Makel des Künstlichen und weniger Komplexen leben, was eine Verallgemeinerung der Ergebnisse ebenfalls erschwert.

Für den folgenden Methodenvergleich wird ein vereinfachter Ansatz nach Messingschlager (2012) herangezogen (siehe hierzu Abbildung 2).¹⁷ Der erste Schritt zeigt den Datensatz in vereinfachter Form. Zu sehen sind gegebene und fehlende Daten (1). Der originale Datensatz wird nun via Bootstrapping vergrößert, indem Stichproben aus ihm gezogen und immer wieder an den ursprünglichen Datensatz angehängt werden (2). Damit erhält man Ziehung für Ziehung einen immer größeren Datensatz. Dies wird so lange wiederholt, bis eine Gesamtheit entsteht, die hinreichend groß ist, um Stichproben gegebener Größe ziehen zu können (Auswahlsatz deutlich unter 0,05). Konsequenterweise behält diese Gesamtheit annähernd die Zusammenhangsstruktur und das Ausfallmuster des ursprünglichen Datensatzes zur Befragung der Lichtenfelser Bürger vor dem Bürgerentscheid (3). Die neue Gesamtheit bildet nun eine valide Basis für den Methodenvergleich.¹⁸ Zu diesem Zweck werden für den simulativen Teil 1.000 Stichproben aus der neuen Gesamtheit gezogen. Die Stichprobengröße entspricht dabei jeweils genau derjenigen des Originaldatensatzes, was beispielsweise bei Bootstrapping angesichts der ohnehin sehr kleinen Stichprobe so nicht durchführbar wäre (4). Mit jeder Stichprobe wird das Modell zur Erklärung des Abstimmungsverhaltens beim Bürgerentscheid geschätzt, wobei CC als eliminierendes Vorgehen zum Einsatz kommt und

17 In der hier verwendeten Variante werden lediglich die Verteilungen der durch die Simulation geschätzten Parameterwerte anhand arithmetischer Mittel und Standardabweichungen miteinander verglichen. Bei Messingschlager (2012) wird hingegen ein Stresstest auf der Grundlage konstruierter wahrer Parameter durchgeführt.

18 Durch dieses Vorgehen wird sichergestellt, eine synthetische Grundgesamtheit zu generieren, deren Korrelationsstruktur derjenigen der empirisch erhobenen Stichprobe entspricht (Stichprobe von $n=220$ aus einer Grundgesamtheit von Wahlberechtigten, dann 1.000 Stichproben vom Umfang $n=220$ aus einer synthetischen Gesamtheit). Es geht an dieser Stelle nicht nur um die Schätzung von Parametern wie es z.B. bei Bootstrapping der Fall wäre, sondern darum, eine einheitliche Ausgangslage für den Methodenvergleich – und damit für die unten folgende Sensitivitätsanalyse – zu schaffen. Mit dieser synthetischen Grundgesamtheit ließen sich auch beliebig andere Analysen erstellen. Aufgrund der Übersichtlichkeit bleibt das Augenmerk jedoch auf dem logistischen Regressionsmodell.

Abbildung 2 Simulationsbeschreibung für den Methodenvergleich



MI als Imputationsverfahren dient (5).¹⁹ Die MI-Schätzung wurde dabei zunächst mit einem Modell unter Kongruenz von Imputations- und Analysemodell (1. MI-Simulation) durchgeführt. Anschließend wurde ein zweites Imputationsmodell spezifiziert (2. MI-Simulation). Dieses beinhaltet zusätzliche Merkmale²⁰ und stellt damit eine größere Informationsquelle für die Ergänzung der fehlenden Werte dar. Hierdurch sollte sich die Stabilität der späteren Parameterschätzungen weiter vergrößern (siehe zur Diskussion über *congeniality* Meng 1994 und Rubin 1996).

Der Vorteil dieses Vorgehens liegt zusammengefasst also darin, dass reale Daten den Ausgangspunkt für den Methodenvergleich bilden, indem ein simulativer Datensatz erzeugt wird, der nahezu gleiche Zusammenhangsstrukturen und identische Ausfallmuster besitzt.²¹ Somit wird eine valide Ausgangsbasis für den Methodenvergleich geschaffen.

- 19 Aus der Modellschätzung fallen allerdings auch hierbei diejenigen Merkmalsträger heraus, die geäußert haben, mit Sicherheit nicht am Bürgerentscheid teilnehmen zu wollen.
- 20 Zusätzliche Merkmale: Stellungnahme, ob FMZ eine Bereicherung darstellt; subjektive Wichtigkeit des Bürgerentscheids; Sympathieskalometer gegenüber Bürgermeisterin Bianca Fischer; Wohngegend in Lichtenfels.
- 21 Anders als bei einer Simulation kann der datengenerierende Mechanismus hier jedoch nicht mehr gesteuert werden. Es könnte sich also bei der Datengrundlage auch um einen Extremfall handeln, was jedoch aufgrund der vorgenommenen Datenerhebung (s.o.) unwahrscheinlich ist.

Die Ergebnisse der Schätzungen durch MI und CC werden im nächsten Kapitel grafisch an ausgewählten Beispielvariablen präsentiert.

6 Empirische Ergebnisse

Für alle exogenen Variablen wurden die jeweiligen 1.000 B-Parameter gemittelt (\bar{B}), deren Standardabweichungen (SD) ermittelt sowie die durchschnittlichen p-Werte (\bar{p}) mitsamt deren Standardabweichungen berechnet, welche den Tabellen 3 und 4 im Anhang zu entnehmen sind. Für drei ausgewählte Merkmale sollen die Ergebnisse exemplarisch erläutert werden:

- *Unterschrift bei Begehren* (1,36 % fehlende Werte), welches einen positiven Effekt auf die Abstimmungsentscheidung im obigen Analysemodell ausübte,
- *Kein Standort besser geeignet* (13,18 % fehlende Werte), welches sich negativ auf den Stopp des Baues auswirkte,
- *Prospektive Wahl der Freien Wähler* (26,81 % fehlende Werte), welche sich nur insignifikant auf den Stimmentscheid auswirkte.

Die Histogramme (Abbildungen 3–5) zeigen die Verteilungen der simulierten 1.000 B-Schätzer sowie deren p-Werte getrennt nach den CC- und den zwei MI-Modellen.

Betrachtet man für die Variable *Unterschrift bei Begehren* die Schätzwerteverteilung der Methoden, fällt zunächst die Breite der Verteilung von CC auf. Die Schätzwerte streuen mit einer Standardabweichung von 2,3649 gegenüber lediglich 0,7360 bei den Simulationen durch das erste MI-Modell. Bei der zweiten MI-Simulation reduziert sich die Standardabweichung der geschätzten Werte noch einmal leicht auf 0,7024. Mit Blick auf die Anzahl der durchgeführten Modellschätzungen wird zudem deutlich, dass nach der Eliminierung der fehlenden Werte 370 Stichproben gar keine Modellschätzungen mehr zuließen (CC).²² Die Interpretation der Schätzwerte wäre zudem sehr ungenau, da zwar ein Großteil der Werte einen positiven Zusammenhang zwischen der Unterschrift und dem Entscheid, jedoch ein nicht unbedeutender Teil der Schätzwerte einen negativen Effekt suggeriert. Bei den MI-Simulationen sind hingegen nur sehr wenige Ausreißer zu verzeichnen. Hieran zeigt sich die politikwissenschaftlich-interpretatorische Relevanz der vorgenommenen Simulation: Die berechneten Effekte der CC-Analysen erweisen sich

22 Schätzparameter werden dann als Extremwerte hinsichtlich ihrer Plausibilität ausgeschlossen, wenn sie Werte kleiner -10 oder größer +10 annehmen. Von den 1.000 simulierten Schätzparametern fallen also 370 CC-Parameter als Extremwerte aus dem Intervall -10 bis +10 heraus.

demnach als äußerst schwankend, wodurch sich keine stabilen Parameterschätzungen erzielen lassen.

Dies spiegelt sich auch in den Verteilungen für die p-Werte wider: Im Durchschnitt erreichen die beiden MI-Simulationen einen p-Wert von 0,0168 bzw. von 0,0135. Die Standardabweichungen der MI-Simulationen signalisieren große Stabilität im Vergleich zu den p-Werten, die aus den CC-Modellen hervorgehen. Unter Verwendung letzterer ist bei möglichen Aussagen zum Einfluss der Variable *Unterschrift* auf die Abstimmungsentscheidung große Vorsicht geboten, da sich die simulierten p-Werte als äußerst instabil erweisen.

Für die nächste Variable *Kein Standort besser geeignet* zeigen sich ganz ähnliche Ergebnisse: Die Standardabweichung der geschätzten Parameter liegt bei CC um ein Vielfaches höher als bei den MI-Simulationen – nämlich 2,1910 gegenüber 0,8664 bzw. 0,7037. Der Anteil der Stichproben unter CC, mit denen eine Modellschätzung überhaupt gelingt (konkret: 642), ist geringfügig höher als bei der ersten Variable *Unterschrift*. Die Abstände zwischen den beiden MI-Simulationen bleiben in etwa gleich, d.h. auch hier weist die zweite MI-Simulation eine minimal schmalere Schätzwerteverteilung auf.

Bei den p-Werten schwächt sich die Stabilität ab: Der durchschnittliche p-Wert für die geschätzten Parameterwerte liegt noch bei 0,1438 bzw. 0,1943. Im Durchschnitt sind unter Verwendung beider MI-Modelle demnach keine signifikanten Parameterschätzer für die Variable zur Standortbewertung zu beobachten.

Das letzte Beispiel anhand der Variable *Prospektive Wahl der Freien Wähler* zeigt die Grenzen der Schätzung bei derartigen Datenausfällen im multivariaten Raum durch MI auf: Die Streuung der geschätzten Parameterwerte durch MI (Standardabweichungen von 1,2204 bzw. 1,0260) ist zwar immer noch geringer als die der geschätzten Werte mittels CC (2,9513), jedoch nimmt die Breite der Verteilung derart zu, dass nicht mehr von einem hinreichend befriedigenden Ergebnis gesprochen werden kann.

Für alle Verfahren werden durchschnittlich keine signifikanten Schätzparameter angegeben, wobei wiederum zu berücksichtigen ist, dass bei CC nun gerade kaum mehr als die Hälfte der gezogenen Stichproben (konkret: 587) zu einer Modellschätzung führt, die realistische Schätzwerte annimmt.

Wie die Tabellen 3 und 4 weiter aufzeigen, ergeben sich unter den CC-Spezifikationen über alle Parameter hinweg größere Standardabweichungen als bei den MI-Modellen, was darauf hindeutet, dass die Schätzer einer hohen Ungenauigkeit unterliegen. Dieser Sachverhalt erhärtet sich mit Blick auf die insgesamt deutlich höhere Streuung der p-Werte. Die MI-Simulationen erweisen sich dem gegenüber für jedes Merkmal als die stabilere Schätzung, wobei die Aufnahme weiterer Varia-

Abbildung 3 B- und p-Werte der 1.000 CC-, MI 1- und MI 2-Simulationen für das Merkmal *Unterschrift* bei *Begehren*

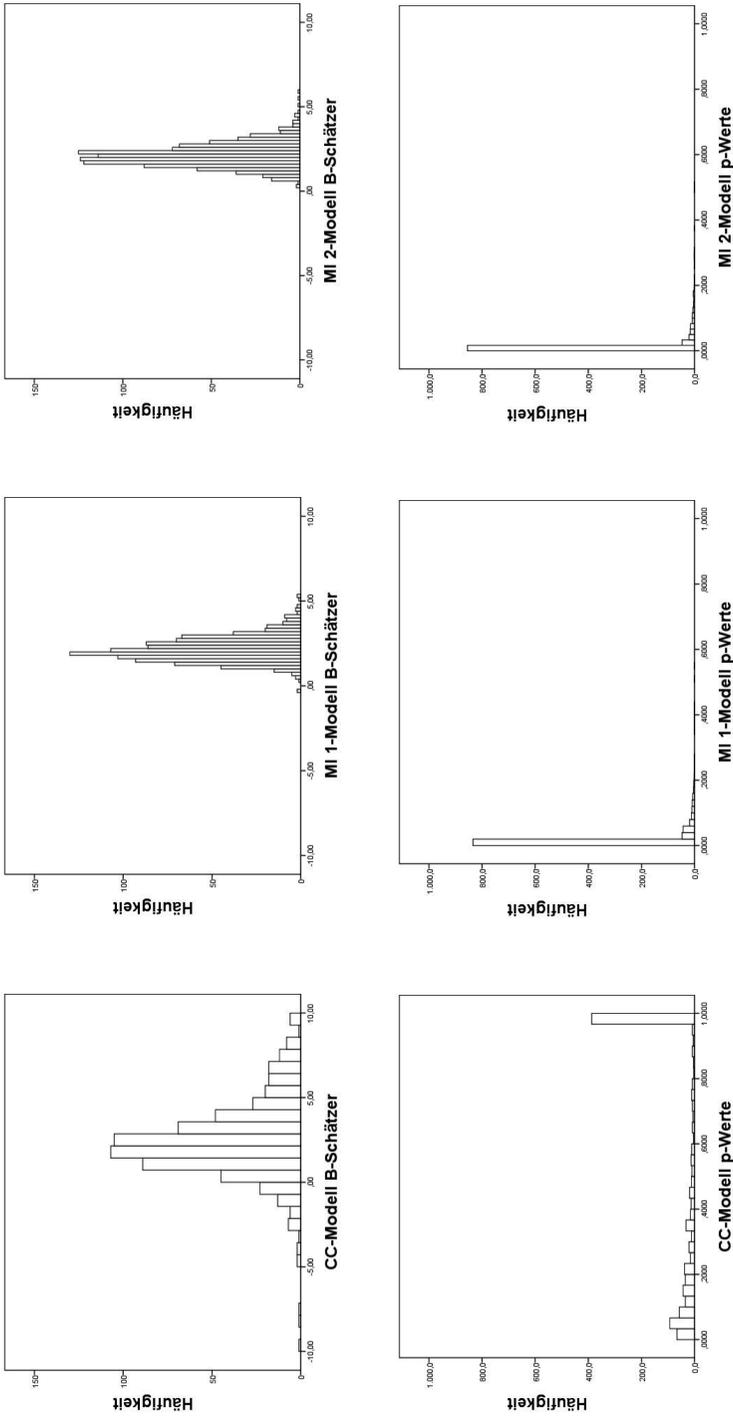


Abbildung 4 B- und p-Werte der 1.000 CC-, MI 1- und MI 2-Simulationen für das Merkmal *Kein Standort besser geeignet*

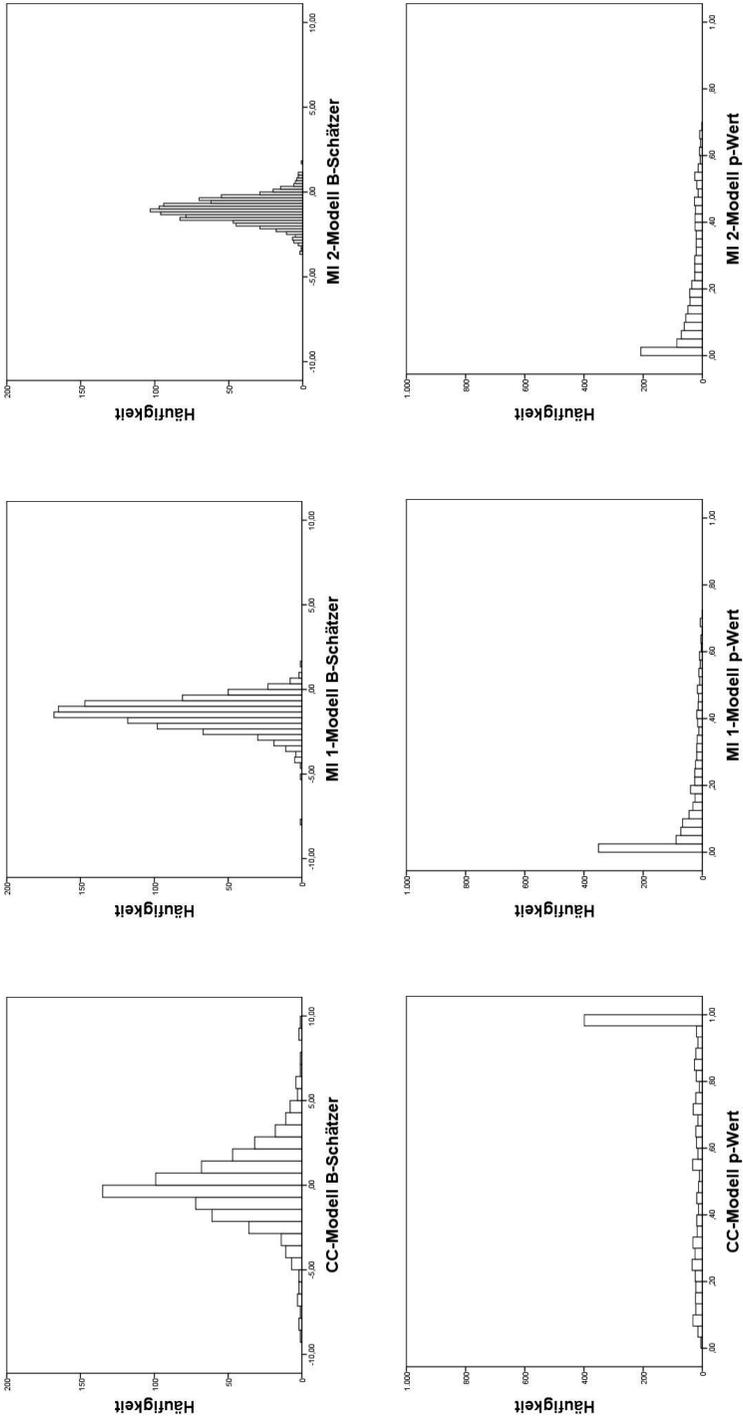
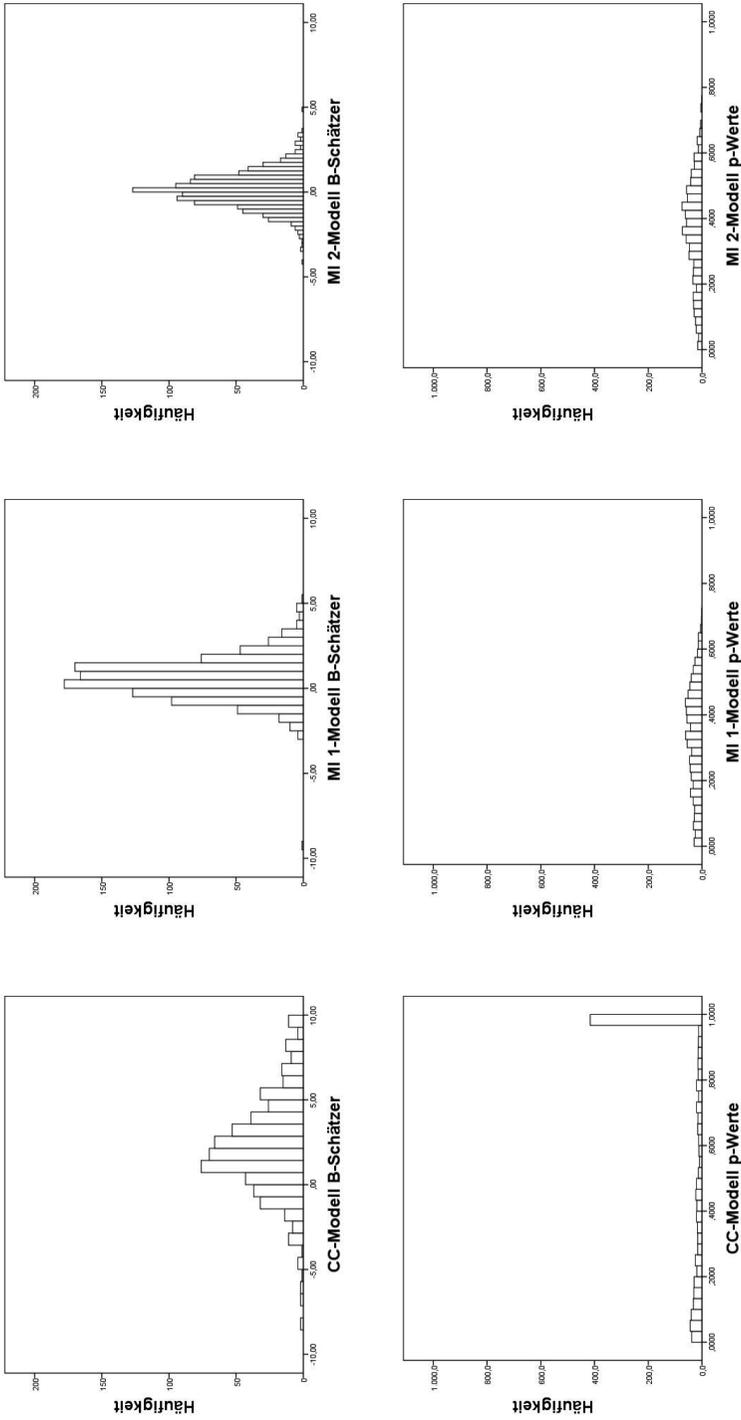


Abbildung 5 B- und p-Werte der 1.000 CC-, MI 1- und MI 2-Simulationen für das Merkmal *Freie Wähler-Wahl* prospektiv



blen im zweiten Modell in nahezu allen Fällen zu genaueren Schätzungen der Parameter führt. Ebenso sinken nach der multiplen Ergänzung der fehlenden Werte die Standardabweichungen der p-Werte, was ebenfalls auf eine stabilere Schätzung hinweist.

7 Fazit und Empfehlungen

Somit lässt sich abschließend konstatieren, dass sich aus politikwissenschaftlicher Sicht besonders die referendumsspezifischen Merkmale auf die Abstimmungsentscheidung ausgewirkt haben. Demnach scheint die Stimmabgabe maßgeblich von der Haltung gegenüber dem in Lichtenfels bereits bestehenden Angebot, dem vorgesehenen Standort des FMZ sowie der vermuteten Arbeitsplatzgefährdung abzuhängen. Ebenso erwiesen sich die Teilnahme am vorangegangenen Bürgerbegehren, das referendumsspezifische Wissen und ein hohes Alter als signifikante Determinanten.

Die auf Grundlage des multipel imputierten Datensatzes berechneten Regressionsschätzer konnten mit Hilfe des Methodenvergleichs auf ihre Sensitivität hin untersucht werden. Die Ergebnisse deuten darauf hin, dass es sich für die Schätzung eines Modells zur Erklärung der Abstimmungsentscheidungen beim Bürgerentscheid in Lichtenfels angesichts der kleinen Stichprobe sehr empfohlen hat, eine MI gegenüber einer CC-Analyse vorzuziehen. Die Simulation zeigt, dass teilweise knapp die Hälfte aller Stichproben unter CC zu keinen annehmbaren oder adäquat interpretierbaren Schätzwerten führt. Dies erscheint umso wichtiger, da der Methodenvergleich so angelegt war, dass kein Verfahren durch das angewandte Vorgehen benachteiligt wird. Die Hinzunahme von Variablen über das Analysemodell hinaus verbessert durchgängig – wenn auch teilweise nur im geringen Maße – die Schätzungen unter MI weiter.

Die Ergebnisse des Methodenvergleichs ermutigen zum Nachdenken über das Problem fehlender Werte und den Umgang damit. Dass die bisher in den Sozialwissenschaften dominierende CC-Analyse im Sensitivitätsvergleich schlechter abschneidet als MI zeigt, dass es sich lohnt, bei der Datenanalyse auch im Umgang mit fehlenden Werten mehr Achtsamkeit an den Tag zu legen. Aufgrund der Implementation von MI in vielen Standardsoftwares bildet die Anwendung richtiger Korrekturmethode keine unüberwindliche Hürde mehr – wie hier ebenfalls gezeigt wurde.

Literatur

- Beatty, P. und D. Herrmann, 2002: To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse. S. 71–85 in: R. M. Groves, D. A. Dillman, J. L. Eltinge und J. A. Roderick (Hg.): *Survey Nonresponse*. New York: John Wiley and Sons.
- Benz, M. und A. Stutzer, 2004: Are Voters Better Informed When They have a Larger Say in Politics? Evidence for the European Union and Switzerland. *Public Choice* 119, 31–59.
- Campbell, A., G. Gurin und W. E. Miller, 1954: *The Voter Decides*. Evanston/Illinois: Row, Peterson and Company.
- Campbell, A., P. E. Converse, W. E. Miller und D. E. Stokes, 1960: *The American Voter*. New York: Wiley.
- Christin, T., S. Hug und P. Sciarini, 2002: Interests and Information in Referendum Voting: An Analysis of Swiss Voters. *European Journal of Political Research* 41: 759–776.
- Clarke, H. D. und A. Kornberg, 1996: Choosing Canada? The 1995 Quebec Sovereignty Referendum. *Political Science and Politics* 29: 676–682.
- Clarke, H. D., A. Kornberg und M. C. Stewart, 2004: Referendum Voting as Political Choice: The Case of Quebec. *British Journal of Political Science* 34: 345–355.
- de Leeuw, E. D., J. J. Hox und D. A. Dillman, 2008: The Cornerstones of Survey Research. S. 1–17 in: E. D. de Leeuw, J. J. Hox und D. A. Dillman (Hg.): *International Handbook of Survey Methodology*. New York: Taylor & Francis.
- de Vreese, C. H., 2006: Political Parties in Dire Straits? Consequences of National Referendums for Political Parties. *Party Politics* 12: 581–598.
- de Vreese, C. H. und H. A. Semetko, 2004: News Matters: Influences on the Vote in the Danish 2000 Euro Referendum Campaign. *European Journal of Political Research* 43: 699–722.
- Decker, F., 2005: Die Systemverträglichkeit der direkten Demokratie. Dargestellt an der Diskussion um die Einführung von plebiszitären Elementen in das Grundgesetz. *Zeitschrift für Politikwissenschaft* 15: 1103–1147.
- Frey, B. S. und G. Kirchgässner, 1993: Diskursethik Politische Ökonomie und Volksabstimmungen. *Analyse & Kritik* 15: 129–149.
- Geman, S. und D. Geman, 1984: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–741.
- Göthlich, S. E., 2007: Zum Umgang mit fehlenden Daten mit großzahligen empirischen Erhebungen. S. 119–134 in: S. Albers, D. Klapper, U. Konradt, A. Walter und J. Wolf (Hg.): *Methodik der empirischen Forschung*. Wiesbaden: Gabler Verlag.
- Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepowski, E. Singer und R. Tourangeau, 2004: *Survey Methodology*. New York: John Wiley and Sons.
- Joslyn, M. R. und D. P. Haider-Markel, 2000: Guns in the Ballot Box: Information, Groups, and Opinion in Ballot Initiative Campaigns. *American Politics Research* 28: 355–378.
- Kirchgässner, G., L. P. Feld und M. R. Savioz, 1999: *Die direkte Demokratie: Modern, erfolgreich, entwicklungs- und exportfähig*. Basel/München: Helbing & Lichtenhahn.
- Kriesi, H., 2002: Individual Opinion Formation in a Direct Democratic Campaign. *British Journal of Political Science* 32: 171–185.
- Kriesi, H., 2006: Role of The Political Elite in Swiss Direct-Democratic Votes. *Party Politics* 12: 599–622.
- Little, R. J. A. und D. B. Rubin, 1987: *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

- Little, R. J. A. und D. B. Rubin, 2002: *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.
- Longford, N. T., 2000: Multiple Imputation in an International Database of Social Science Surveys. *ZA-Information* 46: 72–95.
- Mehr Demokratie e.V.: Übersicht pro Bundesland. <http://www.mehr-demokratie.de/5970.html> (31.01.2012)
- Mendelsohn, M. und F. Cutler, 2000: The Effect of Referendums on Democratic Citizens: Information, Politicization, Efficacy and Tolerance. *British Journal of Political Science* 30: 669–698.
- Meng, X. L., 1994: Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science* 9: 538–573.
- Messingschlager, Martin (2012): *Fehlende Werte in den Sozialwissenschaften – Analyse und Korrektur mit Beispielen aus dem ALLBUS*. Bamberg, University of Bamberg Press.
- Oberreuter, H., 2002: Direkte Demokratie und die repräsentative Verfassung der Bundesrepublik Deutschland. *Zeitschrift für Politik* 49: 290–305.
- Rässler, S., 2000: Ergänzung fehlender Daten in Umfragen. *Jahrbücher für Nationalökonomie und Statistik* 220: 64–94.
- Rubin, D. B., 1976: Inference and Missing Data. *Biometrika* 63: 581–592.
- Rubin, D. B., 1987: *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.
- Rubin, D. B., 1996: Multiple Imputation After 18+ Years. *Journal of the American Statistical Association* 91: 473–489.
- Schoen, H., 2011: Eine eindeutige Entscheidung nach resonanzarmer Kampagne. Überlegungen zum Volksentscheid über den Nichtraucherschutz in Bayern. *Zeitschrift für Politikberatung* 3: 457–461.
- Schoen, H., 2012: Wählen und abstimmen – zwei Seiten einer Medaille? Eine Analyse am Beispiel des Volksentscheids zum Nichtraucherschutz in Bayern. *Politische Vierteljahrszeitschrift – Sonderheft* 45: 514–535.
- Schoen, H. und C. Weins, 2005: Der sozialpsychologische Ansatz zur Erklärung von Wahlverhalten. S. 187–242 in: J. W. Falter, H. Schoen (Hg.): *Handbuch Wahlforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schoen, H., A. Glantz und R. Teusch, 2011a: Abstimmungskampf, Informationsvermittlung und Stimmenscheidung beim Volksentscheid über den Nichtraucherschutz in Bayern. S. 295–320 in: L. P. Feld, P. M. Huber, O. Jung, C. Welzel und F. Wittreck (Hg.): *Jahrbuch für direkte Demokratie 2010*. Baden-Baden: Nomos.
- Schoen, H., A. Glantz und R. Teusch, 2011b: Raucher oder Nichtraucher – ist das die Frage? Wahlbeteiligung und Abstimmungsverhalten beim Volksentscheid über das Rauchverbot in Bayern. *Zeitschrift für Parlamentsfragen* 42: 492–502.
- Selb, P., H. Kriesi, R. Hänggli und M. Marr, 2009: Partisan Choices in a Direct-Democratic Campaign. *Political Science Review* 1: 155–172.
- Smith, M. A., 2002: Ballot Initiatives and the Democratic Citizen. *The Journal of Politics* 64: 892–903.
- Spieß, M., 2008: *Missing-Data-Techniken. Analyse von Daten mit fehlenden Werten*. Hamburg: Lit.
- Stadt Lichtenfels, 2011: Klares Ergebnis beim Bürgerentscheid. <http://www.lichtenfels-city.de/index.php?id=7777,589> (20.09.2012)
- van Buuren, S. und K. Groothuis-Oudshoorn, 2011: mice: Multivariate Imputation by Chained Equations in R. *Journal Of Statistical Software* 45: 1–67.
- Weixner, B. M., 2006: Direkte Demokratie in den Bundesländern. *Aus Politik und Zeitgeschichte* 10: 18–24.

Anschriften der Autoren

Martin Messingschlager
Otto-Friedrich-Universität Bamberg
Fakultät Sozial- und Wirtschaftswissenschaften
Lehrstuhl für Statistik und Ökonometrie
96052 Bamberg
E-Mail: martin.messingschlager@uni-bamberg.de

Marcel Preisung
Otto-Friedrich-Universität Bamberg
Fakultät Sozial- und Wirtschaftswissenschaften
Lehrstuhl für Statistik und Ökonometrie
96052 Bamberg
E-Mail: marcel.preising@uni-bamberg.de

Anhang

Operationalisierung:

- *Unterschrift bei Begehren:* Misst dichotom, ob die Befragten beim bereits durchgeführten Bürgerbegehren eine Unterschrift geleistet hatten (1=Ja, 0=Nein).
- *Subjektive Wichtigkeit des Bürgerentscheids:* Ausprägungen wichtig und sehr wichtig zur 1 zusammengefasst, die restlichen drei Ausprägungen zur 0.
- Drei Wissensfragen, auf die die Befragten mit *Ja* oder *Nein* bzw. *Weiß nicht* antworten konnten: 1. ob im FMZ ein Spielcasino und 2. ein *Media Markt* geplant seien. 3. ob sich die *Verkaufsfläche* in Lichtenfels durch das FMZ verdoppele. Die korrekte Antwort lautet jeweils *Ja*. Bei richtiger Antwort wird die Ausprägung 1 zugeordnet, bei falscher oder keiner Antwort eine 0. Hiernach Bildung eines Index *Wissen*, welcher angibt, wie viele der drei Fragen von den Teilnehmern korrekt beantwortet wurden.
- Sieben sachfragenorientierte Merkmale, gemessen via fünfstufiger Likert-Skalen. Ein erhöhter Skalenpunkt bedeutet immer eine positivere Einstellung zum FMZ (bzw. eine minder negative). Aussagen: 1. „Das FMZ wird Arbeitsplätze in der Innenstadt gefährden.“, 2. „Es gibt in Lichtenfels Standorte, die sich besser für das FMZ eignen als die Mainau.“, 3. „Das FMZ ist eine Bereicherung für Lichtenfels.“, 4. „Lichtenfels braucht kein Geld von auswärtigen Investoren.“, 5. „Die Stadtverwaltung und die Bürgermeisterin haben eine klare Vorstellungen darüber, wie es mit dem Einzelhandel in Lichtenfels weitergehen soll.“, 6. „Die Bürgerinitiative „Rettet das Herz der Lichtenfelser Innenstadt“ vertritt einseitig die Interessen der Einzelhändler in Lichtenfels.“, sowie 7. „Ich bin zufrieden mit dem bestehenden Angebot des Einzelhandels in Lichtenfels.“. Es werden jeweils die beiden Kategorien, welche eine eher oder voll positive Haltung zum FMZ widerspiegeln, zur 1 zusammengefasst. Befragten, die eine der restlichen Ausprägungen besetzen, wird eine 0 zugeordnet.²³
- *Parteieneigung:* siehe Fußnote 5.
- *Bianca Fischer (CSU):* Ursprünglich fünfstufige Sympathiebewertung. Die beiden höchsten Merkmalswerte werden zur 1, die restlichen drei Kategorien zur 0 zusammengefasst.

23 Auf die Erzeugung eines Index wurde verzichtet, da die gemessenen Indikatoren keine gemeinsame Dimension abbilden. Zur Messung der Reliabilität wurde Cronbachs α für verschiedene Merkmalskombinationen berechnet. Dieses fiel jedoch in jedem Fall zu gering für eine Indexbildung aus (geringer als 0,5).

- Soziodemographische Merkmale: *Geschlecht* (Mann=1, Frau=0), *Alter* (eingeteilt in drei Klassen <35, ≥35 & ≤65 sowie >65, wobei jüngste Befragte als Referenzkategorie), höchster *Schulabschluss* (eingeteilt in Abitur oder Hochschulabschluss=1, restliche Abschlüsse=0) sowie der *Wohnort* in Lichtenfels, in dem der Befragte lebt. Alle Befragten aus dem Stadtzentrum zur 1, diejenigen aus nördlichen Stadtteilen – und somit relativ nahe am Mainaugebiet Wohnende – zur 2, sowie Personen aus dem Süden Lichtenfels' zur 3 zusammengefasst.

Tabelle 1 Anteil des Datenausfalls in den verwendeten Merkmalen (aufsteigend)

Merkmal	Anteil fehlender Werte (%)
Wissen	0
Geschlecht	0
Angebot nicht ausreichend	0,09
Unterschrift bei Begehren	1,36
Keine Arbeitsplatzgefährdung	1,82
Bürgerentscheid Teilnahmewahrscheinlichkeit	3,18
Alter	3,64
Hohe Bildung	4,09
Einseitige Vertretung durch Bürgerinitiative	12,73
Kein Standort besser geeignet	13,18
Abstimmungsentscheidung	25,45
Wahl prospektiv	26,81
Investorengelder benötigt	28,64
Kumuliert	62,72

Quelle: Datensatz zur Befragung in Lichtenfels.

Tabelle 2 Multivariate Effekte auf die Wahrscheinlichkeit, für den Stopp des FMZ zu votieren (Datensatz mit Hilfe einer Multiplen Imputation ergänzt)

Merkmal	\hat{B}	$\sqrt{s^2}$
Unterschrift bei Begehren	1,65***	(0,59)
Wissen	-0,92***	(0,32)
Angebot nicht ausreichend	-1,21**	(0,57)
Kein Standort besser geeignet	-1,49***	(0,60)
Keine Arbeitsplatzgefährdung	-1,35***	(0,51)
Investorengelder benötigt	-0,51	(0,61)
Einseitige Vertretung durch BI	-0,33	(0,56)
CSU-Wahl prospektiv	-0,18	(0,73)
SPD-Wahl prospektiv	0,24	(0,79)
Grünen-Wahl prospektiv	0,57	(1,21)
Freie Wähler-Wahl prospektiv	0,69	(1,11)
Geschlecht (Mann)	-0,65	(0,53)
Hohe Bildung	0,41	(0,62)
Mittleres Alter	0,19	(0,70)
Hohes Alter	1,93**	(0,88)
Konstante	3,47***	(1,27)
N		199
Nagelkerke R ²		0,58

Signifikanzniveau: *: p=0,10, **: p=0,05, ***: p=0,01.

Quelle: Datensatz zur Befragung in Lichtenfels.

Tabelle 3 Übersicht: Parameter der Verteilungen der geschätzten B-Werte

Merkmal	CC-Modell			MI 1-Modell			MI 2-Modell		
	\bar{B}	SD	N	\bar{B}	SD	N	\bar{B}	SD	N
Unterschrift bei Begehren	2,4825	2,3649	630	2,1513	0,7360	999	2,1149	0,7024	1000
Wissen	-1,3958	1,6992	700	-0,8621	0,3529	1000	-0,7240	0,3310	1000
Angebot nicht ausreichend	-2,2884	2,3433	634	-1,1114	0,6462	1000	-0,9708	0,5866	1000
Kein Standort besser geeignet	-0,0475	2,1910	642	-1,4374	0,8664	1000	-1,0615	0,7037	1000
Keine Arbeitsplatzfähigkeit	-4,1296	2,2835	577	-1,7020	0,8094	1000	-1,5751	0,6917	1000
Investorengelder benötigt	-1,2416	2,5987	617	-0,9855	0,7685	1000	-0,6425	0,6532	1000
Einseitige Vertretung durch BI	-1,6423	2,1564	639	-0,4856	0,7071	1000	-0,5826	0,6504	1000
CSU-Wahl prospektiv	1,6440	3,1126	635	-0,1903	0,9701	1000	-0,6714	0,7691	1000
SPD-Wahl prospektiv	4,7500	2,7539	499	-0,0824	1,0028	1000	-0,2513	0,8393	1000
Grüne-Wahl prospektiv	2,6563	3,2990	372	1,3548	1,9807	998	0,2385	1,3309	1000
Freie Wähler-Wahl prospektiv	2,2501	2,9513	587	0,5625	1,2204	1000	0,1174	1,0260	1000
Geschlecht (Mann)	-1,2727	2,2564	651	-0,5811	0,6953	1000	-0,4297	0,6132	1000
Hohe Bildung	2,5671	2,3720	612	0,5319	0,8119	1000	0,4301	0,7037	999
Mittleres Alter	2,1549	2,9270	594	0,5198	0,9284	1000	0,3105	0,8027	1000
Hohes Alter	4,4181	2,8009	525	2,8320	1,4045	996	2,5235	1,1820	998
Konstante	1,3793	3,7776	538	3,3542	1,7332	998	3,1573	1,5809	999

Angaben sind arithmetische Mittel und Standardfehler der jeweils 1.000 simulierten B-Schätzer. N entspricht der Anzahl an Parametern, die im Intervall von -10 bis +10 liegen. Quelle: Datensatz zur Befragung in Lichtenfels.

Tabelle 4 Übersicht: Parameter der Verteilungen der geschätzten p-Werte

Merkmal	CC-Modell		MI 1-Modell		MI 2-Modell	
	\bar{p}	SD	\bar{p}	SD	\bar{p}	SD
Unterschrift bei Begehren1	0,5475	0,4065	0,0168	0,0485	0,0135	0,0425
Wissen	0,5125	0,4023	0,0857	0,1345	0,1062	0,1460
Angebot nicht ausreichend	0,5135	0,4261	0,1549	0,1820	0,1810	0,2044
Kein Standort besser geeignet	0,6849	0,3325	0,1438	0,1709	0,1943	0,1831
Keine Arbeitsplatzgefährdung	0,4426	0,4655	0,0631	0,1195	0,0611	0,1174
Investorengelder benötigt	0,6237	0,3718	0,2225	0,1900	0,2897	0,1906
Einseitige Vertretung durch BI	0,5611	0,4056	0,3130	0,2118	0,3007	0,2165
CSU-Wahl prospektiv	0,6621	0,3576	0,3190	0,1690	0,2827	0,1830
SPD-Wahl prospektiv	0,5075	0,4442	0,3345	0,1679	0,3540	0,1785
Grüne-Wahl prospektiv	0,7772	0,3408	0,3715	0,1376	0,3706	0,1342
Freie Wähler-Wahl prospektiv	0,6412	0,3729	0,3247	0,1632	0,3583	0,1587
Geschlecht (Mann)	0,5966	0,3772	0,3011	0,2205	0,3428	0,2219
Hohe Bildung	0,5487	0,4031	0,3386	0,2515	0,3650	0,3038
Mittleres Alter	0,5901	0,3968	0,3541	0,2246	0,4109	0,2209
Hohes Alter	0,5159	0,4441	0,0729	0,1332	0,0754	0,1391
Konstante	0,6606	0,3614	0,1086	0,1566	0,1050	0,1473

Angaben sind arithmetische Mittel und Standardfehler der jeweils 1.000 simulierten p-Werte.

Quelle: Datensatz zur Befragung in Lichtenfels.