



Quelle: Careerbuilder, 2021

Who is working from home and when?

Anna-Sophie Martin

Projektarbeit

Studiengang Bauingenieurwissenschaften

Juni 2021

IVT *Institut für Verkehrsplanung und Transportsysteme*
Institute for Transport Planning and Systems

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Inhaltsverzeichnis

1	Einleitung	1
1.1	Einführung	1
1.2	Zielsetzung und Vorgehensweise	1
2	Literaturüberblick.....	3
2.1	Der Begriff Homeoffice	3
2.2	Beeinflussende Faktoren	3
3	Methoden zum Umgang mit Daten	6
3.1	Datenbasis	6
3.2	Imputation fehlender Daten	7
4	Datenanalyse	12
4.1	Imputation	12
4.2	Analyse der imputierten Datensätze	16
5	Vorbereitung des Choice Models	21
5.1	Discrete-Choice-Model	21
5.2	Wahl der Variablen	22
5.3	Auswahl der Beobachtungen.....	28
6	Aufbau des Choice Models.....	30
6.1	Deskriptive Datenanalyse	30
6.2	Logit-Modell	37
6.3	Probit-Modell.....	41
6.4	Mixed Logit-Modell.....	42
6.5	Finales Modell.....	44
6.6	Interpretation der Parameter	46
6.7	Beschränkungen des Modells und weitere Anpassungsmöglichkeiten.....	47

7	Diskussion der Ergebnisse	50
7.1	Anteile Homeoffice 2020.....	50
7.2	Entwicklungen ab November 2020	51
7.3	Prognosen Homeoffice	52
8	Zusammenfassung.....	56
	Danksagung	58
	Literatur.....	59
	Anhang.....	1
A 1	Anhang zur Imputation der Datensätze	1
A 2	Übersicht elektronischer Anhang.....	6
A 3	Eigenständigkeitserklärung	9

Tabellenverzeichnis

Tabelle 1	Ausgewählte Imputationsmethoden der Funktion mice in R.....	15
Tabelle 2	Mittelwerte der Variablen für die unterschiedlichen Datensätze	20
Tabelle 3	Standardabweichungen der Variablen für die unterschiedlichen Datensätze	20
Tabelle 4	Anzahl der Erwerbstätigen mit Möglichkeiten im Homeoffice zu arbeiten in verschiedenen Zeiträumen.....	30
Tabelle 5	Verwendete Variablen für den Aufbau des Entscheidungsmodells	37
Tabelle 6	Aufbau Logit-Modell	38
Tabelle 7	Kategorien für die Variablen «risk» und «stringency»	39
Tabelle 8	Weitere Anpassungen Logit Modell.....	39
Tabelle 9	Schätzungen der Regressionskoeffizienten für das Logit Model Nr.1040	
Tabelle 10	Schätzungen der Regressionskoeffizienten für das Mixed Logit-Modell	43
Tabelle 11	Schätzungen der Regressionskoeffizienten für das finale Modell.....	45
Tabelle 12	Übersicht Variablen für vorbereiteten Datensatz.....	1
Tabelle 13	Übersicht der erstellten Codes für das Projekt.....	7
Tabelle 14	Übersicht erstellte Datensätze	8

Abbildungsverzeichnis

Abbildung 1	Schematische Darstellung der Imputation mit dem R-Paket Amelia unter Verwendung des EMB-Algorithmus.....	9
Abbildung 2	Anteil fehlender Werte im Datensatz vor Imputation	14
Abbildung 3	Vergleich der Verteilungen der imputierten Werte der Variable «workload.x»	17
Abbildung 4	Vergleich der Verteilungen der imputierten Werte der Variable «workload.y»	18
Abbildung 5	Vergleich der Verteilungen der imputierten Werte der Variable «kurzarbeit.x»	18
Abbildung 6	Vergleich der Verteilungen der imputierten Werte der Variable «kurzarbeit.y»	19
Abbildung 7	COVID- 19 Stringency Index für die Schweiz im zeitlichen Verlauf der COVID-19 Pandemie	23
Abbildung 8	Übersicht über die verwendete Datensätze während des Projekts	25
Abbildung 9	Vorgehensweise zur Zuordnung der Tätigkeit zu den aufgezeichneten Tagen	27
Abbildung 10	Anteil der Teilnehmer im Homeoffice über die Zeit.....	31
Abbildung 11	Anteil der Tätigkeiten im Homeoffice über die Zeit und nach Kanton	32
Abbildung 12	Anteil der Tätigkeiten im Homeoffice über die Zeit und Geschlecht	34
Abbildung 13	Verlauf des Risikos, der Restriktionen und der Anteil der Personen im Homeoffice über die Zeit.....	35

Abbildung 14	Zusammenhang zwischen Restriktionen und Anteil im Homeoffice. 36
Abbildung 15	Anteil der Erwerbstätigen im Homeoffice über die Zeit 50
Abbildung 16	Ergebnisse zur Umfrage Arbeit im Homeoffice im Februar 2021 52
Abbildung 17	Ergebnisse zur Umfrage Homeoffice-Wunsch im Februar 2021 53

Projektarbeit

Who is working from home and when?

Anna-Sophie Martin
IVT
ETH Zürich
CH-8093 Zürich

Betreuer:
Caroline Winkler
Prof. Dr. Kay W. Axhausen
Institut für Verkehrsplanung, IVT
ETH Zürich

Juni 2021

Kurzfassung

Schon seit vielen Jahren wird die Arbeit im Homeoffice als Massnahme zur Verkehrsentslastung vorgeschlagen. Mit Beginn der Covid-19 Pandemie wurde Homeoffice in der Schweiz vermehrt auch als Massnahme zur Eindämmung der Pandemie empfohlen. In dieser Arbeit wird mit Hilfe von aufgezeichneten Bewegungsdaten und Angaben der individuellen Arbeitssituation der beobachteten Personen untersucht, inwiefern sich der Anteil der Personen, die im Homeoffice arbeiten, während der Covid-19 Pandemie verändert hat. Die Auswertung der Daten zeigt einen deutlichen Anstieg der Arbeitstage im Homeoffice. Verglichen mit den Anteilen vor 2020 wird auch nach der Covid-19 Pandemie ein höheres Niveau an Homeoffice-Tagen erwartet. Aus diesem Grund wird ein Entscheidungsmodell erstellt, um die individuellen Tendenzen für die Entscheidung zur Arbeit im Homeoffice zu bestimmen. Neben soziodemographischen Merkmalen werden in das Entscheidungsmodell zudem auch spezifische Faktoren zur Berücksichtigung der Covid-19 Pandemie miteinbezogen. Darüber hinaus werden mögliche Auswirkungen der veränderten Anteile an Homeoffice-Tagen präsentiert und diskutiert.

Schlagworte

Homeoffice, Teleheimarbeit, Entscheidungsmodell, Covid-19

Zitierungsvorschlag

Martin, A. (2021) Who is working from home and when? Projektarbeit, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Semester project

Who is working from home and when?

Anna-Sophie Martin
IVT
ETH Zürich

Supervisor:
Caroline Winkler
Prof. Dr. Kay W. Axhausen
IVT
ETH Zürich

June 2021

Abstract

For many years working from home has been proposed as a policy to reduce congestion. With the beginning of the Covid-19 pandemic, home office was introduced as a measure to contain the corona virus. This project analyzes how the levels for working from home changed during the Covid-19 pandemic based on tracking data and further information like the sociodemographic characteristics and details on the individual working situation of the study participants. The analysis shows that the share of individuals working from home increased during the Covid-19 pandemic. Compared to the share of people working from home before 2020 the levels of home office days are expected to be higher after the pandemic. Therefore, a choice model is set up to predict the individuals' tendency to work from home. In addition to sociodemographic characteristics, variables incorporating the specific situation during the Covid-19 pandemic are included in the set-up of the choice model. Finally, potential effects of changed levels of individuals working from home are presented and discussed.

Keywords

Home office, Teleworking, Choice Model, Covid-19

Preferred citation style

Martin, A. (2021) Who is working from home and when? Projektarbeit, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

1 Einleitung

1.1 Einführung

Teleheimarbeit, oder auch Homeoffice genannt, ist eine Tätigkeit, bei der Erwerbstätige von zu Hause aus arbeiten und Daten via Internet mit dem Arbeit- oder Auftraggeber austauschen. Homeoffice ist in der Schweiz schon rund 20 Jahre bekannt. Mit dem Fortschreiten der Digitalisierung hat während dieser Zeit die Teleheimarbeit an Bedeutung gewonnen. Die Zahl der Erwerbstätigen, die zumindest gelegentlich aus dem Homeoffice gearbeitet haben, hat sich zwischen 2001 und 2019 von 250'000 auf 1.1 Millionen Arbeitnehmende und Selbständige erhöht (Bundesamt für Statistik, 2021). Von den knapp 25 Prozent Erwerbstätigen die 2019 mindestens gelegentlich im Homeoffice arbeiteten, war nur jeder zweite regelmässig und nur jeder achte normalerweise im Homeoffice tätig. Damit liegt die Schweiz im internationalen Vergleich unter dem Durchschnitt (Eurostat, 2018).

Im Rahmen der Massnahmen zur Bekämpfung der Covid-19-Pandemie stieg der Anteil der Erwerbstätigen, die im Homeoffice arbeiten in 2020 markant an. Nach Angaben des Statistischen Bundesamtes stieg die Anzahl der Telearbeit leistenden Personen von 1.1 Millionen auf 1.5 Millionen Personen an. Das entspricht rund 34 Prozent der Erwerbstätigen (Bundesamt für Statistik, 2021).

1.2 Zielsetzung und Vorgehensweise

Erst nach dem Ende der Covid-19 Pandemie wird sich herausstellen, inwiefern die Erfahrungen mit der Arbeit aus dem Homeoffice Auswirkungen auf den dauerhaften Anteil der Teleheimarbeit leistenden Personen haben werden.

In diesem Projekt werden anhand von Erhebungen und aufgezeichneten Bewegungsdaten vor und während der Covid-19 Pandemie unterschiedliche Einflüsse auf den Anteil der Teleheimarbeit leistenden Erwerbstätigen analysiert. Anhand der Daten wird untersucht, wie hoch der Anteil der Erwerbstätigen im Homeoffice zu unterschiedlichen Zeitpunkten der Covid-19-Pandemie war. Zudem werden die Effekte unterschiedlicher soziodemographischer Merkmale sowie die Auswirkungen weiterer Einflüsse kontrolliert.

Hierfür werden in Kapitel 2 zunächst die wichtigsten Grundlagen definiert und die soziodemographischen Merkmale, die eine erwartete Auswirkung auf die Wahl des Arbeitsortes haben, eingeführt.

In Kapitel 3 werden die Grundlagen der verwendeten Daten erläutert. Zudem wird auf die Methodik zur Datenaufbereitung eingegangen. Der unvollständige Datensatz wird mit Hilfe unterschiedlicher Verfahren imputiert. Das Vorgehen sowie die Auswertung der Imputationen werden in Kapitel 4 erläutert.

Der Kern dieser Projektarbeit ist die Modellierung eines Entscheidungsmodells. Die Vorbereitung und die benötigten Grundlagen werden in Kapitel 5 vorgestellt. In Kapitel 6 werden unterschiedliche Entscheidungsmodelle aufgebaut und geschätzt. Die Analyse der Modelle wird ebenfalls dokumentiert. Das finale Entscheidungsmodell wird anschliessend präsentiert.

In Kapitel 7 werden die Ergebnisse der vorangegangenen Kapitel diskutiert. Zudem wird ein Ausblick auf die weiteren Entwicklungen gegeben. Überdies werden in Kapitel 7 die Auswirkungen von Homeoffice und der zukünftigen Entwicklung auf Grundlage der Literatur erläutert. Schliesslich wird in Kapitel 8 der Bericht mit einer Zusammenfassung komplettiert.

2 Literaturüberblick

2.1 Der Begriff Homeoffice

Gemäss der Definition des Staatssekretariats für Wirtschaft (SECO) wird unter dem Begriff Homeoffice «vorliegend jene Arbeit verstanden, die Arbeitnehmende ganz oder teilweise, regelmässig oder unregelmässig von zu Hause aus verrichten. Dabei ist der häusliche Arbeitsplatz normalerweise mit dem betrieblichen Arbeitsplatz durch elektronische Kommunikationsmittel verbunden.» (Staatssekretariat für Wirtschaft SECO, 2019)

Neben dem Begriff «Homeoffice» wird oft der Begriff «Telearbeit» synonym verwendet. Ein wesentliches Kriterium der Telearbeit ist die Nutzung von Telekommunikationsmitteln zur Integration des Telearbeitenden in die Organisation des Arbeitgebers. Die Arbeit wird aus der Ferne und somit ausserhalb der Geschäftsräumlichkeiten des Arbeitgebers verrichtet (Domening, 2016). Im Homeoffice wird die Arbeit grundsätzlich dem Wortlaut nach zu Hause ausgeführt. Gemäss dem Duden wird das Homeoffice als Arbeitsplatz im privaten Wohnraum, der mit Kommunikationstechnik ausgestattet ist, definiert.

Zunehmend wird der Begriff Telearbeit seltener genutzt, während Homeoffice mittlerweile im Berufsalltag ein gefestigter Anglizismus für die Leistungserbringung in den Räumlichkeiten des Arbeitnehmers zu Hause ist. Eine weitere Unterscheidung, ob die Arbeitsleistung tatsächlich am Wohnort des Arbeitnehmers oder an einem anderen Ort, beispielsweise einer Ferienwohnung, erbracht wird, findet oft nicht mehr statt (Domening, 2016).

Für dieses Projekt wird nicht weiter zwischen den Begriffen Homeoffice und Telearbeit unterschieden. Im Folgenden wird untersucht, wie sich die Anteile der im Homeoffice arbeitenden Erwerbstätigen verändert. Dabei werden die Tage erfasst, an denen die Person an ihrer Tätigkeitsstätte, in den Räumlichkeiten des Arbeitgebers, tätig wurde. Und zum anderen werden die Tage dokumentiert, an denen die Person die Arbeit von zu Hause verrichtet hat. Eine weitere Differenzierung der Ausgestaltung des Arbeitsplatzes im privaten Wohnraum, sowie die verwendeten Kommunikationsmittel findet nicht statt.

2.2 Beeinflussende Faktoren

In der Literatur sind verschiedene Konzepte und Ansätze zur Arbeit im Homeoffice zu finden. Zudem wurden schon vor Beginn der Covid-19 Pandemie, die unterschiedlichen Faktoren, die einen Einfluss auf die Arbeit im Homeoffice haben, mehrfach untersucht. Insbesondere wurden

die Auswirkungen unterschiedlicher soziodemographischer Merkmale analysiert. Da diese Merkmale im späteren Verlauf dieses Projekts in den Aufbau des Entscheidungsmodells einbezogen werden, werden im Folgenden die wichtigsten Einflussfaktoren dargestellt.

Auch wenn Arbeitsgeber ihren Angestellten die Arbeit im Homeoffice ermöglichen, nutzen viele Erwerbstätige das Angebot nicht. In der Literatur sind unterschiedliche Gründe hierfür zu finden.

Nachteile, die oft benannt werden, ist die Angst vor einer geringeren Produktivität zu Hause als am primären Arbeitsplatz. Neben der Ablenkung zu Hause spielt auch die fehlende Selbstdisziplin eine Rolle (Graaff und Rietveld, 2007). Ablenkung zu Hause entsteht oft durch weitere Familienmitglieder und fehlende Rückzugsorte zum Arbeiten. Ein wesentliches Merkmal ist daher die Haushaltsgrösse der jeweiligen Person. Eine weitere Schwäche des Homeoffice sehen viele Erwerbstätige in einer möglichen Benachteiligung bei Beförderungen und verringerte Aufstiegschancen. Aufgrund der geringeren Anwesenheit am Arbeitsplatz, könnten Arbeitsgeber die geleistete Arbeit nicht umfänglich wahrnehmen. Auch die Vernetzung und der Austausch mit anderen Kollegen ist durch die Arbeit im Homeoffice erschwert (Graaff und Rietveld, 2007). Um die Auswirkungen auf die Karrieremöglichkeiten zu betrachten, können zudem auch das Alter, das aktuelle Einkommen und der Bildungsgrad des Erwerbstätigen betrachtet werden. Insbesondere zeigten Untersuchungen, dass Personen mit niedrigerem Einkommen ein grösseres Interesse an der Arbeit im Homeoffice zeigten (Bloom *et al.*, 2015).

Neben den Nachteilen, die aus der Arbeit im Homeoffice entstehen können, gibt es auch Vorteile und Anreize für Erwerbstätige von zu Hause aus zu arbeiten. Die Zeit kann flexibler genutzt werden. Bei der Betreuung von beispielsweise Kindern kann Homeoffice eine Möglichkeit bieten, Beruf und Familie zu kombinieren (Graaff und Rietveld, 2007). Auch hierfür kann die Haushaltsgrösse als Hinweis verwendet werden.

Neben den Auswirkungen auf die Produktivität der Erwerbstätigen, kann Homeoffice aber auch Auswirkungen auf das Wohlbefinden der Person haben. Von zu Hause aus zu arbeiten erspart das Pendeln zur Tätigkeitsstätte und den damit verbundenen Stress. Die eingesparte Zeit kann flexibler genutzt werden (Graaff und Rietveld, 2007). Andererseits haben Menschen, die gerne mit ihren Kollegen zusammenarbeiten, weniger Kontakt und Kommunikationsmöglichkeiten mit diesen. Die Auswirkungen von Homeoffice auf das Wohlbefinden der Erwerbstätigen hängt damit stark von den persönlichen Präferenzen und weniger von allgemeinen soziodemographischen Merkmalen ab.

Ein oft benannter Vorteil des Homeoffice sind die besseren Ausgestaltungsmöglichkeiten der Work- Life-Balance. Homeoffice ermöglicht eine grössere Flexibilität in der Gestaltung der Freizeit und auch der Arbeitszeiten. Die Zeit, die durch das Pendeln entfällt, kann frei genutzt werden. Während der Standardarbeitszeiten können Erwerbstätige im Homeoffice auch unbezahlter Arbeit, beispielsweise kleine Arbeiten im Haushalt, nachgehen (Giménez-Nadal *et al.*, 2019). Auch können Eltern ihre Kinder während der Arbeitszeit betreuen.

In Untersuchungen zeigte sich, dass männliche Erwerbstätige im Homeoffice glücklicher und zufriedener sind als ihre Kollegen, die täglich zur Arbeit pendeln. Bei weiblichen Erwerbstätigen lässt sich hingegen kein Unterschied zwischen Pendlern und Personen im Homeoffice feststellen (Giménez-Nadal *et al.*, 2019). Eine mögliche Begründung könnte darin liegen, dass Frauen ihre neugewonnene Freizeit für unbezahlte Arbeit im Haushalt nutzen, während Männer mehr Zeit in ihre Freizeit investieren (Giménez-Nadal *et al.*, 2019). Die Auswirkungen von Homeoffice variieren demzufolge zwischen den Geschlechtern.

Auch das Alter der Person kann eine Rolle spielen, wie wohl sich Erwerbstätige im Homeoffice fühlen. Insbesondere neue Telekommunikationsmöglichkeiten, die bei der Arbeit im Homeoffice vermehrt genutzt werden, sind jüngeren Erwerbstätigen geläufiger. Die Anpassung an die Arbeitsumstände im Homeoffice gelingt dann schneller und die Person fühlt sich im Homeoffice schneller wohl.

Neben den hier vorgestellten Faktoren werden in Kapitel 5.2.1 weitere Faktoren vorgestellt, die die besonderen Umstände der Covid-19-Pandemie berücksichtigen.

3 Methoden zum Umgang mit Daten

3.1 Datenbasis

3.1.1 Forschungsprojekt MOBIS: COVID-19

Die für dieses Projekt verwendeten Daten wurden im Zeitraum zwischen März 2020 und November 2020 im Rahmen des Forschungsprojekts «MOBIS: COVID-19» erhoben. Das Forschungsprojekt MOBIS: COVID-19 ist eine Weiterführung des Forschungsprojektes «Mobilitätsverhalten in der Schweiz» (MOBIS), welches zwischen September 2019 und Januar 2020 durchgeführt wurde. Das ursprüngliche Forschungsprojekt zum Mobilitätsverhalten in der Schweiz ist eine Zusammenarbeit der ETH Zürich, der Universität Basel und der Zürcher Hochschule für Angewandte Wissenschaften.

Im Rahmen der MOBIS Studie wurden 90'000 zufällig ausgewählte Bewohner der Schweizer Agglomerationen zu einer Onlinebefragung eingeladen. Teilnehmer, die die Befragung beantworteten und die Voraussetzungen erfüllten, wurden anschliessend zu einer vertiefenden zweimonatigen Smartphone-Studie und einer weiteren abschliessenden Befragung eingeladen. Mit Hilfe der GPS-Logger- und Reisetagebuch-App 'Catch-My-Day' wurden die Reisen der Teilnehmer aufgezeichnet. An der initialen Befragung nahmen 90'909 Personen teil.

Mit Beginn der COVID-19 Pandemie wurde beschlossen, die ursprüngliche Studie als MOBIS: COVID-19 weiterzuführen, um die Auswirkungen der Covid-19 Pandemie auf das Mobilitätsverhalten der Schweizer Bevölkerung zu untersuchen. Am 16. März 2020 wurden die 3'690 Teilnehmer, die die MOBIS-Studie zwischen September 2019 und Januar 2020 abgeschlossen hatten, erneut eingeladen, ihr Mobilitätsverhalten aufzuzeichnen.

Im Zeitraum zwischen März und November 2020 zeichneten 1'300 Teilnehmer ihr Mobilitätsverhalten zum zweiten Mal auf. Von diesen 1'300 Teilnehmern zeichneten 250 Teilnehmer ihr Mobilitätsverhalten zudem auch zwischen den beiden Studien, im Zeitraum von Januar bis März 2020, auf.

3.1.2 Verfügbare Daten und Abgrenzung

Für dieses Projekt werden für die Erstellung des Entscheidungsmodells die aufgezeichneten Daten der Teilnehmer zwischen Februar und November 2020 verwendet. Zudem stehen vier

Umfragen zur Verfügung. Neben der initialen und der finalen Befragung der MOBIS Studie werden die beiden Befragungen, die die Teilnehmer während der MOBIS: COVID-19 Studie beantwortet haben, genutzt. Die initiale Befragung des MOBIS Projektes diente primär der Erhebung der soziodemographischen Daten wie beispielsweise Alter, Geschlecht und Haushaltsgrösse. In der finalen Befragung der MOBIS Studie wurden dieselben Parameter erneut erhoben, um Veränderungen der soziodemographischen Daten festzustellen. Für dieses Projekt werden die erhobenen Daten der initialen Befragung mit den Daten der finalen Befragung abgeglichen. Haben sich soziodemographische Daten der Teilnehmer verändert, so werden die Daten aktualisiert. Während der anschliessenden MOBIS: COVID-19 Studie wurde zu Beginn im April 2020 eine weitere Befragung der Teilnehmer durchgeführt. Zudem fand eine zweite Befragung im Juli 2020 statt. Die beiden Erhebungen beinhalteten insbesondere Fragen zum Beschäftigungsverhältnis sowie der Möglichkeit der Kurzarbeit.

Da die initiale Befragung der MOBIS-Studie eine Voraussetzung zur Teilnahme an der folgenden Studie war, liegen für alle 3'690 Teilnehmer die Antworten der Befragung vor. An der ersten Erhebung während der MOBIS: COVID-19 Studie nahmen 1'165 Personen teil. Die zweite Erhebung, die insbesondere aktualisierte Daten der Beschäftigungsverhältnisse erfragt, beantworteten 425 Teilnehmer.

3.2 Imputation fehlender Daten

3.2.1 Einführung in die Imputation

Wie im vorangegangenen Abschnitt beschrieben, liegen nicht für alle Teilnehmer Daten aus allen drei Erhebungen vor. Um Analysen an einem unvollständigen Datensatz durchzuführen, gibt es verschiedene Ansätze. Eine Möglichkeit sind die «Complete Case» Analysen bei denen nur Beobachtungen, bei denen alle Variablen vollständig vorliegen, berücksichtigt werden (Spiess, 2010). Da in diesem Fall nur 405 Datensätze vollständig vorliegen, ist die Anwendung dieses Ad-hoc-Verfahrens nicht sinnvoll.

Eine andere Möglichkeit besteht darin, die fehlenden Werte durch möglichst plausible Werte zu ersetzen. Dieses Vorgehen wird als Imputation bezeichnet. Bei den Imputationen werden zwei Gruppen unterschieden. Die einfachen Imputationsverfahren erzeugen für jeden fehlenden Wert genau eine Imputation, wohingegen die multiplen Imputationsverfahren für jeden Wert mehrere Imputationen erzeugen. Für einen Datensatz entstehen dementsprechend mehrere

vollständige Exemplare, bei denen die beobachteten Werte identisch sind, während sich die imputierten Werte unterscheiden (Spiess, 2010).

Multiple Imputationsverfahren bestehen aus drei Schritten. Zunächst werden mit Hilfe des gewählten Verfahrens m imputierte Datensätze erstellt. Die unterschiedlichen Datensätze werden anschliessend einzeln analysiert. Der dritte Schritt kombiniert die Einzelergebnisse zu einem Gesamtergebnis. Beispielsweise kann der Mittelwert der m Einzelergebnisse gebildet werden (Spiess, 2010).

Durch mehrmaliges Imputieren werden die Unsicherheiten in den Daten berücksichtigt und dementsprechend die wahre Varianz der Daten besser abgebildet. Nach Spiess (2010) liefern multiple Imputationen zudem im Vergleich zu einfachen Imputationen bessere Ergebnisse. Um die fehlenden Daten in diesem Projekt zu schätzen wird daher ein multiples Imputationsverfahren angewandt.

3.2.2 Expectation-Maximization-Algorithmus

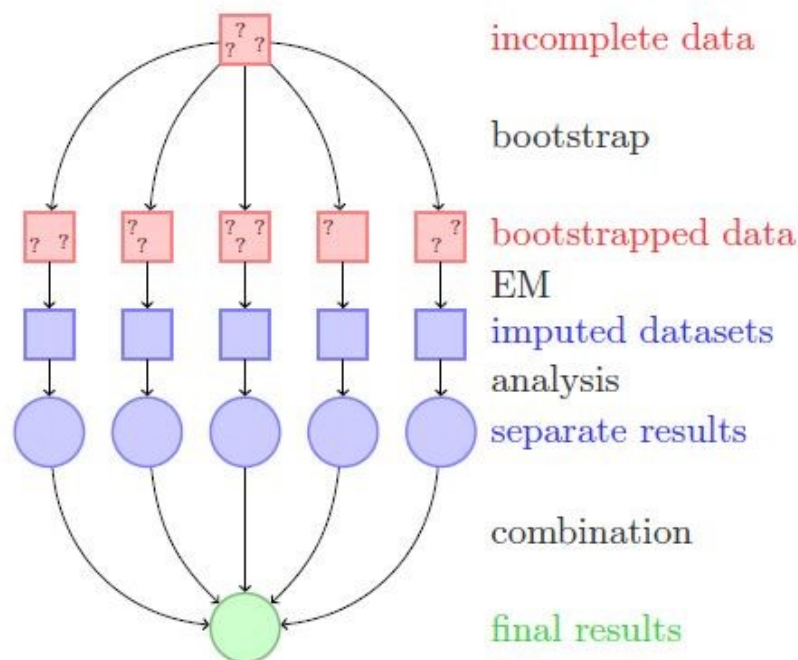
Der Expectation-Maximization-Algorithmus (EM-Algorithmus), auch als Erwartungswert-Maximierungs-Verfahren bekannt, ist eine Methode zur Imputation. Er besteht aus dem Estimation-Schritt (E-Schritt) und dem Maximization-Schritt (M-Schritt). Im E-Schritt werden die fehlenden Beobachtungen ergänzt. Mit Hilfe der Maximum-Likelihood-Methode wird anschliessend im M-Schritt der Parameter des Modells geschätzt (Dempster *et al.*, 1977).

Um den EM-Algorithmus anzuwenden, steht in R das Paket Amelia II zur Verfügung. Dieses Paket verwendet zur Imputation den EMB-Algorithmus (expectation-maximization with bootstrapping). Für die beobachteten (D_{obs}) sowie die fehlenden Daten (D_{mis}) wird eine multivariate Normalverteilung mit Mittelwert μ und der Kovarianzmatrix Σ angenommen.

Im ersten Schritt wird ein Bootstrapping-Verfahren zur Erstellung der m verschiedenen Datensätze verwendet. Dabei werden aus dem vorliegenden unvollständigen Datensatz m verschiedene Datensätze zufällig gezogen. Mit dem anschliessenden mehrfachen und zufälligen Resampling stellt die EMB-Technik Variation zwischen den Datensätzen her (Honaker *et al.*, 2011). Anschliessend werden mit Hilfe des EM-Algorithmus die fehlenden Werte in die m Datensätze imputiert.

Abbildung 1 stellt den schematischen Ablauf der Imputation mit dem Paket Amelia unter Verwendung des EMB-Algorithmus dar, wobei die Analyse und Kombination der imputierten Datensätze nicht in der Funktion Amelia enthalten sind.

Abbildung 1 Schematische Darstellung der Imputation mit dem R-Paket Amelia unter Verwendung des EMB-Algorithmus



Quelle: Honaker *et al.*, 2011 S.5 Abbildung 1

Zudem wird die MAR-Annahme getroffen. Die MAR-Annahme (missing at random) besagt, dass das Muster der fehlenden Werte nur mit den beobachteten Daten zusammenhängt. Mit Hilfe dieser Annahme können die fehlenden Werte aufgrund der verfügbaren Informationen korrigiert werden (Honaker *et al.*, 2011). Die Daten dieses Projekts basieren auf Befragungen der Teilnehmenden. Es kann davon ausgegangen werden, dass Teilnehmer zufällig die Befragungen nicht ausgefüllt haben. Das heisst, dass die Teilnehmer beispielsweise aus Zeitgründen oder mangels Interesses den gesamten Fragebogen nicht beantwortet haben und nicht bestimmte Fragen absichtlich nicht beantwortet haben. Dementsprechend kann für die vorliegenden Daten die MAR-Annahme getroffen werden.

Unter Verwendung der Standardeinstellung des Pakets werden für jeden fehlenden Wert fünf Werte erzeugt. Für den Fall, dass die Anzahl der fehlenden Werte hoch ist, kann die Anzahl der Imputationen auch erhöht werden (Honaker *et al.*, 2011). Das genaue Vorgehen und die Wahl der einzelnen Parameter werden in Kapitel 4.1 beschrieben.

3.2.3 Fully Conditional Specification (FCS)

Neben dem EM-Algorithmus gibt es weitere Verfahren, die für multiple Imputationen geeignet sind. Ein weiteres Verfahren ist das FCS-Verfahren (fully conditional specification). Das im vorangegangenen Kapitel beschriebene Verfahren basiert auf dem Ansatz des «joint modelling». Dementsprechend wird eine gemeinsame Verteilung für den gesamten Datensatz modelliert. Beim FCS-Verfahren werden hingegen für jede unvollständige Variable bedingte Verteilungen definiert. Einer binären Variablen wird dann ein logistisches Schätzmodell zugewiesen, während für eine ordinale Variable ein multinominal logistisches Schätzmodell verwendet wird (Wahl und Urban, 2020). Für den verwendeten Datensatz in diesem Projekt bietet sich aufgrund einer hohen Anzahl an binären Variablen hierdurch ein Vorteil im Vergleich zum EMB-Algorithmus.

Eine mögliche Umsetzung des FCS-Verfahrens bietet das R-Paket «mice». Mice steht hierbei für «multivariate imputation by chained equations». Der zugrundeliegende Algorithmus der Funktion mice kann nach Wahl und Urban (2020) in drei wesentliche Schritte unterteilt werden:

1. Nach der Zuweisung der einzelnen Schätzmodelle der Variablen, wird der unvollständige Datensatz zunächst mit Platzhaltern vervollständigt. Die Platzhalter werden aus einer einfachen Stichprobe der beobachteten Werte erstellt.
2. Im zweiten Schritt werden die Platzhalter einer Variable Y wieder gelöscht, während die Platzhalter der anderen Variablen erhalten bleiben.
3. Für die unvollständige Variable Y wird anschliessend ein Modell unter Berücksichtigung der Skalierung der Variablen spezifiziert. Mit diesem Modell werden dann die fehlenden Werte der Variable geschätzt und imputiert.

Anschliessend werden die Schritte zwei und drei für jede Variable im Datensatz, die imputiert werden soll, durchgeführt. Wenn für alle Variablen, die fehlenden Werte imputiert wurden, hat das Verfahren eine Iteration durchlaufen. Der gesamte Prozess wird so lange wiederholt, bis die angegebene Anzahl an Iterationen durchlaufen wurde. Die Wahl der Iterationen hängt von

der Grösse des Datensatzes sowie der Anzahl der fehlenden Werte ab. Es wird empfohlen, zwischen 5 und 20 Iterationen zu benutzen (Wahl und Urban, 2020).

Aufgrund der hohen Anzahl an fehlenden Werten wird der unvollständige Datensatz im Folgenden vorbereitet und mit der Funktion *Amelia* sowie mit der Funktion *mice* imputiert. Anschliessen werden die beiden Imputationen verglichen und die geeignetere Imputation für den Aufbau des Entscheidungsmodells gewählt. In Kapitel 4.1.1 wird auf die Wahl der einzelnen Parameter und die Durchführung eingegangen.

4 Datenanalyse

4.1 Imputation

4.1.1 Vorbereitung der Daten

Für multiple Imputationsverfahren ist es zunächst notwendig, die Variablen für das Imputationsmodell auszuwählen. Im Imputationsmodell sollten alle Variablen und Verknüpfungen enthalten sein, die auch später im Modell der Analyse enthalten sind. Die Verwendung zusätzlicher Variablen kann zudem das Vorhersagemodell verbessern (Honaker *et al.*, 2011).

In diesem Projekt sollen die drei Erhebungen zu einem Gesamtdatensatz zusammengefasst werden. Die fehlenden Werte in dem Datensatz werden imputiert. Für die Imputationen der fehlenden Daten gibt es zwei Möglichkeiten. Ein mögliches Vorgehen wäre zunächst die initiale MOBIS Umfrage und die erste Umfrage der MOBIS: COVID-19 Studie zusammenzufassen und die fehlenden Daten aus diesem Datensatz zu imputieren. Anschliessend wird der zusammengefasste Datensatz um die Daten der zweiten Erhebung der MOBIS: COVID-19 Studie ergänzt und die wiederum fehlenden Daten mit Hilfe einer Imputation ergänzt. Eine andere Möglichkeit wäre die Daten aller drei Erhebungen von Anfang in einem Datensatz zusammenzufassen und mit nur einer Imputation alle fehlenden Beobachtungen zu vervollständigen. Der Vorteil dieses Vorgehens ist die grössere Breite an Variablen, die dadurch zur Verfügung stehen. Da die Verwendung zusätzlicher Variablen das Vorhersagemodell verbessern kann, wird dieses Vorgehen in diesem Fall angewandt.

Zunächst wird der Datensatz der initialen MOBIS Erhebung mit den neuen Daten der finalen Befragung der MOBIS Studie aktualisiert. Anschliessend wird der Datensatz mit allen verfügbaren Beobachtungen aus den beiden Erhebungen der MOBIS: COVID-19 Studie gefüllt. Anschliessend werden doppelte Variablen und Variablen, die zur Imputation nicht benötigt werden, aus dem Datensatz gelöscht. Variablen werden zudem zusammengefasst, sodass die Informationen erhalten bleiben, die Anzahl der fehlenden Werte jedoch reduziert wird.

Neben den Daten aus den Erhebungen werden dem Datensatz zwei weitere Variablen hinzugefügt. Diese Variablen berücksichtigen die Bewegungsaktivität der Teilnehmer. Aus den vorliegenden Bewegungsdaten werden Aktivitätsquoten der einzelnen Teilnehmer berechnet. Anhand der Bewegungen wird für jeden Teilnehmer für jeden einzelnen Tag erfasst, ob dieser aktiv war, das heisst eine Bewegung aufgezeichnet wurde, oder nicht. Aus diesen Daten werden durchschnittliche Aktivitätsquoten für jeden Teilnehmer berechnet. Der Wert 1 bedeutet, dass

der Teilnehmer durchschnittlich an einem Tag in der Woche aktiv war, wohingegen der Wert 6 Aktivitäten an 6 aus 7 Tagen in der Woche ausdrückt. Die Berechnung der Variablen erfolgt für zwei Zeiträume. Eine Variable berücksichtigt den Zeitraum vor der Pandemie, von November 2019 bis Anfang März 2020. Die zweite Variable charakterisiert den Zeitraum während der Pandemie ab März 2020.

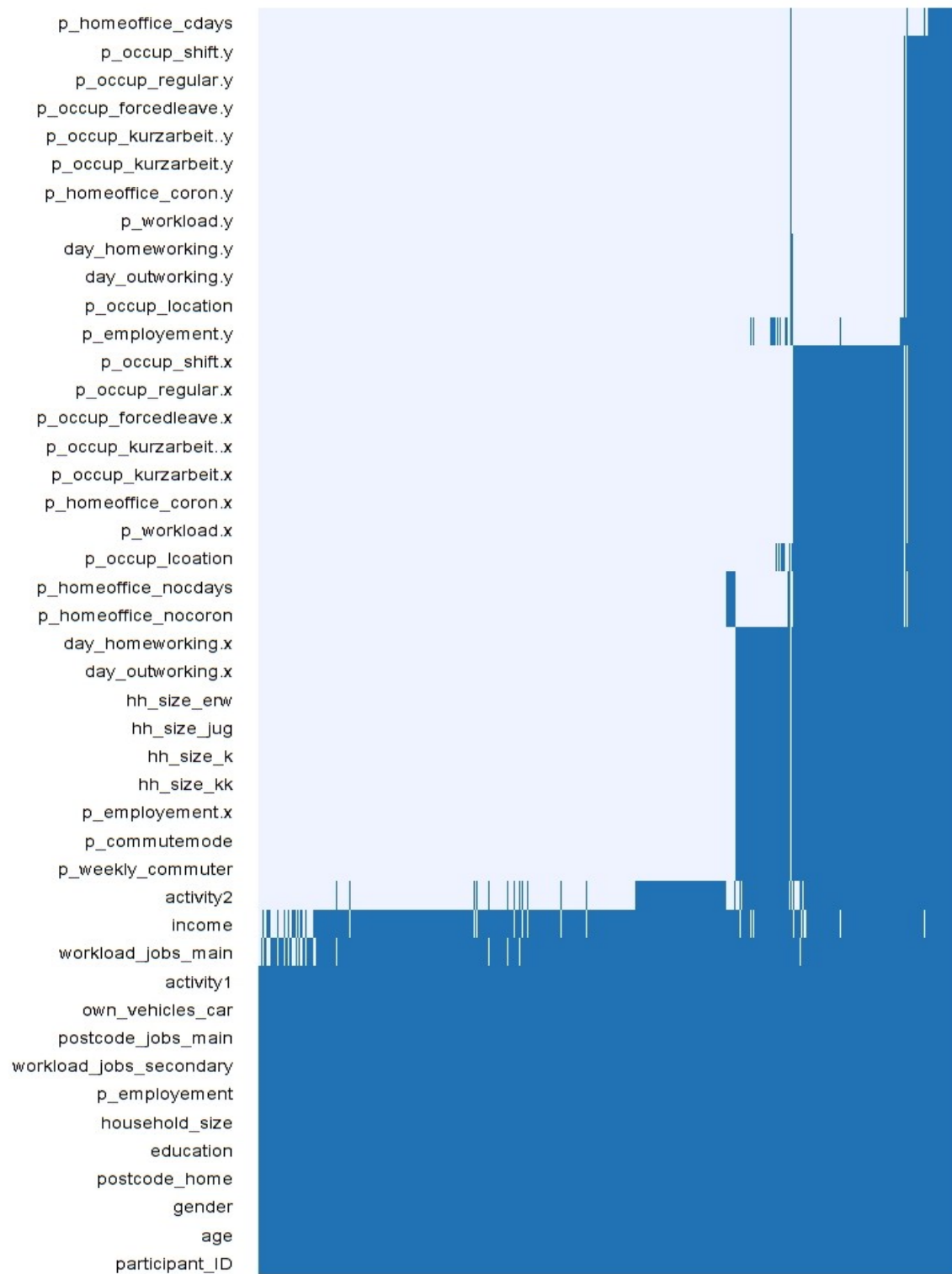
4.1.2 Imputation mit Amelia

Da die Imputation zweifach ausgeführt wird, wird der zusammengefasste Datensatz anschliessend jeweils an die Anforderungen der beiden Funktionen angepasst.

Für die Imputation mit dem Paket Amelia müssen zudem die kategorischen Variablen in numerische Variablen umgewandelt werden. Beispielsweise wird die Unterscheidung des Geschlechts für weiblich (female) und männlich (male) in 0 und 1 geändert. Zusätzlich muss der Funktion für die Durchführung der multiplen Imputation mit dem Paket Amelia II die Information übergeben werden, bei welchen Variablen es sich um ordinale beziehungsweise nominale Variablen handelt. In Anhang A 1.1 gibt Tabelle 12 einen Überblick über die vorhandenen Variablen und die zugehörigen Ausprägungen.

Der vorbereitete Datensatz wird anschliessend mit Hilfe des R-Pakets Amelia II imputiert. Vor der Imputation kann der unvollständige Datensatz analysiert werden, um die einzelnen Parameter für die Imputation zu wählen. Abbildung 2 zeigt die fehlenden Daten im vorbereiteten Datensatz. Der Datensatz beinhaltet 45 verschiedene Variablen. Der dunkelblaue Bereich stellt die vorhandenen Daten (43%) dar, während der hellblaue Bereich die fehlenden Beobachtungen (57%) repräsentiert.

Abbildung 2 Anteil fehlender Werte im Datensatz vor Imputation



Mit Hilfe der Analyse des vorliegenden Datensatzes können die Parameter für die Imputation mit dem R-Paket Amelia II festgelegt werden. Aufgrund des Umfangs an fehlenden Beobachtungen werden zehn imputierte Datensätze erstellt, die anschliessend analysiert und zu einem Datensatz zusammengefasst werden. Neben der Verwendung der Standardeinstellungen der Funktion Amelia können weitere Parameter nach Honaker (2019) selbst gewählt werden. Mit dem Parameter «idvars» wird beispielsweise festgelegt, dass die Variable «participant_ID» für die Imputation nicht genutzt wird, zur Identifikation jedoch erhalten bleibt. Der Parameter «empri» ermöglicht Imputationen bei Datensätzen mit einer hohen Anzahl an fehlenden Werten, wie dem Vorliegenden, mit Hilfe einer A-priori-Verteilung. Die Anzahl der gewählten Werte sollte klein gehalten werden und maximal zehn Prozent der Beobachtungen betragen (Honaker *et al.*, 2019). Unter Berücksichtigung der fehlenden Datenmenge wurde der Parameter «empri» auf 300 gesetzt.

4.1.3 Imputation mit mice

Um die Imputation mit der Funktion mice durchzuführen, werden die kategorischen Variablen als Typ «factor» im Datensatz gespeichert. Da im FCS-Verfahren die Verteilung für jede einzelne Variable definiert wird, müssen die jeweiligen Verteilungen der Funktion übergeben werden (van Buuren und Oudshoorn 2000). Tabelle 1 stellt die für den Datensatz ausgewählten Imputationsmethoden dar. Die Zuordnung der Methoden zu den einzelnen Variablen unter der Verwendung von mice ist ebenfalls in Tabelle 12 im Anhang A 1.1 dokumentiert.

Tabelle 1 Ausgewählte Imputationsmethoden der Funktion mice in R

Skalierung der Variable	Methode	Regressionsmodell
numerisch	pmm	predictive mean matching
binär	logreg	logistische Regression
nominal (> 2 Level)	polyreg	multinominales Logit-Modell
ordinal	polr	kumulatives Logit-Modell

Quelle: van Buuren und Oudshoorn, 2000

Wie schon im vorangegangenen Abschnitt erläutert, ist die Anzahl der fehlenden Werte sehr hoch, weshalb auch für die Funktion mice zehn Imputationen durchgeführt werden. Weiterhin können die Anzahl der Iterationen (Standardwert 5) erhöht werden (van Buuren und Groothuis-Oudshoorn, 2021). Auch hier wird die Anzahl auf 20 erhöht.

4.2 Analyse der imputierten Datensätze

Unter Verwendung der beiden Funktionen entstehen jeweils zehn imputierte Datensätze, bei denen die vorher beobachteten Werte D_{obs} in allen Datensätzen identisch sind, während sich die imputierten Werte unterscheiden. Die imputierten Daten werden auf Unstimmigkeiten und Ausreisser überprüft. Anschliessend müssen die einzelnen Datensätze zu jeweils einem Datensatz zusammengefasst werden. Hierfür existieren verschiedene Möglichkeiten. Mit Hilfe einer Regression können die imputierten Datensätze angepasst werden. Ein weiteres mögliches Vorgehen ist die Bildung der Mittelwerte der imputierten Daten.

Für dieses Projekt wird für beide Imputationsverfahren jeweils ein Datensatz aus den Mittelwerten der zehn imputierten Datensätze gebildet. Für die kategorialen Variablen im Datensatz wird der Modalwert verwendet. Diese beiden Datensätze werden nur zur Auswahl der Imputationsmethode verwendet. Die beiden resultierenden Datensätze der beiden Methoden werden verglichen und die Methode mit den plausibleren Ergebnissen wird ausgewählt. Um genauere Ergebnisse zu erhalten wird anschliessend das Entscheidungsmodell für alle zehn imputierten Datensätze der ausgewählten Methode erstellt. Abschliessend werden dann die einzelnen Regressionskoeffizienten gemittelt.

Um die Güte der imputierten Datensätze zu beurteilen, wird geprüft, inwiefern die Verteilung der nicht imputierten Werte mit den imputierten Werten übereinstimmt. Der unvollständige Datensatz beinhaltet 45 verschiedene Variablen. Um das Vorhersagemodell zu verbessern wurden dem Datensatz mehr Variablen hinzugefügt, als später für den Aufbau des Entscheidungsmodells benötigt werden. Die Auswertung der Imputationen erfolgt für die Variablen, die für den Aufbau des Entscheidungsmodells anschliessend verwendet werden. Im Folgenden werden die Verteilungen vier verschiedener Variablen des Datensatzes für die beiden Imputationsmethoden dargestellt und verglichen. Die beiden Variablen «workload.x» und «workload.y» enthalten die Angaben des allgemeinen Arbeitspensums zu den jeweiligen Zeitpunkten der beiden Erhebungen. Die Variablen «kurzarbeit.x» und «kurzarbeit.y» beinhalten das angegebene Arbeitspensum während der Kurzarbeit zu den Zeitpunkten der beiden Erhebungen.

Jede Abbildung stellt für eine Variable die imputierten Werte der Funktion «Amelia» den imputierten Werten der Funktion «mice» gegenüber. Die imputierten Werte werden jeweils durch die hellblauen Punkte dargestellt. Die dunkelblauen Punkte repräsentieren die Verteilungen der die Werte des ursprünglichen, unvollständigen Datensatzes. Die horizontale Ausbreitung der

beiden Verteilungen entlang der x-Achse erfolgt nur zur besseren Sichtbarkeit der einzelnen Werte und hat keine weitere Bedeutung. Anhand der Verteilungen lässt sich erkennen, dass die Verteilung des Datensatzes nach der Imputation mit dem Paket mice besser zu der Verteilung des unvollständigen Datensatzes passt.

Abbildung 3 Vergleich der Verteilungen der imputierten Werte der Variable «workload.x»

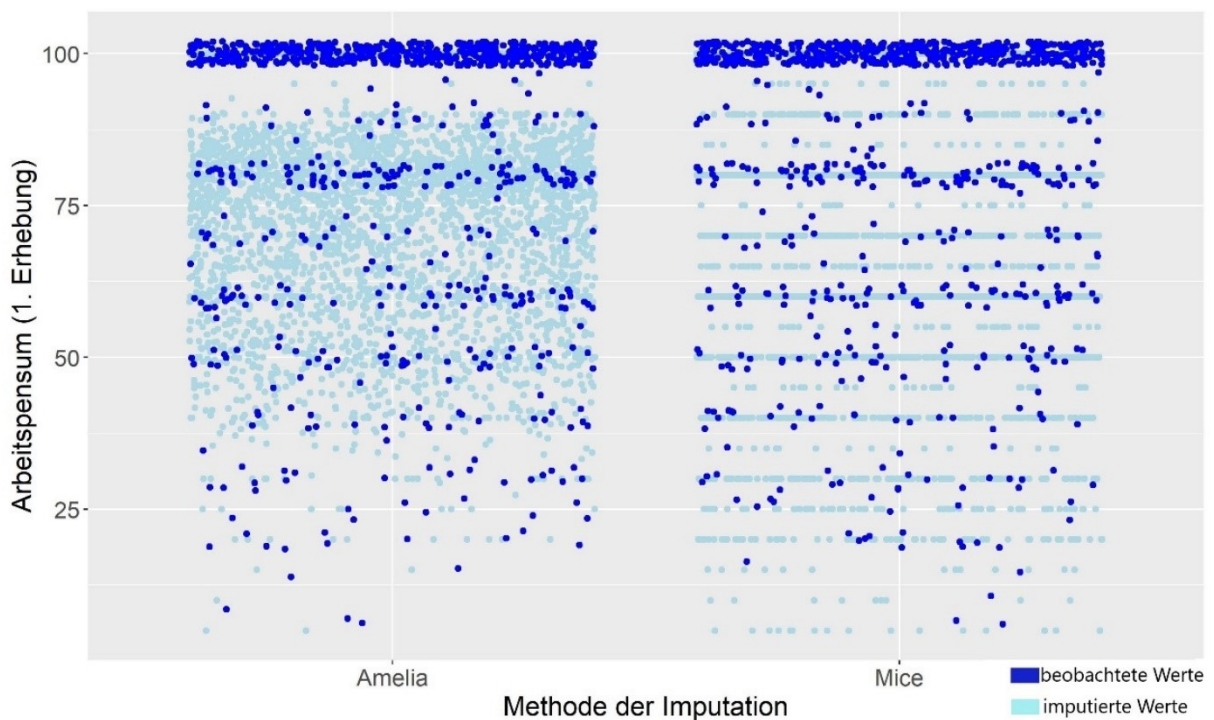


Abbildung 4 Vergleich der Verteilungen der imputierten Werte der Variable «workload.y»

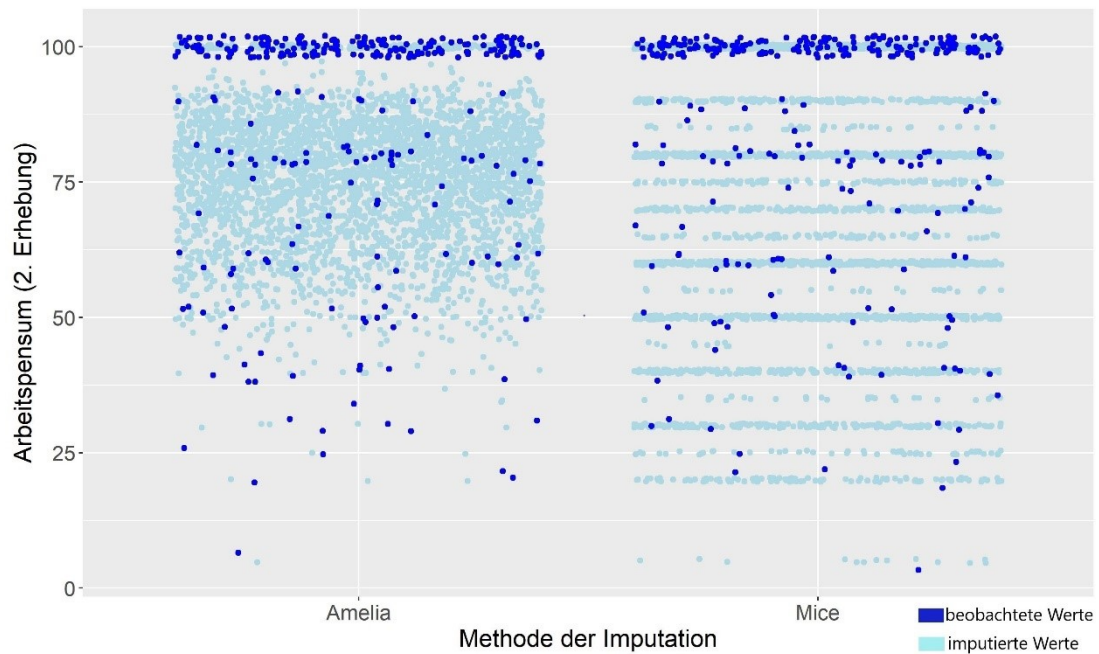


Abbildung 5 Vergleich der Verteilungen der imputierten Werte der Variable «kurzarbeit.x»

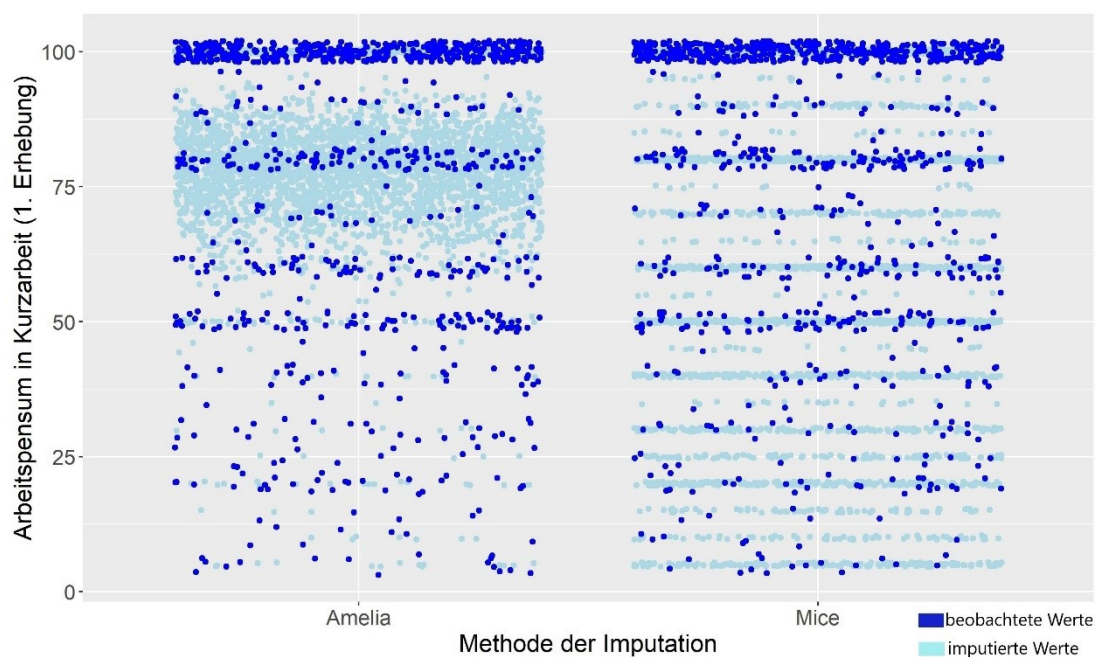
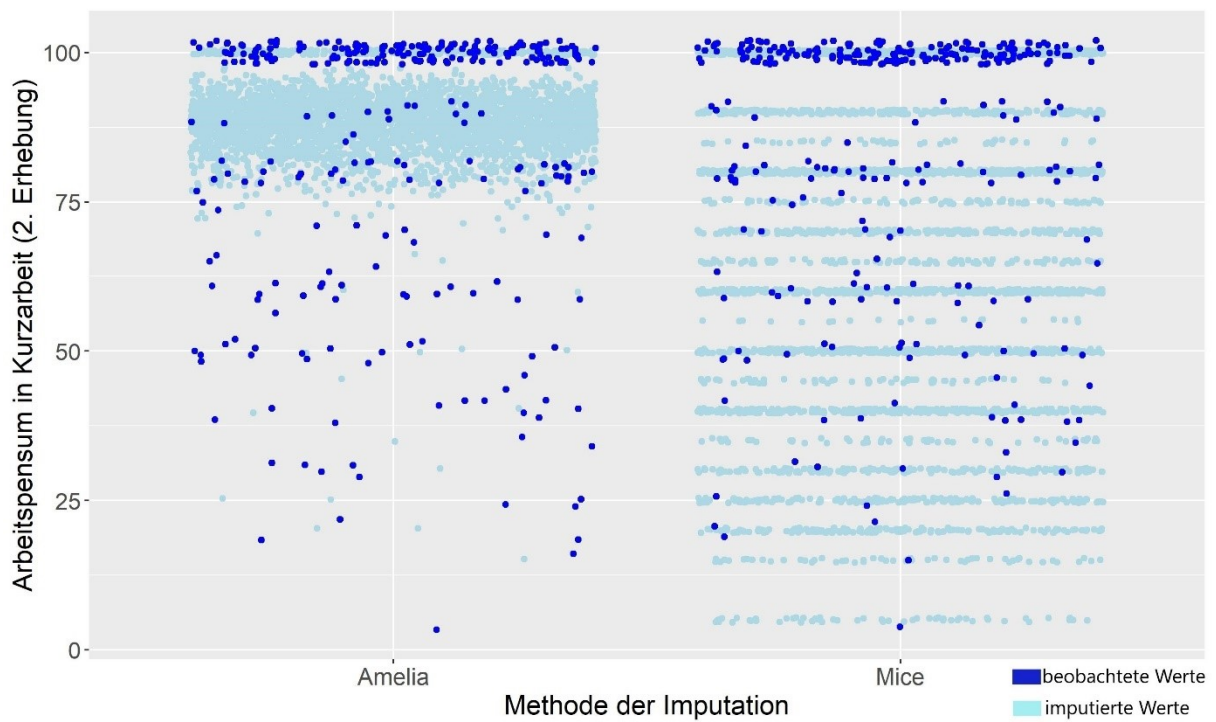


Abbildung 6 Vergleich der Verteilungen der imputierten Werte der Variable «kurzarbeit.y»



In Tabelle 2 und Tabelle 3 werden die Lage- und Streumasse der untersuchten Variablen zusammengefasst. Da die Variablen «workload» und «kurzarbeit» metrische Variablen sind, eignet sich als Lagemass der Mittelwert. Für das Streumass werden die Standardabweichung in Tabelle 3 angegeben. Aufgrund der Verteilungen wird der Datensatz, der mit Hilfe des FSC-Verfahrens und der Funktion «mice» erstellt wurde, für die weiteren Berechnungen verwendet.

Tabelle 2 Mittelwerte der Variablen für die unterschiedlichen Datensätze

Bezeichnung Variable	Originaler Datensatz	Imputation mit Mice	Imputation mit Amelia
workload_jobs_main	68.16	68.73	68.63
p_workload.x	85.05	74.59	73.79
p_workload.y	86.68	72.36	76.11
p_occup_kurzarbeit..x	77.90	65.48	80.51
p_occup_kurzarbeit..y	84.45	65.26	88.58
p_homeoffice_coron.x	0.33	0.49	0.45
P_homeoffice_coron.y	0.34	0.50	0.48

Tabelle 3 Standardabweichungen der Variablen für die unterschiedlichen Datensätze

Variable	Originaler Datensatz	Imputation mit Mice	Imputation mit Amelia
workload_jobs_main	40.66	40.28	39.92
p_workload.x	22.62	24.31	17.22
p_workload.y	21.34	24.75	12.16
p_occup_kurzarbeit..x	27.63	31.38	14.08
p_occup_kurzarbeit..y	22.61	26.94	6.19
p_homeoffice_coron.x	0.48	0.50	0.50
P_homeoffice_coron.y	0.48	0.50	0.50

5 Vorbereitung des Choice Models

5.1 Discrete-Choice-Model

Ziel dieser Arbeit ist die Wahrscheinlichkeit zu bestimmen, dass Personen in Abhängigkeit unabhängiger Variablen von zu Hause aus arbeiten. Die Wahlentscheidung hängt von verschiedenen Einflussgrößen ab. Mit Hilfe von sogenannten Choice-Modellen wird eine Entscheidung (abhängige Variable) unter Berücksichtigung bestimmter Umwelteinflüsse (unabhängige Variablen) modelliert (Temme, 2007).

Da die Entscheidung immer nur bestimmte Werte annehmen kann, handelt es sich bei der abhängigen Variablen nicht um eine stetige, sondern um eine diskrete Variable. Im vorliegenden Fall kann das Ergebnis der Entscheidung immer nur zwei Werte annehmen, die Person arbeitet von zu Hause oder nicht. Dementsprechend ist die abhängige Variable «WFH» binär. Um die Abhängigkeit diskreter Variablen zu modellieren, können Discrete-Choice-Modelle verwendet werden, die den Zusammenhang einer kategorial abhängigen Variablen und einer oder mehrerer unabhängiger Variablen, die sowohl metrisch als auch kategorial skaliert sein können, untersuchen (Temme, 2007).

Discrete-Choice-Modelle basieren auf dem Ansatz der logistischen Regression.

Mit Hilfe von beobachteten Daten berechnen diese die Wahrscheinlichkeit, dass ein Individuum n aus einer Anzahl an Möglichkeiten (Choice Set C) Alternative i wählt. Die Wahrscheinlichkeit $P_n(i)$ für die Wahl von Alternative i hängt von dem persönlichen Nutzen U_n der Alternative ab. Da der Nutzen der Alternative U_{in} nicht beobachtbar ist, wird dieser über beobachtbare Eigenschaften des Individuums sowie der Alternativen operationalisiert (Temme, 2007).

Da im vorliegenden Fall nur zwei Alternativen zur Auswahl stehen, kann ein binäres Logit-Modell für die Modellierung verwendet werden. Mit Hilfe der Daten aus den drei Erhebungen sowie den aufgezeichneten Bewegungsdaten wird ein Datenmodell aus Beobachtungen erstellt. Auf Basis dieser Beobachtungen kann anschliessend mit Hilfe des linearen Regressionsansatzes das Modell erstellt werden.

Für beide Alternativen werden die Koeffizienten mit Hilfe des linearen Regressionsansatzes berechnet. Beispielsweise ergibt sich für die Alternative «WFH» dann:

$$y_{WFH} = \sum_{i=1}^j b_i * x_i$$

Die Wahrscheinlichkeit für die Arbeit aus dem Homeoffice berechnet sich dann nach:

$$P_{WFH} = \frac{e^{y_{WFH}}}{e^{y_{WFH}} + e^{y_{\overline{WFH}}}}$$

Die Koeffizienten der Regression werden mit dem Statistikprogramm R berechnet.

5.2 Wahl der Variablen

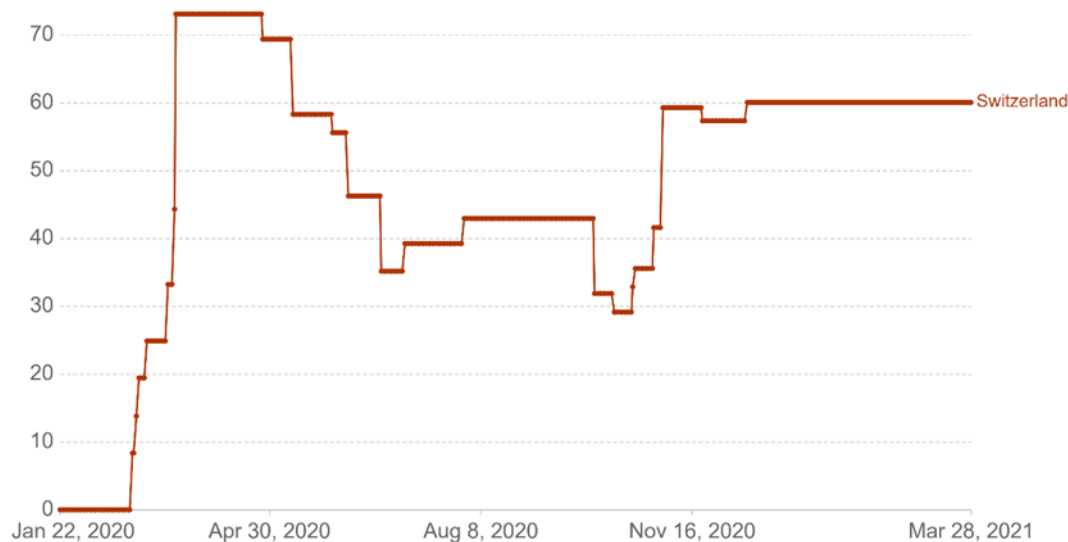
5.2.1 Auswahl der unabhängigen Variablen

Wie in Kapitel 2.3 dargestellt hängt die Entscheidung von zu Hause zu arbeiten von verschiedenen soziodemografischen Eigenschaften ab. Für den Aufbau des Entscheidungsmodells wurden die unabhängigen Variablen Geschlecht, Alter, und Haushaltsgrösse ausgewählt. Anstatt der Grösse des Haushalts könnte auch die Anzahl der im Haushalt lebenden Kinder mit einbezogen werden. Insbesondere für Familien, bei denen die Betreuungsmöglichkeiten ihrer Kinder eingeschränkt wurden, hätte dies eine Auswirkung auf die Entscheidung von zu Hause aus zu arbeiten. Da diese Variable imputiert wurde, die Haushaltsgrösse aber von allen teilnehmenden Personen in der ersten MOBIS Erhebung beantwortet wurde, wurde hier die Haushaltsgrösse als unabhängige Variable gewählt, um die Auswirkungen der Unsicherheiten aus den Imputationen zu reduzieren. Zudem werden die Merkmale Einkommen und Bildungsgrad als unabhängige Variablen ausgewählt.

Für jeden Teilnehmer ist das genaue Alter dokumentiert. Für die Schätzungen des Entscheidungsmodells werden Altersgruppen in kategorialen Variablen zusammengefasst. Die Aufteilung der Altersgruppen wird in Abschnitt 6.2 dargestellt.

Neben den soziodemografischen Variablen werden zwei weitere unabhängige Variablen eingeführt. Eine Variable repräsentiert die Massnahmen und Verordnungen der Regierung zur Eindämmung der COVID-19 Pandemie. Eine Möglichkeit, die Massnahmen einzuordnen und zu kategorisieren, bietet hierbei der «Oxford COVID-19 Government Response Tracker» (OxCGRT). Dieser beinhaltet 20 Indikatoren für Massnahmen, die von der Regierung zur Bekämpfung der COVID-19 Pandemie ergriffen wurden. Hierzu zählen beispielsweise Reisebeschränkungen, Schulschliessungen und Maskentrageverordnungen. Aus dem Datensatz werden vier verschiedene Indizes für die untersuchten Länder erstellt. Der «stringency index» bildet die Schärfe der Restriktionen im zeitlichen Verlauf der COVID-19 Pandemie ab (Hale *et al.*, 2021). Abbildung 7 zeigt den zeitlichen Verlauf des Indizes für die Schweiz.

Abbildung 7 COVID- 19 Stringency Index für die Schweiz im zeitlichen Verlauf der COVID-19 Pandemie



Quelle: Hale *et al.*, 2021

Da die Restriktionen massgeblich die Bürger in ihrem Verhalten beschränken und dementsprechend auch einen Einfluss auf die Entscheidung haben, ob Teilnehmer von zu Hause arbeiten, kann dieser Index als unabhängige Variable für das Entscheidungsmodell genutzt werden.

Eine detaillierte Betrachtung der Restriktionen für die einzelnen Schweizer Kantone ermöglicht der «KOF Stringency Index» der Konjunkturforschungsstelle der ETH Zürich. Dieser wird aufbauend auf dem stringency index der Oxford University berechnet. Die Werte gehen von 0 (= keine Massnahmen) bis 100 (= vollständiger Lockdown). Der Index beinhaltet die folgenden neun Subindizes: Schulschliessungen, Betriebsschliessungen, Absage öffentlicher Veranstaltungen, Versammlungseinschränkungen, reduzierter Betrieb des öffentlichen Verkehrs, Vorschriften, zu Hause zu bleiben, landesweite Ausgangsbeschränkungen, internationale Reisebeschränkungen und Informationskampagnen für die Bevölkerung. Neben diesen neun Subindizes berücksichtigt der «KOF Stringency-Plus Index» die Gesichtbedeckungen als zusätzlichen Indikator (Konjunkturforschungsstelle KOF, 2021). Da beispielsweise eine Maskentragepflicht in öffentlichen Verkehrsmitteln auf dem Weg an den Arbeitsplatz oder auch eine Pflicht zum Tragen der Mund-Nasen-Bedenkung am Arbeitsplatz selbst einen Einfluss auf die Entscheidung haben kann, ob die Person sich zur Arbeit im Homeoffice entscheidet, wird der «KOF Stringency Index Plus» als unabhängige Variable verwendet. Da die am Arbeitsort geltenden

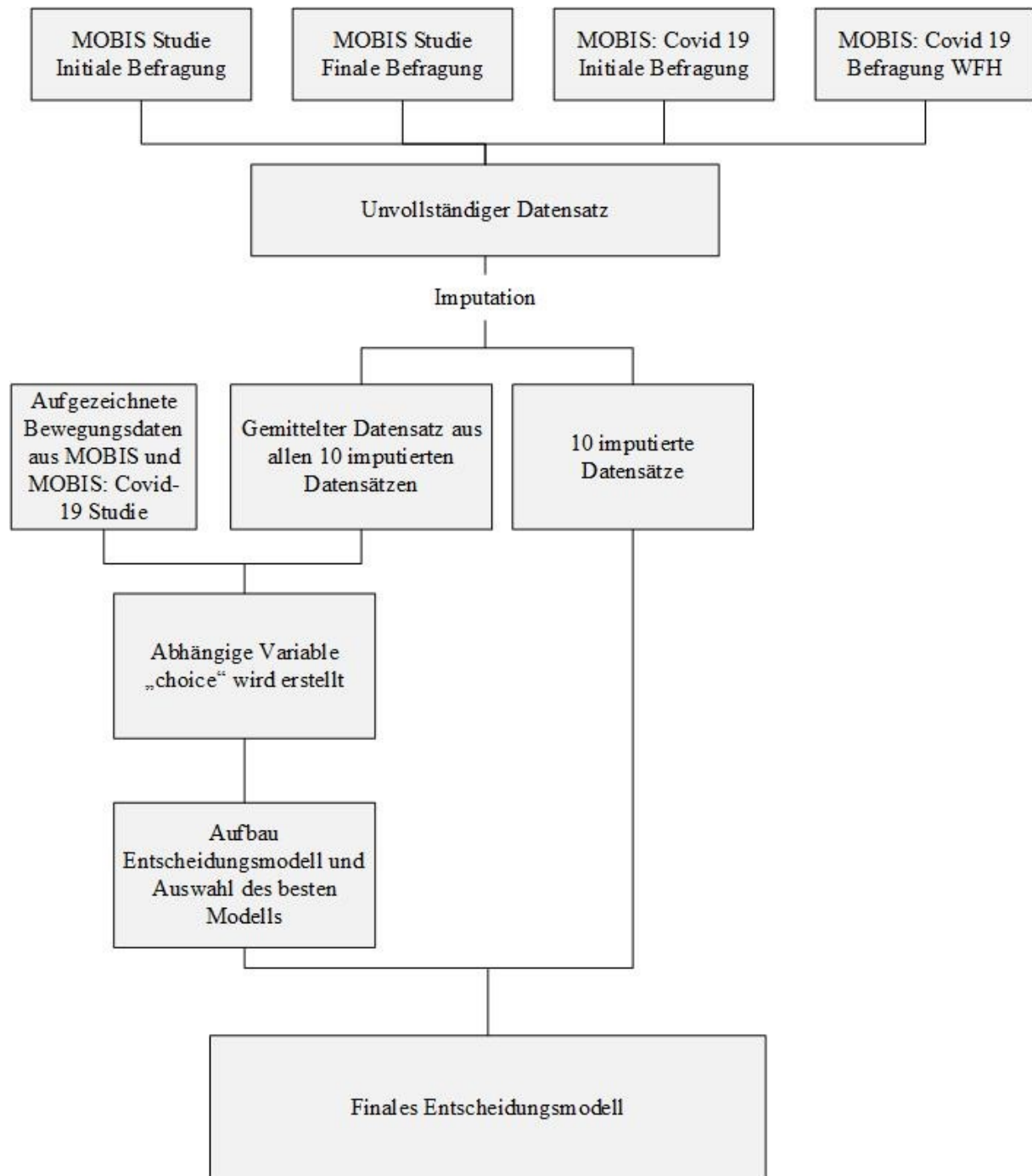
Restriktionen massgebend für den Arbeitsalltag sind, werden die Restriktionen immer auf den Kanton des Arbeitsortes bezogen.

Um den Verlauf des Risikos der COVID-19 Pandemie innerhalb der Schweizer Gesellschaft zu modellieren, wird eine weitere Variable eingeführt. Hierfür werden die Infektionszahlen innerhalb der Schweiz verwendet. Die Zahlen werden täglich vom Bundesamt für Gesundheit (BAG) veröffentlicht. Neben den täglich gemeldeten laborbestätigten Fällen werden auch die Inzidenzwerte für die letzten sieben Tage dokumentiert. Die Sieben-Tage Inzidenz gibt die Zahl der Neuinfektionen pro 100'000 Einwohner über die letzten sieben Tage an (Bundesamt für Gesundheit BAG, 2021a). Die Inzidenzwerte werden vom Bundesamt für Gesundheit sowohl für die gesamte Schweiz als auch für die einzelnen Kantone veröffentlicht. Für dieses Projekt werden die kantonalen Inzidenzwerte verwendet. Anders als beim Stringency Index werden die Inzidenzwerte auf den Kanton des Wohnortes der Person bezogen.

5.2.2 Auswahl der abhängigen Variable

Die abhängige «WFH»-Variable wird mit Hilfe der vorliegenden Daten modelliert. Diese liegt nicht direkt als Beobachtung vor. Aus den aufgezeichneten Bewegungen der Teilnehmer und den zugehörigen Angaben wird bestimmt, ob die Person aus dem Homeoffice gearbeitet hat oder nicht. Zur besseren Übersichtlichkeit, welche Daten verwendet wurden ist der Zusammenhang der einzelnen Datensätze in Abbildung 8 dargestellt. Nach der Imputation wird wie bereits beschrieben der gemittelte Datensatz erstellt. Zudem werden alle zehn imputierten Datensätze gespeichert. Die abhängige Variable wird mit den aufgezeichneten Bewegungsdaten und den Informationen über die jeweilige Arbeitsstation aus dem imputierten Datensatz erstellt. Für die Berechnung der abhängigen Variable und den anschliessenden Aufbau des Entscheidungsmodells wird zunächst der gemittelte Datensatz verwendet. Anschliessend werden die Berechnungen für die abhängige Variable für alle zehn imputierten Datensätze durchgeführt. Das finale Entscheidungsmodell wird ebenfalls mit den Daten aus allen zehn imputierten Datensätzen durchgeführt.

Abbildung 8 Übersicht über die verwendete Datensätze während des Projekts



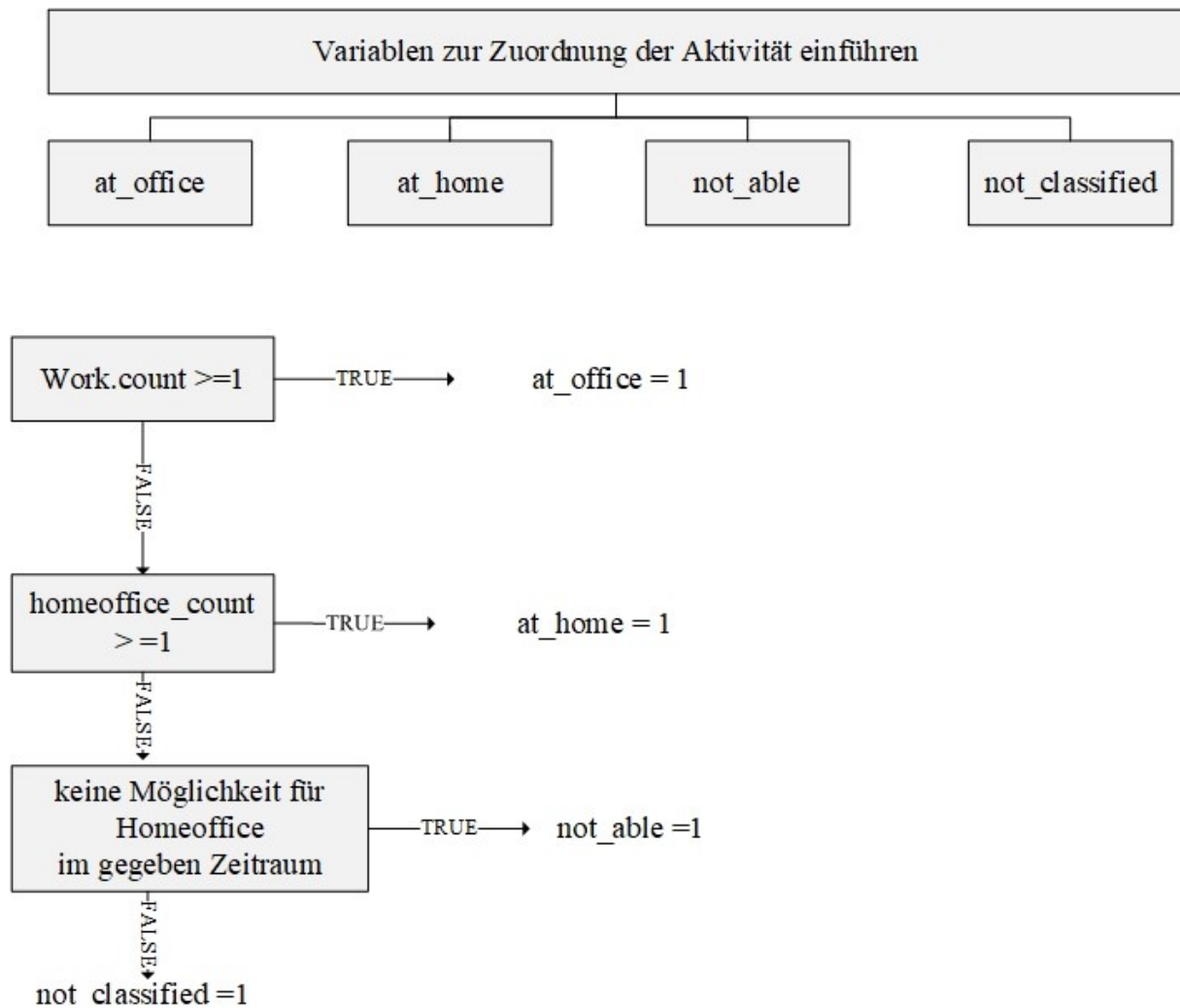
Zunächst werden drei Zeiträume definiert. Da die Erhebungen nur Momentaufnahmen darstellen, muss zunächst definiert werden, für welche Zeitspannen die Angaben übernommen werden. Der erste Zeitraum umfasst die Zeit vor der Corona-19-Pandemie in der Schweiz. Der erste Zeitraum endet am 16. März 2020. An diesem Tag erklärte der Bundesrat die «ausserordentliche Lage» und verschärfte die Massnahmen in der Schweiz (Bundesamt für Gesundheit BAG, 2020a). Zudem wird die Bevölkerung aufgerufen, alle unnötigen Kontakte zu vermeiden und Abstand zu halten. Für diesen Zeitraum werden die Daten aus den beiden Umfragen der MOBIS-Studie verwendet.

Die Daten der ersten Erhebung der MOBIS: COVID-Studie werden für den zweiten Zeitraum zwischen dem 17. März 2020 und dem 5. Juni 2020 übernommen. Der Bundesrat beschloss weitgehende Lockerungen aufgrund der positiven epidemiologischen Entwicklung per 6. Juni 2020. Die Empfehlungen zu Home-Office blieben jedoch bestehen (Bundesamt für Gesundheit BAG, 2020b). Der dritte Zeitraum umfasst dementsprechend alle erhobenen Daten ab dem 6. Juni 2020. Für diesen Zeitraum werden die Daten der zweiten Erhebung der MOBIS: COVID-19 Studie verwendet, die im Juli 2020 erhoben wurden.

Die Daten der Erhebungen, inklusive der imputierten Daten, werden anschliessend mit den Bewegungsdaten der einzelnen Tage verknüpft. Für jeden Tag werden anhand der definierten Zeiträume und der Daten aus den Umfragen neue Variablen bestimmt. Diese enthalten dann beispielsweise die Informationen über das Arbeitspensum zu diesem Zeitpunkt oder ob die Person zu diesem Zeitpunkt die Möglichkeit hatte im Homeoffice zu arbeiten.

Anschliessend werden aus den gesammelten Daten die Tage bestimmt, an denen die Person an ihrer Arbeitsstelle tätig war sowie die Tage, an denen die Person aus dem Homeoffice gearbeitet hat. Hierzu werden dem Datensatz zunächst vier binäre Variablen hinzugefügt. Die Zuweisung der Werte ist in Abbildung 9 veranschaulicht. In den Bewegungsdaten zeigt die Variable «work.count» an, wenn die Person an diesem Tag an ihrer Arbeitsstätte tätig war. Hierbei wird angenommen, dass die Aufzeichnung der Tage an denen die Person an ihrer Arbeitsstätte tätig war, vollständig sind, das heisst, dass jede Tätigkeit an der Arbeitsstätte aufgezeichnet wurde.

Abbildung 9 Vorgehensweise zur Zuordnung der Tätigkeit zu den aufgezeichneten Tagen



Die Aufzeichnungen über die Homeoffice Aktivitäten werden hingegen als unvollständig angenommen. Der Datensatz der Bewegungsdaten enthält ebenfalls eine Variable «homeoffice.count», die die Tage repräsentiert, an denen eine Aktivität im Homeoffice festgestellt wurde. Die Schlussfolgerung ist hier jedoch nur möglich, wenn der Teilnehmer die Tätigkeit selbst vermerkt hat, da das Arbeiten aus dem Homeoffice keinen Weg zum Arbeitsplatz beinhaltet, der in den Bewegungsdaten aufgezeichnet werden konnte. Dementsprechend wird angenommen, dass Teilnehmer an mehr Tagen im Homeoffice gearbeitet haben, als aufgezeichnet wurde. Aus diesem Grund werden alle aufgezeichneten Tage betrachtet und bestimmt, an welchen dieser Tage die Personen zusätzlich im Homeoffice gearbeitet haben.

Zunächst werden alle Tage ausgeschlossen, an denen es dem Teilnehmer nicht möglich war, im Homeoffice zu arbeiten. Die aufgezeichneten Tage der Personen, die keine Möglichkeit hatten im betroffenen Zeitraum im Homeoffice zu arbeiten, werden dementsprechend ausgeschlossen. Für die verbleibenden aufgezeichneten Tage wird anschliessend bestimmt, ob die Person im Homeoffice tätig war oder ob die Person nicht gearbeitet hat.

Dazu werden zunächst nur die Wochentage Montag bis Freitag und die Personen, die ihrer Beschäftigung regelmässig nachgehen (kein Schichtdienst) betrachtet. Die aufgezeichneten Daten werden für jede Kalenderwoche und Person gruppiert. Mit Hilfe des angegebenen Arbeitspensums wird die Anzahl der angenommenen Arbeitstage in der jeweiligen Woche berechnet. Anschliessend werden von den möglichen Arbeitstagen die schon zugeordneten Arbeitstage abgezogen. Die Differenz ergibt die Anzahl der Arbeitstage, die zusätzlich zu den aufgezeichneten Arbeitstagen im Homeoffice gearbeitet wurde.

In der Variablen «choice» wird anschliessend die Zuordnung des aufgezeichneten Tages festgehalten. Die Variable bekommt den Wert 1, wenn die Person an diesem Tag im Homeoffice gearbeitet hat. Den Wert 2 bekommt die Variable, wenn die Person an diesem Tag an ihrer Arbeitsstätte gearbeitet hat. Für alle Tage, an denen die Person nicht gearbeitet hat wird der Variable der Wert 3 zugewiesen. Diese Beobachtungen werden anschliessend für die Schätzung des Entscheidungsmodells ausgeschlossen.

5.3 Auswahl der Beobachtungen

Die Koeffizienten des Entscheidungsmodells werden auf Basis der Beobachtungen aus der MOBIS sowie der MOBIS: COVID-19 Studie geschätzt. Hierfür müssen die geeigneten Beobachtungen aus dem Datensatz ausgewählt werden. Eine Zeile des Datensatzes beinhaltet die aufgezeichneten Daten eines Teilnehmers an einem Tag. Der Datensatz mit allen aufgezeichneten Daten umfasst 454'796 Beobachtungen.

Für die Beobachtungen werden die aufgezeichneten Bewegungsdaten der Phase 3 bis Phase 7 verwendet. Zudem wird die Phase 1, die bereits während der MOBIS-Studie aufgezeichnet wurde, als Basis miteinbezogen. Die Phase 3 ist der Zeitraum zwischen den Aufzeichnungen der MOBIS-Studie und dem Start der COVID-19 Pandemie in der Schweiz. Mit Phase 4 beginnt die COVID-19 Pandemie in der Schweiz. Nach dem ersten Lockdown der Schweiz beginnt am 11.05.2020 die Phase 5. Ab dem 06.07.2020 gilt eine Maskentragepflicht in öffentlichen Verkehrsmitteln. Ab dann werden die Beobachtungen der Phase 6 zugeordnet. Der Start der zweiten Welle der COVID-19 Pandemie in der Schweiz markiert den Beginn der Phase 7.

Um die Wahrscheinlichkeit zu berechnen, ob eine Person aus dem Homeoffice arbeitet oder nicht, muss die Person einer Tätigkeit nachgehen. Daher werden in den Datensatz der Beobachtungen nur die Personen einbezogen, die bei der ersten Befragung angegeben haben, angestellt oder selbstständig zu sein. Zudem werden Teilnehmer, die sich in einer Ausbildung befinden in den Datensatz miteinbezogen. Studenten, Pensionäre und nicht erwerbstätige Personen werden nicht betrachtet.

Eine weitere wichtige Voraussetzung ist die Wahlmöglichkeit der Person. Es werden nur Beobachtungen von Teilnehmern verwendet, die die Möglichkeit zur Arbeit aus dem Homeoffice hatten. Ermöglicht die Tätigkeit oder der Arbeitsgeber beispielsweise keine Arbeit von zu Hause aus, so werden diese Beobachtungen nicht einbezogen.

Die Wahlentscheidung wurde nicht direkt beobachtet. Wie in Abschnitt 5.2.2 beschrieben, wurde die Wahlentscheidung anhand anderer beobachteter Werte bestimmt. Da diese Zuordnung nur für Personen mit regelmässiger Tätigkeit ausreichend genau getroffen werden konnte, werden nur Beobachtungen von Teilnehmern mit regelmässigen Arbeitszeiten in den Modelllaufbau miteinbezogen. Beobachtungen von Teilnehmern im Schichtbetrieb werden nicht miteinbezogen. Ausserdem werden nur Beobachtungen an Wochentagen (Montag-Freitag) betrachtet. Beobachtungen am Wochenende werden nicht berücksichtigt. Aufgrund der starken Abweichungen vom allgemeinen Arbeitspensum infolge der Feiertage werden zudem die Kalenderwoche 52 im Jahr 2019 und die Kalenderwoche 1 und 2 in 2020 nicht berücksichtigt.

Insgesamt können für die Schätzung des Datenmodells 20'403 Beobachtungen verwendet werden, die alle vorher genannten Kriterien erfüllen.

6 Aufbau des Choice Models

6.1 Deskriptive Datenanalyse

Die aufgezeichneten Bewegungsdaten der Teilnehmer wurden um weitere Daten aus den Befragungen der Teilnehmer ergänzt. Da die Befragungen nicht vollständig waren, wurden die Daten der Erhebungen, wie in Kapitel 4 beschrieben, imputiert. Insgesamt wurden zehn imputierte Datensätze erstellt. Zudem wurde ein Datensatz als Mittel der zehn Datensätze erstellt. Sowohl die folgende Analyse des Datensatzes des Entscheidungsmodells als auch der Aufbau des Choice Models erfolgen mit dem gemittelten Datensatz. Im Anschluss an die Auswahl des Choice Models werden die Parameter des Modells für alle zehn imputierten Datensätze geschätzt. Um das finale Modell zu erstellen werden anschliessend die geschätzten Koeffizienten der zehn imputierten Datensätze gemittelt.

Insgesamt wurden 454'796 Beobachtungen von 1'300 verschiedenen Teilnehmern während der MOBIS und der MOBIS: COVID-19 Studie dokumentiert. Von diesen Beobachtungen können für den Modellaufbau nach den gewählten Kriterien 39'359 Beobachtungen von 827 unterschiedlichen Teilnehmern genutzt werden. Tabelle 4 enthält die Angaben über Anzahl der Teilnehmer sowie die Anzahl der Teilnehmer, denen es grundsätzlich möglich war, im Homeoffice zu arbeiten. Die Angaben in Tabelle 4 beziehen sich auf die Angaben während der Befragungen. Da nur 1'300 Teilnehmer während beider Studien ihre Bewegungsarten aufgezeichnet haben, ist die Anzahl der Teilnehmer, deren Daten verwendet werden können, geringer als die Anzahl der ermittelten Teilnehmer in Tabelle 4.

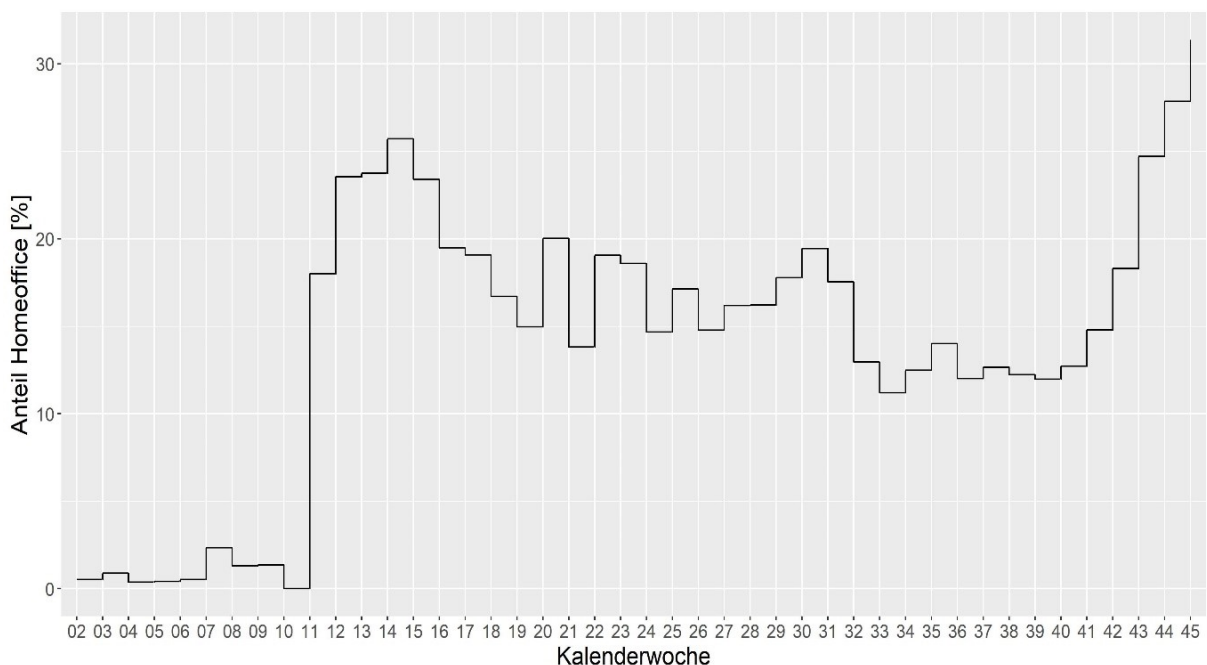
Tabelle 4 Anzahl der Erwerbstätigen mit Möglichkeiten im Homeoffice zur arbeiten in verschiedenen Zeiträumen

Tätigkeit	Gesamter Datensatz	WFH vor Corona bis 16. März 20	WFH 1 17. März 20 bis 5. Juni 20	WFH 2 ab 6. Juni 2020
Arbeitnehmer	2635	1110	1405	1404
Selbständige	232	102	100	85
Auszubildende	63	8	27	40
Total	2930	1220	1532	1529

Wie bereits in Abschnitt 5.2.2 beschrieben, wurden auf Basis der aufgezeichneten Bewegungsdaten und den zusätzlichen Angaben der Befragten die Tage bestimmt an denen ein Teilnehmer an der Tätigkeitsstätte gearbeitet hat. Zudem wurden die Tage, an denen der Teilnehmer im Homeoffice erwartet wurde mit den beschriebenen Berechnungen ermittelt.

Abbildung 10 zeigt den Anteil der Tätigkeiten im Homeoffice im Verlauf des Jahres 2020. Die aufgezeichneten Daten enden im November 2020. Während der durchschnittliche Anteil der Teilnehmer im Homeoffice zu Beginn des Jahres unter einem Prozent lag, ist in Kalenderwoche 11 (9.März 2020) ein starker Anstieg der Personen im Homeoffice bis zum vorläufigen Höchstpunkt in der Kalenderwoche 14 (bis 5.April 2020) erkennbar. In Kalenderwoche 14 liegt der Anteil bei 25 Prozent. Anschliessend fällt der Anteil der Homeoffice Tätigkeiten wieder auf einen durchschnittlichen Anteil von 15 Prozent ab. Ab der Kalenderwoche 40 (28.September 2020) lässt sich erneut ein starker Anstieg bis auf über 30 Prozent in Kalenderwoche 45 (bis 8. November 2020) erkennen. In Kalenderwoche 45 Enden die Aufzeichnungen.

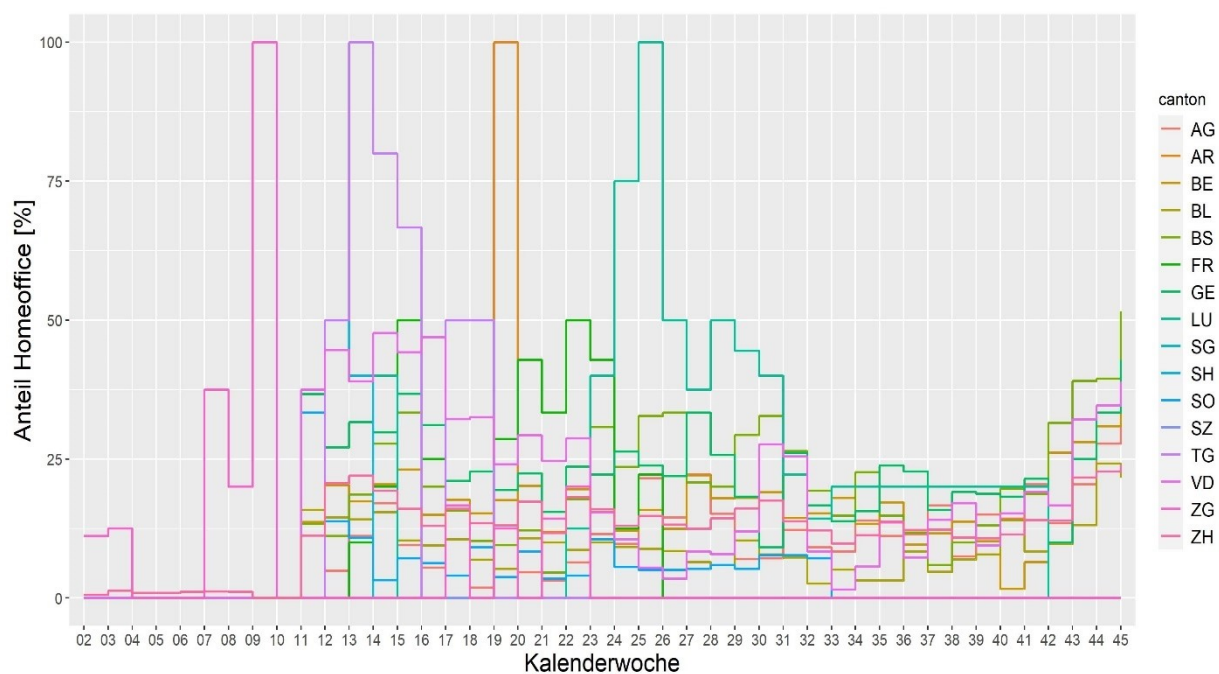
Abbildung 10 Anteil der Teilnehmer im Homeoffice über die Zeit



Während der Sommermonate sind Schwankungen im Anteil der Tätigkeiten aus dem Homeoffice zu erkennen. Neben tatsächlichen Änderungen der Anteile, könnten die Schwankungen auch durch falsch erfasste Homeoffice Tage während der Sommerferien begründet sein.

Zudem wird der Datensatz auf weitere Unterschiede, die von den bisher miteinbezogenen Variablen nicht erfasst werden, untersucht. Eine Möglichkeit könnten beispielsweise regionale Unterschiede sein. Dafür werden die Anteile der Teilnehmer im Homeoffice noch einmal für die Kantone getrennt dargestellt.

Abbildung 11 Anteil der Tätigkeiten im Homeoffice über die Zeit und nach Kanton



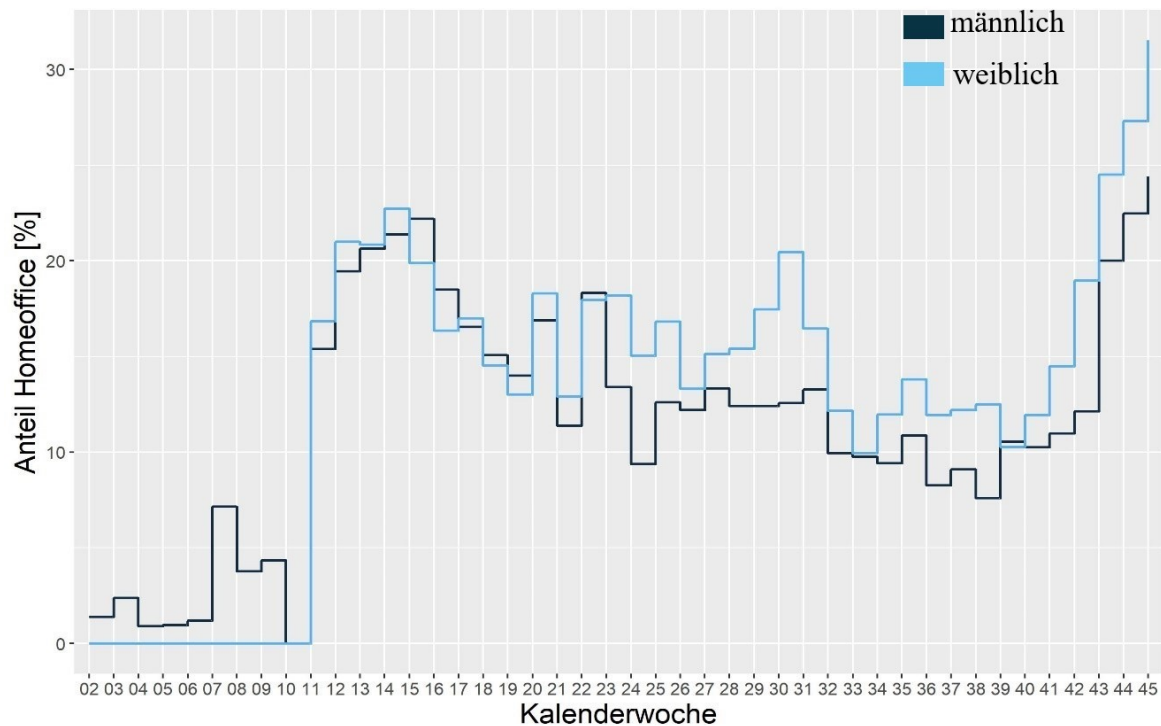
In Abbildung 11 werden die Anteile der Tätigkeiten im Homeoffice erneut über die Zeit dargestellt. Im Vergleich zu Abbildung 10 werden hier die Anteile jedoch getrennt nach Kanton dargestellt. In der Darstellung lassen sich einzelne Spitzen verschiedener Kantone erkennen. Beispielsweise wird für den Kanton Zug in Kalenderwoche 9 ein Anteil von 100 Prozent dargestellt. Vergleicht man die allerdings die Anzahl der Messwerte fällt auf, dass im Kanton Zug für die Kalenderwoche 9 nur zwei Beobachtungen dokumentiert sind. Während im Kanton Zürich 12'953 Beobachtungen dokumentiert sind, können im Kanton Schaffhausen nur 13 Beobachtungen verwendet werden. Eine Differenzierung nach Regionen ist aufgrund der grossen Unterschiede in der Anzahl der Beobachtungen daher nicht möglich.

Eine weitere Möglichkeit wäre die Berücksichtigung der Wochentage im Entscheidungsmodell. Dafür müsste untersucht werden, ob die Teilnehmer bestimmte Muster innerhalb einer Woche aufweisen. Beispielsweise könnte ein Teilnehmer immer montags an die Tätigkeitsstätte fahren, und mittwochs immer im Homeoffice arbeiten. Diese Muster könnten dann in das Entscheidungsmodell über weitere Variablen miteinbezogen werden. Da die Tage jedoch nicht direkt beobachtet wurden, sondern die Variable «choice» aus weiteren Variablen abgeleitet und über die gesamte Woche berechnet wurde, ist es nicht mehr möglich, nachzuvollziehen, an welchen tatsächlichen Wochentagen die Teilnehmer im Homeoffice gearbeitet haben. Daher ist die Berücksichtigung der Wochentage im Entscheidungsmodell nicht möglich.

In Kapitel 5.2.1 wurden bereits die unabhängigen Variablen beschrieben, die in das Entscheidungsmodell miteinbezogen werden sollen. Um Korrelationen der erklärenden Variablen mit der zu erklärenden Variablen zu untersuchen, werden zunächst die einzelnen Variablen der abhängigen Variablen gegenübergestellt und analysiert. Zudem werden mögliche Kollinearitäten zwischen den erklärenden Variablen ausgeschlossen.

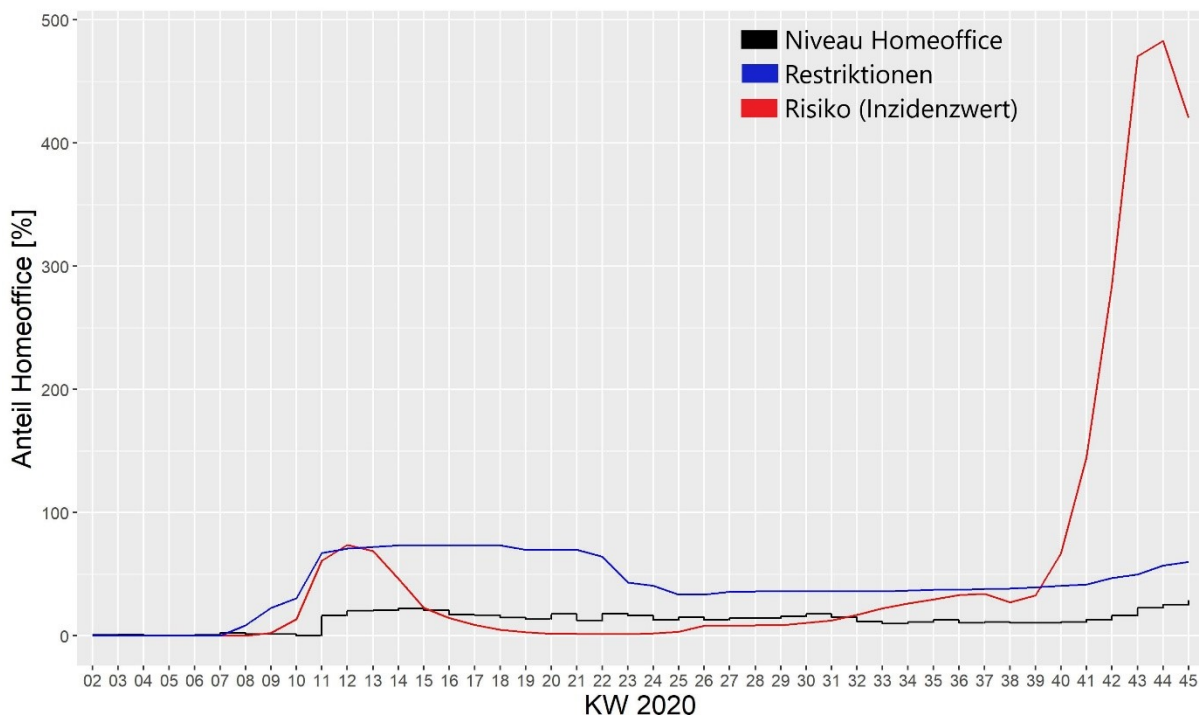
Als eine unabhängige Variable wurde das Geschlecht für den Aufbau des Entscheidungsmodells ausgewählt. In schwarz werden die Anteile der Tätigkeiten im Homeoffice der weiblichen Teilnehmer dargestellt, während die Anteile der männlichen Teilnehmer in blau dargestellt werden. Während bis Kalenderwoche 11 lediglich Frauen im Homeoffice tätig waren, steigt der Anteil der Männer ab Kalenderwoche 11 stark an. Während der folgenden Monate ist der Anteil der Männer, die im Homeoffice arbeiten, um durchschnittlich ein bis zwei Prozent höher als der Anteil der Frauen, die im Homeoffice arbeiten.

Abbildung 12 Anteil der Tätigkeiten im Homeoffice über die Zeit und Geschlecht



Während der COVID-19 Pandemie wurden vom Bundesrat und den einzelnen Kantonen Massnahmen zur Bekämpfung der Pandemie erlassen. Diese werden durch die Variable «stringency» repräsentiert. Die Variable «risk» stellt das Risiko, wie in Abschnitt 5.2.1 beschrieben, dar. Abbildung 13 stellt die Entwicklung dieser Variablen über die Zeit dar. In blau wird der Verlauf der Restriktionen dargestellt. Der rote Verlauf repräsentiert das Risiko. Die vorher bereits dargestellten Anteile der Tätigkeiten im Homeoffice, sind hier erneut in schwarz dargestellt. Vor Beginn der Pandemie im März 2020 sind Risiko wie auch Massnahmen nicht vorhanden. Auch der Anteil der Personen, die im Homeoffice tätig waren, lag nur zwischen null und zwei Prozent. Ab KW 9 (25. Februar 2020) steigt das Risiko. Zudem werden die ersten Massnahmen ergriffen und die Variable «stringency» steigt an. Während das Risiko schon ab KW 12 wieder sinkt, bleiben die Restriktionen weiterhin auf einem hohen Niveau. Auch die Anteile der Personen, die im Homeoffice arbeiten bleibt weiterhin auf einem erhöhten Niveau. In Kalenderwoche 23 (1. Juni 2020) sinken dann die Restriktionen. Auch der Anteil der Personen im Homeoffice sinkt. Ab Juli 2020 kann man einen Anstieg des Risikos erkennen, während sich das Niveau der Restriktionen kaum verändert. Erst ab Kalenderwoche 40 (28. September 2020) ist ein signifikanter Anstieg der Personen im Homeoffice zu erkennen. Gleichzeitig steigt das Risiko stark an. Auch die Restriktionen steigen schwach.

Abbildung 13 Verlauf des Risikos, der Restriktionen und der Anteil der Personen im Homeoffice über die Zeit



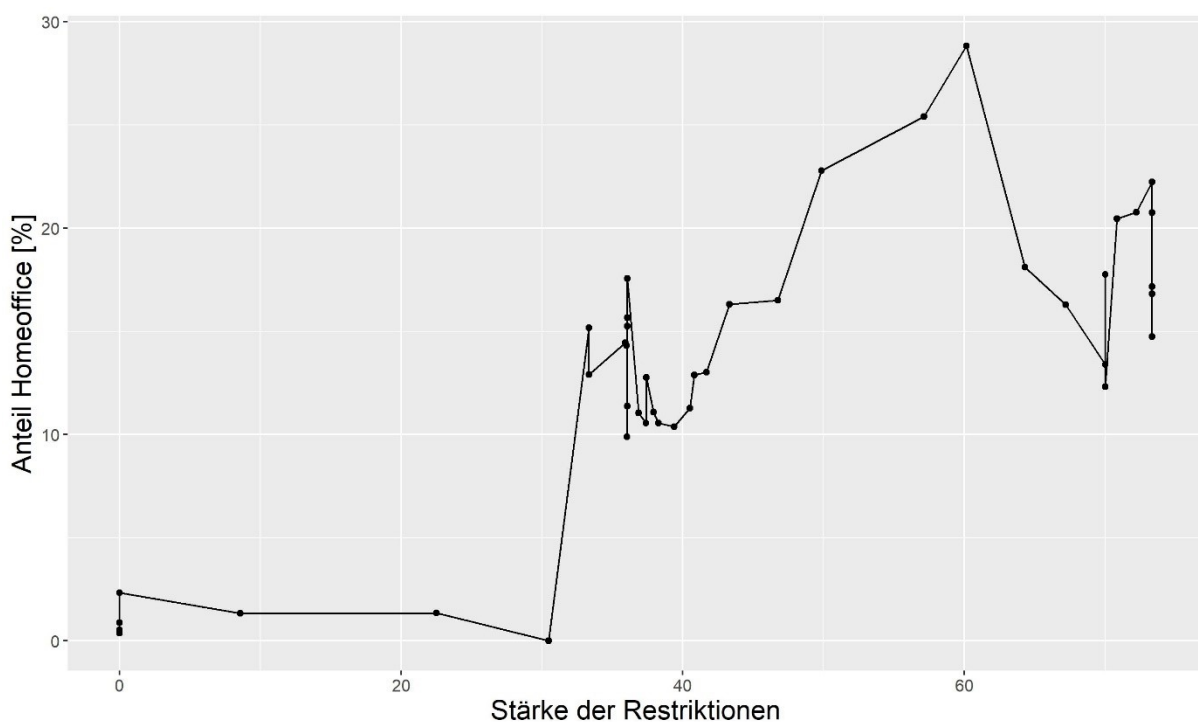
In Abbildung 14 wird zudem der Zusammenhang zwischen der Stärke der Restriktionen und dem Anteil der Personen im Homeoffice gezeigt. Die Restriktionen werden auf einer Skala zwischen 0 und 100 bewertet. Tendenziell steigt der Anteil der Personen im Homeoffice mit der Stärke der Restriktionen an. Allerdings zeigen sich auch Abweichungen von dieser Tendenz. Insbesondere bei Restriktionen über 60 sinkt der Anteil im Homeoffice wieder.

Dieser Verlauf könnte verschiedene Ursachen haben. Die Restriktionen werden durch den KOF Stringency Index repräsentiert. In diesem werden alle Massnahmen zusammengefasst und bewertet. Die Restriktionen hängen demnach von allen Massnahmen ab. Ob eine verschärfte Massnahme die Entscheidung zur Arbeit im Homeoffice betrifft, kann im Modell nicht erfasst werden. Daher ist es möglich, dass die Erhöhung der Restriktionen in diesem Fall keinen Einfluss auf die Entscheidung und damit auf den Anteil der Arbeitstage im Homeoffice hat.

Der Abfall des Homeoffice-Niveaus könnte dann durch andere Einflüsse im selben Zeitraum erklärt werden. Beispielsweise könnte das Risiko im beobachteten Zeitraum geringer gewesen sein oder geänderte persönliche Präferenzen könnten eine Ursache für den Rückgang der Homeoffice Anteile sein.

Zudem gab es im beobachteten Zeitraum lediglich eine Empfehlung für die Arbeit im Homeoffice. Eine Homeoffice Pflicht wurde erst nach Ende der aufgezeichneten Daten im Januar 2021 eingeführt. Möglicherweise haben Erwerbstätige ihre persönlichen Massnahmen angepasst um weiterhin eine gewisse soziale Interaktion beizubehalten. Wurden die Massnahmen im privaten Bereich verschärft, arbeiteten die Teilnehmer lieber an ihrem Arbeitsplatz um mit der sozialen Interaktion mit Arbeitskollegen die reduzierte soziale Interaktion im privaten Bereich auszugleichen.

Abbildung 14 Zusammenhang zwischen Restriktionen und Anteil im Homeoffice



6.2 Logit-Modell

Zunächst wird für den Aufbau des Entscheidungsmodells ein Logit-Modell verwendet, dem nacheinander die unabhängigen Variablen hinzugefügt werden. Die kategorialen Variablen werden in binär kodierte Dummy-Variablen aufgelöst. In Tabelle 5 sind die verwendeten Variablen für den Modellaufbau zusammengefasst. Die kursiv gedruckten Variablen, sind die jeweiligen Referenzvariablen. Die abhängige Variable «choice» kann die Werte 0 (Arbeit an der Tätigkeitsstätte) und 1 (Homeoffice) annehmen.

Tabelle 5 Verwendete Variablen für den Aufbau des Entscheidungsmodells

unabhängige Variable	im Model verwendete Variablen
Geschlecht	male
Alter	age1: < 20 Jahre age2: 20- 29 Jahre age3: 30-39 Jahre age4: 40-49 Jahre age5: 50-59 Jahre age6: ≥60 Jahre
Haushaltsgrösse	household1: Eine Person household2: 2 Personen household3: 3 Personen household4: Personen household5: 5 oder mehr Personen
Einkommen	income1: ≤4'000 CHF income2: 4'001-8'000CHF income3: 8'001-12'000CHF income4: 12'001-16'000CHF income5: > 16'000CHF
Bildungsgrad (Höchster Bildungsabschluss)	education1: Obligatorische Schule education2: Weiterführende Ausbildung education3: Universität, Fachhochschule
Risiko	risk: metrisch
Restriktionen	stringency: metrisch [0, 100]

Für den Aufbau des Entscheidungsmodells wird zunächst ein Logit-Modell verwendet. Die unabhängigen Variablen werden nacheinander dem Modell hinzugefügt und die Koeffizienten geschätzt. Nachdem das Modell geschätzt wurde, muss die Güte des Modells geprüft werden. Neben der globalen Prüfung des Modells werden auch die einzelnen Variablen auf ihre statistische Signifikanz überprüft. Zur Schätzung der logistischen Regression wird die Maximum-Likelihood-Methode verwendet. Daher ist ein möglicher Wert zur Überprüfung der Güte des Gesamtmodells der maximierte Log-Likelihood LL (Backhaus *et al.*, 2018). Auf diesem Wert basiert auch der wichtigste Test für die Prüfung der Güte der logistischen Regression. Der Likelihood-Ratio-Test (LLR) gibt die Differenz der Log-Likelihood für das Modell ohne Variablen (nur Konstante) und der Log-Likelihood des vollständigen Modells an.

$$LLR = -2 * (LL_0 - LL_V)$$

Je grösser die Differenz der beiden Werte, desto besser ist das Modell (Backhaus *et al.*, 2018). Mit dem Likelihood-Ratio-Test kann auch überprüft werden, ob ein Modell durch Hinzufügen einer weiteren Variablen eine signifikante Verbesserung erbringt. Dazu wird die Differenz des vollständigen und des erweiterten Modells berechnet. In diesem Fall wird die LLR genutzt, um zu bestimmen, ob das Hinzufügen einer weiteren Variablen sinnvoll ist. Dementsprechend werden die Variablen nacheinander dem Modell hinzugefügt und die Differenz zwischen dem alten und dem neuen Modell (mit der hinzugefügten Variablen) berechnet. In Tabelle 6 ist der Aufbau dokumentiert. In der Tabelle ist der mit dem LLR-Test berechnete p-Wert angegeben. Ist der p-Wert < 0.05 gilt die Verbesserung des Modells als statistisch signifikant (Backhaus *et al.*, 2018).

Tabelle 6 Aufbau Logit-Modell

Modell	inkludierte Variablen							p-Wert
	male	age	income	hhsz	edu	string.	risk	
1	x							
2	x	x						$1.05 \cdot 10^{-11}$
3	x	x	x					$2.2 \cdot 10^{-16}$
4	x	x	x	x				$2.2 \cdot 10^{-16}$
5	x	x	x	x	x			$3.24 \cdot 10^{-11}$
6	x	x	x	x	x	x		$2.2 \cdot 10^{-16}$
7	x	x	x	x	x	x	x	$2.2 \cdot 10^{-16}$

Tabelle 6 zeigt, dass mit jeder hinzugefügten Variable das Modell signifikant verbessert werden kann. Zunächst wurden alle soziodemografischen Variablen in kategoriale Variablen überführt. Die Variablen «stringency» und «risk» wurden als metrische Variablen hinzugefügt. Eine weitere Verbesserung des Modells könnte mit der Verwendung einer kategorialen Variablen, beziehungsweise Dummy-Variablen, für das Risiko erfolgen. Das Risiko wird in sechs Kategorien unterteilt. Anschliessend wird das Modell erneut aufgestellt und berechnet.

Tabelle 7 Kategorien für die Variablen «risk» und «stringency»

unabhängige Variable	im Model verwendete Variablen
Risiko	risk0: 0
	risk1: 1-30
	risk2: 31-60
	risk3: 61-100
	risk4: 101-200
	risk5: >200
Restriktionen	string0: 0
	string1: 1-25
	string2: 26-50
	string3: 51-75
	string4: 76-100

Ersetzt man die metrische Variable «risk» durch die Dummy-Variablen aus Tabelle 7 so kann man das Model noch einmal signifikant verbessern (p-Wert: $<2.2 \cdot 10^{-16}$). Weitere Anpassungsmöglichkeiten wurden überprüft und sind in Tabelle 8 zusammengefasst.

Tabelle 8 Weitere Anpassungen Logit Modell

Modell	Anpassung	p-Wert
8	metrische Variabel «risk» durch Dummy-Variablen (Risiko-Kategorien ersetzt)	$<2.2 \cdot 10^{-16}$
9	metrische Variabel «stringency» durch Dummy-Variablen	$9.7 \cdot 10^{-13}$
10	neue Variable: (stringency*risk) hinzugefügt	$2.2 \cdot 10^{-16}$

Tabelle 9 Schätzungen der Regressionskoeffizienten für das Logit Model Nr.10

unabhängige Variable	Variable in Model	Schätzung	p-Wert
Konstante	intercept	-5.30341	$<2.2 \cdot 10^{-16}$
Geschlecht	male	0.17895	$1.72 \cdot 10^{-6}$
Alter	age1	0.71580	0.016
	age2	0.57991	$1.00 \cdot 10^{-15}$
	age4	-0.04122	0.435
	age5	-0.09323	0.079
	age6	0.06011	0.36
Haushaltsgrösse	household2	0.45960	$2.91 \cdot 10^{-13}$
	household3	0.98743	$<2.2 \cdot 10^{-16}$
	household4	0.75541	$<2.2 \cdot 10^{-16}$
	household5	0.50717	$5 \cdot 10^{-9}$
Einkommen	income1	1.19277	$<2.2 \cdot 10^{-16}$
	income3	-0.01122	0.825
	income4	0.21975	$4.85 \cdot 10^{-5}$
	income5	0.10508	0.079
Bildungsgrad	education1	-0.04092	0.75
	education3	0.25506	$1.16 \cdot 10^{-11}$
Risiko	risk1	11.15131	0.913
	risk2	11.37245	0.911
	risk3	11.59617	0.910
	risk4	12.38824	0.904
	risk5	11.82755	0.894
Restriktionen	string1	1.88209	0.011
	string2	-9.02444	0.929
	string3	-8.47791	0.934
Restriktionen * Stringency	string1*risk1	-13.65341	0.89
	string2*risk1	0.34361	0.002
	string2*risk2	-0.21300	0.16
	string2*risk3	-0.03498	0.86
	string2*risk4	-0.87170	$1.00 \cdot 10^{-5}$

Für Model 10 sind die Schätzungen der Regressionskonstanten in Tabelle 9 zusammengefasst. Neben den Schätzungen sind auch die p-Werte der einzelnen Parameter angegeben. Eine Variable gilt als statistisch signifikant, wenn der p-Wert kleiner 0.05 ist.

Neben der Überprüfung der einzelnen Variablen muss die Güte des Gesamtmodells geprüft werden. Für die logistische Regression können Pseudo- R^2 Werte genutzt werden. Sie ähneln dem R^2 der linearen Regression. Die Pseudo R^2 können Werte zwischen 0 und 1 annehmen, wobei ein höherer Wert einer höheren Güte entspricht (Backhaus *et al.*, 2018). Für die Überprüfung des Gesamtmodells wird hier das McFadden R^2 verwendet.

$$McF R^2 = 1 - \frac{LL_v}{LL_0}$$

Im Unterschied zu LR-Statistik wird bei der Berechnung des R^2 der Quotient der Log-Likelihood gebildet. Ist der Unterschied der beiden Modelle nur gering, so ist der Quotient nahe 1 und das McFadden R^2 liegt nahe 0. Zur Anwendung des McFadden R^2 gilt als Faustregel, dass bereits Werte von 0.2 bis 0.4 eine gute Modellanpassung bedeuten (Backhaus *et al.*, 2018).

Für das oben beschriebene Logit-Modell ergibt sich ein McFadden R^2 von 0.1129.

6.3 Probit-Modell

Mit Hilfe des McFadden R^2 wurde die Modellgüte des Logit-Modells überprüft. Ein R^2 von 0.1129 weist auf eine ungenügende Modellgüte hin. Daher ist eine weitere Möglichkeit die Verwendung des Probit-Modells für den Aufbau des Entscheidungsmodells.

Der Aufbau des Probit- Modells erfolgt analog dem vorherigen Vorgehen für das Logit-Modell. Auch hier kann die Güte des Gesamtmodells mit dem McFadden R^2 beurteilt werden.

Für das Probit-Modell, das die gleichen Variablen und die gleiche Ausgangsfunktion wie das vorher ausgewählte Logit-Model beinhaltet, wird das McFadden R^2 zu 0.1124 berechnet. Eine Verbesserung der Modellgüte ist durch das Probit-Modell dementsprechend nicht zu erreichen. Weitere Anpassungen im Probit-Modell werden daher nicht untersucht.

6.4 Mixed Logit-Modell

Eine weitere Möglichkeit ist die Anwendung sogenannter Mixed-Logit Modelle. Diese bilden eine Erweiterung der grundlegenden Logit-Modelle und können Panel-Daten berücksichtigen. Als Paneldaten werden Daten bezeichnet, die aus mehrmaligen Befragungen derselben Person resultieren. Wird eine Person mehrfach zu einem Sachverhalt befragt oder werden Stated-Choice Untersuchungen für eine Person mehrfach durchgeführt, liegen Panel-Befragungen vor. Für dieses Projekt wurden über einen längeren Zeitraum immer wieder dieselben Personen beobachtet, es liegen dementsprechend Panel-Daten vor.

Im grundlegenden Logit-Modell wird jede Beobachtung als einzelne Beobachtung gewertet. Das bedeutet, dass jede Beobachtung so gewertet wird, als wäre sie von einer anderen Person. Korrelationsanteile, die von nicht beobachtbaren Präferenzen der Person ausgehen, können im Logit-Modell nicht erfasst werden. Im Mixed Logit-Modell können individuelle, unbeobachtete Effekte als zufällig Effekte erfasst werden. Mixed Logit-Modelle können dementsprechend Effekte, die zwischen den einzelnen Personen variieren und zusätzlich beobachtbare Variablen für die einzelnen Personen schätzen.

Für den Aufbau des Mixed Logit-Modells wird das vorhergehende Logit-Modell verwendet. Dieses beinhaltet alle «fixed effects». Zusätzlich wird dem Modell ein Faktor hinzugefügt, der die individuellen, nicht beobachtbaren Effekte, sogenannte «random effects», für die einzelnen Teilnehmer erfasst.

Die Schätzung des Mixed Logit-Modells erfolgt in der Statistiksoftware R mit Hilfe des Pakets «lme4». Die geschätzten Koeffizienten sind in Tabelle 10 zusammengefasst.

Vergleicht man die einzelnen Regressionsparameter mit den Regressionsparametern des Logit-Modells, fällt auf, dass im Mixed Logit-Modell deutlich weniger Parameter statistisch signifikant sind. Die nicht beobachtbaren Effekte haben dementsprechend einen starken Einfluss auf das Modell. Die persönlichen Präferenzen der einzelnen Teilnehmer haben also einen stärkeren Einfluss auf die Entscheidung, als die allgemeinen soziodemographischen Merkmale.

Tabelle 10 Schätzungen der Regressionskoeffizienten für das Mixed Logit-Modell

unabhängige Variable	Variable in Model	Schätzung	p-Wert
Konstante	intercept	-11.950519	$<2.2 \cdot 10^{-16}$
Geschlecht	male	-0.110153	0.949854
Alter	age1	-0.493792	0.877331
	age2	0.341528	0.756330
	age4	-0.395568	0.627151
	age5	-0.989207	0.307269
	age6	0.290069	0.924106
Haushaltsgrösse	household2	0.800123	0.494340
	household3	2.146613	0.043599
	household4	1.923199	0.090708
	household5	1.953060	0.249526
Einkommen	income1	4.207228	0.029024
	income3	0.331263	0.608677
	income4	2.269483	0.008641
	income5	1.275450	0.167160
Bildungsgrad	education1	0.126812	0.873172
	education3	-0.003201	0.988568
Risiko	risk1	18.36214	0.002208
	risk2	19.083684	0.000499
	risk3	19.470350	0.000529
	risk4	19.612150	0.000391
	risk5	19.598811	$2.75 \cdot 10^{-5}$
Restriktionen	string1	-0.168699	0.051946
	string2	-15.689315	0.513898
	string3	-14.821259	0.152077
Restriktionen* Risiko	string1*risk1	-18.482557	0.30358
	string2*risk1	0.538450	0.00446
	string2*risk2	-0.494215	0.04664
	string2*risk3	-0.651154	0.05416
	string2*risk4	-0.783391	0.03133

6.5 Finales Modell

Das Akaike-Informationskriterium (AIC) ist ein Schätzer für Vorhersagefehler und damit für die relative Qualität statistischer Modelle für einen bestimmten Datensatz. Für dieses Project kann das AIC verwendet werden, um das Logit-Modell mit dem Mixed Logit-Modell zu vergleichen. Das AIC des Logit-Modells beträgt 24'406, während das AIC des Mixed Logit-Modells nur 11'720 beträgt. Das Mixed Logit-Modell mit dem kleineren AIC ist das Modell mit der höheren Güte.

Vergleicht man allerdings die statistische Signifikanz der Parameter in beiden Modellen, fällt auf, dass die Variablen im Mixed Logit-Modell grösstenteils nicht statistisch signifikant sind, wohingegen die Variablen im Logit-Modell eine Auswirkung auf die Entscheidung haben. Für die Untersuchung von Auswirkungen verschiedener Faktoren ist daher das Logit-Modell geeigneter. Aufgrund der fehlenden statistischen Signifikanz der unabhängigen Variablen im Mixed Logit-Modell, könnte man die Auswirkungen von veränderten Variablen schlechter untersuchen als mit dem Logit-Modell.

Der Vergleich des Mixed Logit-Modells mit dem einfachen Logit-Modell zeigt, dass die persönlichen nicht beobachtbaren Effekte einen starken Einfluss auf die Entscheidung haben. Allerdings konnten diese Effekte möglicherweise nur beim Aufbau dieses Modells nicht beobachtet werden. Werden zusätzliche Variablen berücksichtigt, könnten persönliche Präferenzen erklärt werden. Das Logit-Modell bietet daher auch das Potenzial unter Berücksichtigung weiterer Variablen weiter entwickelt zu werden.

Aus diesem Grund wird für dieses Projekt das Logit-Modell als finales Modell gewählt. Trotzdem wird zum Vergleich auch für das Mixed Logit-Modell ein Modell erstellt, das auf allen zehn imputierten Datensätzen aufbaut. Der Code sowie das geschätzte Modell sind im digitalen Anhang zu finden.

Es muss jedoch bedacht werden, dass dementsprechend die persönlichen Präferenzen nicht berücksichtigt werden, obwohl diese einen starken Einfluss auf das Modell haben.

Für das finale Modell werden die einzelnen Datensätze der zehn Imputationen verwendet. Mit Hilfe der zehn Datensätze wird das Modell erneut geschätzt. Die Parameter sind in Tabelle 11 dokumentiert.

Für das finale Logit-Modell ergibt sich ein McFadden R^2 von 0.1185.

Tabelle 11 Schätzungen der Regressionskoeffizienten für das finale Modell

unabhängige Variable	Variable in Model	Schätzung	p-Wert
Konstante	constant	-5.34153	$<2 \cdot 10^{-16}$
Geschlecht	male	-0.03649	0.001142
Alter	age1	-0.81797	$2.76 \cdot 10^{-7}$
	age2	0.19453	$1.06 \cdot 10^{-13}$
	age4	0.11782	$7.14 \cdot 10^{-13}$
	age5	-0.03973	0.0158
	age6	0.32348	$<2 \cdot 10^{-16}$
Haushaltsgrösse	household2	0.27074	$<2 \cdot 10^{-16}$
	household3	0.89099	$<2 \cdot 10^{-16}$
	household4	0.68636	$<2 \cdot 10^{-16}$
	household5	0.20231	$2.26 \cdot 10^{-13}$
Einkommen	income1	0.94722	$<2 \cdot 10^{-16}$
	income3	0.18459	$<2 \cdot 10^{-16}$
	income4	0.26885	$<2 \cdot 10^{-16}$
	income5	0.28311	$<2 \cdot 10^{-16}$
Bildungsgrad	education1	0.06760	0.051578
	education3	0.39706	$<2 \cdot 10^{-16}$
Risiko	risk1	0.63320	0.009622
	risk2	0.87746	0.000342
	risk3	1.19611	$1.30 \cdot 10^{-6}$
	risk4	1.76761	$8.54 \cdot 10^{-13}$
	risk5	1.26996	$2.24 \cdot 10^{-7}$
Restriktionen	string1	1.71735	$2.37 \cdot 10^{-13}$
	string2	1.59610	$1.17 \cdot 10^{-10}$
	string3	2.04322	$<2 \cdot 10^{-16}$
Restriktionen * Stringency	string1*risk1	-2.31076	$6.43 \cdot 10^{-14}$
	string2*risk1	0.25286	$<2 \cdot 10^{-16}$
	string2*risk2	-0.16212	0.161054
	string2*risk3	-0.22883	0.001083
	string2*risk4	-0.81057	$<2 \cdot 10^{-16}$

6.6 Interpretation der Parameter

Insbesondere die Vorzeichen der einzelnen Regressionsparameter geben eine Information darüber, welchen Einfluss die Variable auf die Entscheidung nimmt. Die binäre Entscheidungsvariable «choice» kann die Werte 0 und 1 annehmen. Der Wert 0 bedeutet, dass die Person an der Arbeitsstelle war, wohingegen der Wert 1 die Arbeit aus dem Homeoffice repräsentiert.

Ist das Vorzeichen des Koeffizienten negativ, erhöht sich mit dieser Variablen die Wahrscheinlichkeit, dass die Person ihre Arbeit an der Arbeitsstelle verrichtet. Positive Vorzeichen bedeuten, dass die Variabel die Wahrscheinlichkeit für die Arbeit aus dem Homeoffice erhöht.

Die Referenz für das Entscheidungsmodell ist eine weibliche Person, im Alter zwischen 30 und 39 Jahren mit einem monatlichen Einkommen zwischen 4'001 und 8'000 CHF, welche eine weiterführende Ausbildung genossen hat und in einem Einzelhaushalt lebt. Des Weiteren beträgt das Risiko durch die Covid-19 Pandemie null und es sind keine Einschränkungen in Form von Massnahmen aktiv.

Die negative Konstante des Modells besagt, dass die Wahrscheinlichkeit für die Arbeit an der Tätigkeitsstätte höher ist als die Wahrscheinlichkeit für die Arbeit im Homeoffice. Des Weiteren sinkt die Wahrscheinlichkeit für die Verrichtung der Arbeit im Homeoffice bei männlichen Personen. Während Personen, die jünger als 20 Jahre sind, eher an der Arbeitsstätte ihre Arbeit verrichten, steigt die Wahrscheinlichkeit für die Wahl des Homeoffice bei älteren Personen. Da insbesondere jüngere Erwerbstätige noch in der Ausbildung sind oder zumindest Anleitungen benötigen, könnte die Wahl für die Arbeit am primären Arbeitsplatz begründen. Zudem ist das Risiko durch eine Covid-19 Erkrankung für ältere Personen höher, wodurch sich diese stärker schützen möchten.

Steigt die Anzahl der im Haushaltlebenden Personen, steigt auch die Wahrscheinlichkeit im Homeoffice zu arbeiten. Insbesondere sind die Koeffizienten für die Haushalte mit drei und vier Personen höher, als die Koeffizienten der anderen Haushalte. Da beispielsweise Kinder zusätzlich während der Covid-19 Pandemie durch geschlossene Schulen und Kindergärten zu Hause von den Eltern betreut werden mussten.

Auch mit steigenden Restriktionen und steigenden Risiken, steigt die Wahrscheinlichkeit für die Entscheidung zum Homeoffice. Allerdings fällt hierbei auf, dass die Wahrscheinlichkeit bei einem Risiko mit einer Inzidenz zwischen 100 und 200 höher ist, als bei einer Inzidenz mit mehr als 200 Neuinfektionen.

6.7 Beschränkungen des Modells und weitere Anpassungsmöglichkeiten

Mit Hilfe der Pseudo R^2 Werte lässt sich die Gesamtgüte des Modells bemessen. Für das verwendete McFadden R^2 gilt als Faustregel, dass bereits Werte von 0.2 bis 0.4 eine gute Modelanpassung bedeuten. Für das finale Modell dieses Projektes liegt der Wert mit 0.1185 unterhalb der gewünschten Grenze. Dass das Modell keine gute Anpassungsgüte besitzt, kann unterschiedliche Gründe haben.

Die in diesem Projekt verwendete Datenbasis ist unvollständig. Da zu Beginn nur 43 Prozent der Daten im Datensatz beobachtet wurden, musste ein grosser Teil der Daten imputiert werden. Je mehr Daten imputiert werden müssen, desto ungenauer ist das Ergebnis der Imputation. Zum Vergleich wurde das Logit-Modell auch mit dem Datensatz, der ausschliesslich beobachtete und keine imputierten Daten enthält, berechnet. Die Tage, die die Teilnehmer im Homeoffice waren, wurde auch hierbei mit dem beschriebenen Vorgehen ermittelt. Das McFadden R^2 wird bei diesem Modell zu 0.1321 berechnet.

Zudem stellen die Erhebungen, insbesondere die Erhebungen während der MOBIS: COVID-19 Studie nur Momentaufnahmen dar. Die Zeiträume, für die die Erhebungen übernommen wurden, wurden in diesem Projekt festgelegt. Allerdings können die Zeiträume, in denen die Teilnehmer beispielsweise in Kurzarbeit waren, variieren und müssen nicht mit den festgelegten Zeiträumen übereinstimmen. Um die Zeiträume genauer zu erfassen, wären weitere rückwirkende Erhebungen der Teilnehmer notwendig. Bei diesen Befragungen müssten die genauen Zeiträume, in denen Teilnehmende beispielsweise im Homeoffice oder in Kurzarbeit waren, erfasst werden.

Eine weitere Schwierigkeit beim Aufbau des Entscheidungsmodells war die beobachtete Entscheidung. Während die Tage, an denen der Teilnehmer an der Arbeitsstätte tätig war, durch Bewegungsaufzeichnungen genau definiert werden konnten, ist dies bei denen Tagen, an denen der Teilnehmer im Homeoffice gearbeitet hat, nicht ohne weiteres möglich. Da die Tätigkeit im Homeoffice keine Bewegung aufzeichnet, sind die Tage schwieriger nachzuvollziehen. Auch wenn eine Person sich den ganzen Tag zu Hause aufgehalten hat, ist dies kein Nachweis dafür, dass die Person von zu Hause aus auch gearbeitet hat.

In diesem Projekt wurde deshalb ein Vorgehen entwickelt, welches die Tage im Homeoffice bestimmt. Dafür wurden das Arbeitspensum, eine eventuelle Kurzarbeit und weitere Angaben berücksichtigt. Mit Hilfe der beobachteten Tage, an denen der Teilnehmer an der

Tätigkeitsstätte gearbeitet hat, und der Anzahl der erwarteten Tage, die der Teilnehmer in der Woche insgesamt gearbeitet hat, wurden anschliessend die Anzahl der Tage im Homeoffice zugewiesen. Um die Genauigkeit dieses Vorgehens zu erhöhen, wurden nur Teilnehmer mit regelmässigen Arbeitszeiten betrachtet. Schichtarbeiter und Beobachtungen an Wochenenden wurden ausgeschlossen. Andere beeinflussende Faktoren konnten aufgrund der vorhandenen Daten nicht berücksichtigt werden. Beispielsweise wurden Urlaubstage der Teilnehmer nicht erfasst. Daher können Tage fälschlicherweise als Tage im Homeoffice definiert werden, obwohl die Person an diesem Tag Urlaub hatte und gar nicht gearbeitet hat. Insbesondere in den Sommermonaten ist dies möglich.

Zudem wurde die Annahme getroffen, dass die Teilnehmer immer einen Tag vollständig oder gar nicht gearbeitet haben. Auch hier konnte aufgrund der vorhandenen Daten keine geeignetere Annahme getroffen werden. Dementsprechend wurden beispielsweise für Personen, die ein Arbeitspensum von 60 Prozent angegeben haben, drei Arbeitstage pro Woche angenommen. Allerdings kann es sein, dass diese Personen statt an drei Tagen jeweils 8 Stunden an vier Tagen jeweils 6 Stunden arbeiten. Dementsprechend wurden bei diesen Personen zu wenige Homeoffice Tage ermittelt werden.

In den beiden Erhebungen der MOBIS: COVID-19 Studie wurde auch nach der Anzahl der Tage, die die Person in der letzten Woche im Homeoffice gearbeitet hat, gefragt. Insbesondere bei Unternehmen, die während der Pandemie mit Wechselmodellen gearbeitet haben, führen auch diese Angaben zu Verzerrungen.

Mit verbesserten Daten, durch beispielsweise eine weitere rückwirkende Erfassung von Daten, könnten Unsicherheiten in der Datengrundlage verringert werden. Zudem könnten weitere Variablen zur Verbesserung des Modells genutzt werden.

Die persönlichen, nicht beobachtbaren Effekte könnten verringert werden, indem weitere Variablen miteinbezogen werden. Zusätzliche Variablen könnten persönliche Präferenzen möglicherweise erklären. Beispielsweise das Tätigkeitsfeld könnte in das Entscheidungsmodell integriert werden.

Eine mögliche Berücksichtigung der Branche der Erwerbstätigen sowie der Art der Tätigkeit könnten im Modell erfasst werden, um zu untersuchen, ob gewisse Berufsfelder eine höhere Tendenz zur Arbeit im Homeoffice aufweisen. Wenn die Homeoffice Tage direkt beobachtet werden würden, könnten auch die Wochentage im Entscheidungsmodell berücksichtigt werden. Damit könnte untersucht werden, ob beispielsweise an Montagen weniger Personen im

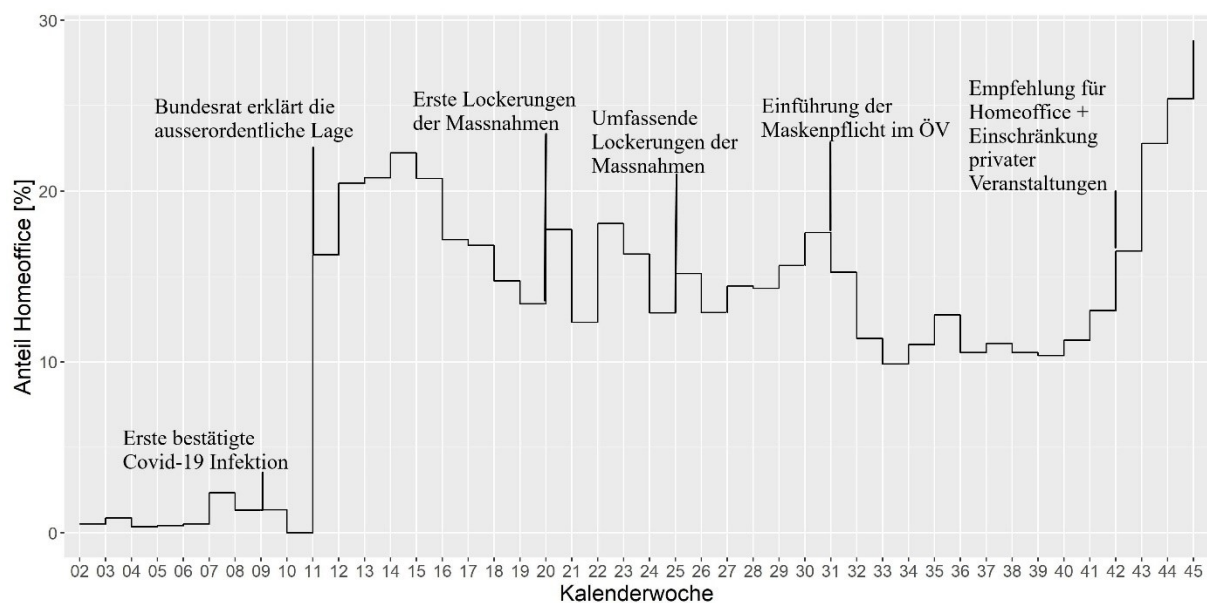
Homeoffice gearbeitet haben als an Freitagen. Insbesondere für die Verwendung des Entscheidungsmodells nach dem Ende der Covid-19- Pandemie, wäre die Berücksichtigung einer Variablen der Distanz vom Wohnort bis zur Arbeitsstelle sinnvoll, da hiermit auch die Distanz und die damit verbundene eingesparte Zeit im Falle von Homeoffice berücksichtigt werden kann.

7 Diskussion der Ergebnisse

7.1 Anteile Homeoffice 2020

In Abbildung 15 werden noch einmal die Anteile der Erwerbstätigen im Homeoffice über die Zeit dargestellt. Zudem beinhaltet die Abbildung die wichtigsten Ereignisse während der Covid-19 Pandemie im Jahr 2020.

Abbildung 15 Anteil der Erwerbstätigen im Homeoffice über die Zeit



Im Verlauf der Covid-19 Pandemie kann man grosse Unterschiede der Erwerbstätigen im Homeoffice erkennen. Zu Beginn der Pandemie steigt der Anteil der Erwerbstätigen im Homeoffice stark an. Während die ersten Wochen der Pandemie der Homeoffice Anteil ansteigt, fällt er ab Kalenderwoche 16 ab. Bemerkenswert ist dabei, dass die Massnahmen erst ab Kalenderwoche 20 gelockert werden. Eine mögliche Begründung könnte die anfängliche Unsicherheit über das neuartige Virus gewesen sein. Die unvollständige Informationslage hat zunächst zu mehr Vorsicht geführt. Nach einer ersten Zeit konnten sowohl Erwerbstätige als auch Arbeitgeber die Lage besser bewerten. Dies könnte zu einem Rückgang der Personen im Homeoffice geführt haben.

Trotz umfassender Lockerungen im Juni 2020, kehrt das Homeoffice Niveau nicht auf das Niveau vor der Covid-19 Pandemie zurück. Hier könnten mehrere Gründe ausschlaggebend sein. Zum einen könnten Erwerbstätige und Unternehmen weiterhin vorsichtig sein und die Arbeit aus dem Homeoffice zur Sicherheit beibehalten. Zum anderen könnten Erwerbstätige die Möglichkeiten des Homeoffice erkannt haben und wollen deren Vorteile weiterhin nutzen. Das würde dafürsprechen, dass auch nach der Covid-19-Pandemie der Anteil der Erwerbstätigen, die im Homeoffice arbeiten, mutmasslich höher sein wird als vor der Pandemie. Während der Sommermonate haben viele Erwerbstätige zudem Urlaub. Da Urlaubstage in der Datengrundlage nicht erfasst wurden, könnten auch Homeoffice Tage falsch erfasst worden sein, was zu einem erhöhten Level führen würde.

Zu Beginn der zweiten Welle im Oktober 2020 werden in Kalenderwoche 42 erneut Massnahmen beschlossen. Neben der Einschränkung von privaten Veranstaltungen spricht der Bundesrat eine Homeoffice-Empfehlung aus. Auch der Anteil der Erwerbstätigen im Homeoffice steigt wieder an.

7.2 Entwicklungen ab November 2020

Die Aufzeichnung der Daten, die für dieses Projekt verwendet wurden, endet im November 2020. Über diesen Zeitpunkt hinausgehende Daten wurden für dieses Projekt nicht verwendet und der Untersuchungszeitraum endet im November 2020. Zu diesem Zeitpunkt stiegen die Infektionszahlen wieder und die zweite Welle der Covid-19 Pandemie begann.

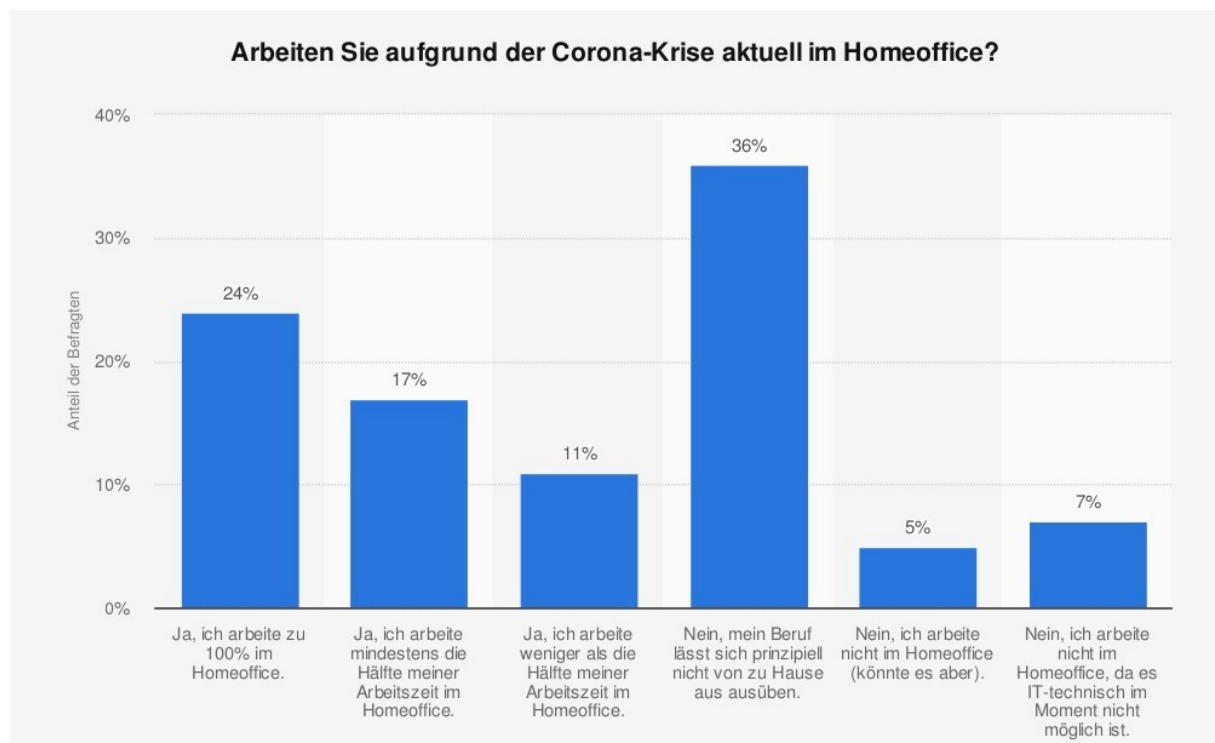
Als weitere Massnahme zur Eindämmung der Covid-19-Pandemie wurde vom Bundesrat ab dem 18. Januar 2021 eine Homeoffice-Pflicht beschlossen. Mit der Homeoffice-Pflicht sind Arbeitgeber verpflichtet, Homeoffice «überall dort anzuordnen, wo dies aufgrund der Art der Aktivität möglich und mit verhältnismässigem Aufwand umsetzbar ist.» (Bundesamt für Gesundheit BAG, 2021b) Zuvor hatte es nur Empfehlungen zum Homeoffice gegeben.

Bei einer Umfrage von Deloitte im Februar 2021 wurden 1'699 Erwerbstätige befragt, ob sie aufgrund der Corona-Krise aktuell im Homeoffice arbeiten. 24 Prozent der Befragten gaben an, dass sie 100 Prozent ihrer Arbeitszeit im Homeoffice verbringen. Insgesamt 41 Prozent der Erwerbstätigen verbringen mindestens die Hälfte ihrer Arbeitszeit im Homeoffice (Statista, 2021b) . In Abbildung 16 sind die Ergebnisse der Umfrage zusammengefasst.

Vergleicht man den Anteil der Erwerbstätigen, ist zu erkennen, dass der Anteil der Menschen, die ihre Arbeitszeit grösstenteils im Homeoffice verbringen höher ist, als die ermittelten Anteile

dieses Projektes für das Jahr 2020. Unter anderem lässt sich dieser Anstieg mit der Homeoffice-Pflicht erklären.

Abbildung 16 Ergebnisse zur Umfrage Arbeit im Homeoffice im Februar 2021



Quelle: Statista, 2021b

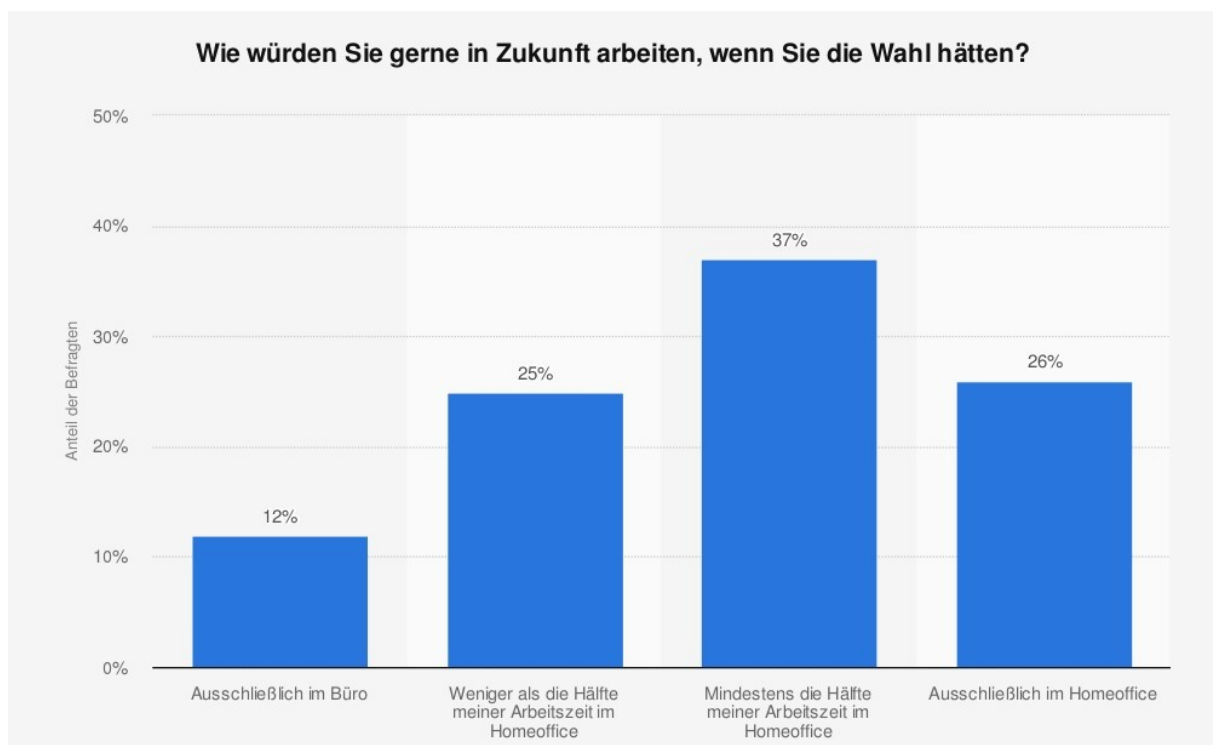
7.3 Prognosen Homeoffice

7.3.1 Homeoffice in der Zukunft

Erst nach dem Ende der Covid-19 Pandemie wird sich herausstellen, inwiefern die Erfahrungen mit der Arbeit aus dem Homeoffice Auswirkungen auf den dauerhaften Anteil der Erwerbstätigen im Homeoffice haben werden. Der erhöhte Anteil der Erwerbstätigen, die im Sommer trotz der Lockerungen weiterhin ihre Arbeitszeit zumindest teilweise im Homeoffice verbracht haben, lässt die Schlussfolgerung zu, dass Erwerbstätige die Möglichkeit im Homeoffice zu arbeiten verstärkt nutzen werden.

Im Februar 2021 wurde vom Unternehmen Deloitte eine Umfrage durchgeführt, bei welcher 1'093 Erwerbstätige in der Schweiz befragt wurden, ob sie auch nach der Covid-19 Pandemie weiterhin im Homeoffice arbeiten wollen. In Abbildung 17 sind die Ergebnisse der Umfrage zusammengefasst. 88 Prozent aller Erwerbstätigen wollen zumindest gelegentlich ihre Arbeitszeit im Homeoffice verbringen. 63 Prozent wollen mindestens die Hälfte ihrer Arbeitszeit im Homeoffice verbringen (Statista, 2021a).

Abbildung 17 Ergebnisse zur Umfrage Homeoffice-Wunsch im Februar 2021



Quelle: Statista, 2021a

7.3.2 Auswirkungen Homeoffice

Die Umfrage lässt darauf schliessen, dass der Anteil der Erwerbstätigen, die zumindest einen Teil ihrer Arbeitszeit im Homeoffice verbringen wollen, auch nach der Covid-19-Pandemie auf einem höheren Niveau bleibt.

Schon vor der Pandemie gab es unterschiedliche Untersuchungen und Studien, welche Auswirkungen die Verlagerung der Arbeitszeit in das Homeoffice haben kann.

Die Ausweitung von Homeoffice wird oft mit dem Potential verknüpft den Verkehr zu entlasten. Erwerbstätige, die von zu Hause aus arbeiten, müssen nicht zu ihrer Arbeitsstätte pendeln und entlasten damit den Verkehr. Die Bewertung dieses Potentials ist in der Literatur jedoch umstritten und hängt von verschiedenen Faktoren ab. Wird die Arbeitszeit während eines Arbeitstages vollständig im Homeoffice verbracht, kann dies Studien zu Folge insbesondere den Verkehr in den Hauptverkehrszeiten entlasten (Lachapelle *et al.*, 2018). Wird die Arbeitszeit jedoch in Form von Telearbeit an einem anderen Ort ausserhalb der regelmässigen Tätigkeitsstätte aber nicht zu Hause erbracht, sinkt die mögliche Entlastung des Verkehrs. Zudem hängt die Entlastung des Verkehrs insbesondere von der Entfernung zwischen Arbeitsstätte und Wohnort sowie der Verkehrsmittelwahl ab (Mokhtarian, 2004).

Weitere Untersuchungen zeigten, dass Erwerbstätige, die ihre Arbeit im Homeoffice verrichten auch ihre weiteren Aktivitäten zeitlich neu planen. Während Erwerbstätige die für die Verrichtung der Arbeit die Arbeitsstätte aufsuchen, weitere Aktivitäten um die Arbeitszeit herum planen, ändern Personen im Homeoffice die Zeiten dieser Aktivitäten. Es zeigt sich, dass dementsprechend auch die Freizeitaktivitäten gleichmässiger über den Tag verteilt werden (Pendyala *et al.*, 1991).

Mokhtarian (2004) führt die Eigenschaft des Menschen an, sich an neue Lebensumstände anzupassen. Menschen passen sich äusseren Umständen und Einwirkungen an. Insbesondere haben sich Erwerbstätige den Umständen des Pendelns angepasst. Die Auswirkungen des Pendelns wurden durch beispielsweise eine angenehmere Gestaltung der Fahrtzeit oder auch durch einen Wohnsitz- oder Arbeitsplatzwechsel minimiert. Betrachtet man diese Flexibilität, ist es möglich, dass Erwerbstätige im Umkehrschluss auch ihren Wohnsitz anpassen, sobald sie vermehrt im Homeoffice arbeiten (Graaff und Rietveld, 2007). Erhöhen sich die Pendeldistanzen, erhöhen sich demnach auch die Verkehrsbelastungen. Das Potential der Entlastung des Verkehrs verringert sich dementsprechend.

Auch Studien, die vor der Covid-19-Pandemie den Zusammenhang zwischen den Pendeldistanzen und der Arbeit im Homeoffice untersucht haben, kamen zu dem Ergebnis, dass Homeoffice-Nutzung mit längeren Distanzen zwischen Wohn- und Arbeitsort einhergeht (Rüger *et al.*, 2021). Sollte die durch Homeoffice eingesparte Fahrtzeit einen Anreiz zur Vergrösserung der Pendeldistanzen bieten, könnte eine Abwanderung der Einwohner von den Städten ins Umland entstehen (Rüger *et al.*, 2021). Da für die Arbeit im Homeoffice auch ein eingerichteter

Arbeitsplatz, im besten Fall ein Arbeitszimmer, benötigt wird, könnten die aktuell geringeren Immobilienpreise im Umland einen weiteren Anreiz für einen Umzug ins dieses schaffen.

Neben der Verringerung des Verkehrsaufkommens, würde eine stärkere Homeoffice-Nutzung auch zu einem geringeren Bedarf an Arbeitsplätzen am Standort des Arbeitsgebers führen. Vorzugsweise wenn die Tage, an denen Mitarbeiter im Homeoffice sind, sich unterscheiden. Insbesondere in teuren Ballungszentren, könnten Unternehmen teure Büroflächen einsparen (Garnadt *et al.*, 2020). Zudem böten sich Entwicklungsmöglichkeiten in Regionen mit derzeit geringerer ökonomischer Aktivität. Regionale Ungleichheiten könnten dadurch reduziert werden (Garnadt *et al.*, 2020). Um die soziale Isolierung im Homeoffice zu verhindern, könnten öffentliche Coworking-Spaces in der Nähe der neuen Siedlungsgebiete entstehen (Garnadt *et al.*, 2020).

Um die Auswirkungen von einem Anstieg des Niveaus der im Homeoffice verbrachten Arbeitstage zu erfassen, sind weitere Analysen des Datensatzes nötig. Auf Grundlage der ermittelten Homeoffice Tage könnte analysiert werden, wie sich die Anzahl der Fahrten für Personen, die im Homeoffice gearbeitet haben, von den Personen, die an der Arbeitsstelle gearbeitet haben, unterscheidet. Auch sollten die Längen der einzelnen Etappen verglichen werden. Diese Vergleiche zeigen die kurzfristigen Veränderungen während der Covid-19 Pandemie.

Aufgrund der besonderen Situation während Pandemie können anhand der aufgezeichneten Bewegungsdaten jedoch keine direkten Annahmen für die Auswirkungen auf den Verkehr nach der Pandemie abgeleitet werden. Während der Pandemie waren Freizeitaktivitäten nur eingeschränkt möglich. Ein Anstieg dieser Möglichkeiten führt auch zu mehr Bewegung. Die Anzahl der Fahrten könnte nach Ende der Pandemie dementsprechend wieder ansteigen. Eine weitere Möglichkeit wäre daher der Vergleich der prozentualen Änderungen der beiden Gruppen (Homeoffice und Personen die am Arbeitsplatz gearbeitet haben).

Durch weitere bereits erläuterte Änderungen, wie beispielsweise vergrößerte Pendeldistanzen, können langfristige Auswirkungen auf den Verkehr erst zu einem späteren Zeitpunkt beobachtet und untersucht werden.

8 Zusammenfassung

In dieser Projektarbeit wurden die Auswirkungen der Covid-19 Pandemie auf die Anzahl der Personen, die zumindest einen Teil ihrer Arbeit im Homeoffice verbringen, untersucht. Der untersuchte Zeitraum beginnt im November 2019 und endet im November 2020. Zur Untersuchung wurden aufgezeichnete Bewegungsdaten sowie zusätzliche Befragungen der Teilnehmer der MOBIS Studie sowie der MOBIS: Covid-19 Studie verwendet.

Die Daten der vier Erhebungen wurden zunächst aufbereitet. Anschliessend wurden die fehlenden Daten mit Hilfe von zwei verschiedenen Imputationsverfahren ergänzt. Die Daten aus den plausibleren Datensätzen wurden ausgewählt und für die weiteren Untersuchungen verwendet.

Der imputierte Datensatz wurde mit den aufgezeichneten Bewegungsdaten der Teilnehmer verknüpft. Neben den Bewegungsdaten wurden auch die Wege zur Arbeitsstätte aufgezeichnet. Anhand dieser Aufzeichnungen konnten die Tage bestimmt werden, an denen der Teilnehmer an seinem Arbeitsplatz gearbeitet hat. Da die Tage, die ein Teilnehmer im Homeoffice verbracht hat, nicht direkt aufgezeichnet wurden, wurde anschliessend ein Vorgehen entwickelt, dass bestimmt, an welchen Tagen der Teilnehmer im Homeoffice erwartet wurde. Mit den ermittelten Arbeitstagen am Arbeitsplatz sowie den prognostizierten Arbeitstagen im Homeoffice wurde anschliessend der Verlauf des Anteils der Personen im Homeoffice über die Zeit dargestellt und analysiert. Die ausgewerteten Daten zeigen deutlich, dass mehr Personen prinzipiell im Homeoffice gearbeitet haben. Zudem stieg auch der Anteil der im Homeoffice verbrachten Tage an.

Des Weiteren wurde mit den ermittelten Arbeitstagen am Arbeitsplatz sowie den prognostizierten Arbeitstagen im Homeoffice ein Entscheidungsmodell aufgebaut und geschätzt. Dabei wurden nur Personen berücksichtigt, die die Möglichkeit hatten im Homeoffice zu arbeiten. Neben soziodemographischen Merkmalen werden in das Entscheidungsmodell zudem auch spezifische Faktoren zur Berücksichtigung der Covid-19 Pandemie miteinbezogen. Zunächst wurde ein Logit-Modell entwickelt, das alle ausgewählten unabhängigen Variablen berücksichtigt. Da für das Modell keine ausreichende Güte nachgewiesen werden konnte, wurde zudem ein Mixed Logit-Modell entwickelt. Dieses kann neben den unabhängigen Variablen beobachtbare Effekte berücksichtigen. In diesem Projekt wurde dem Mixed Logit-Modell ein Faktor zur Berücksichtigung der beobachtbaren, individuellen Präferenzen hinzugefügt. Mit Hilfe des AIC wurde das Mixed Logit-Modell mit dem vorher erstellten Logit-Modell verglichen. Demnach ist die Anpassung des Mixed Logit-Modells besser als die Anpassung des Logit Modells. Der Vergleich der geschätzten Koeffizienten ergibt, dass im Mixed Logit-Modell nur wenige unabhängige

Variablen statistisch signifikant sind. Im einfachen Logit-Modell sind hingegen alle verwendeten Variablen statistisch signifikant. Aus dieser Beobachtung resultiert die Annahme, dass die persönlichen Präferenzen, die nicht beobachtet werden können, einen stärkeren Einfluss auf die Entscheidung haben als die allgemeinen soziodemographischen Merkmale. Die beiden vorgestellten Modelle erreichen beide keine hohe Anpassungsgüte. Zur weiteren Verwendung der Modelle, müssten diese weiter verbessert werden. Die Anpassung der Datengrundlage und die Berücksichtigung weiterer Variablen wären hierfür nötig.

Die berechneten Anteile der Arbeitstage, die im Homeoffice verbracht wurden, zeigen einen deutlichen Anstieg des Niveaus während der Covid-19 Pandemie. In den Sommermonaten 2020 waren zeitweise nur noch wenige Restriktionen in Kraft. Zudem sanken auch die gemeldeten Infektionen deutlich. Trotzdem waren die Anteile, der Personen die im Homeoffice gearbeitet haben, höher als das Homeoffice Niveau vor 2020. Diese Beobachtungen sowie Befragungen von Erwerbstätigen lassen vermuten, dass auch nach Ende der Pandemie der Anteil der Personen im Homeoffice höher sein wird als der Anteil der Personen, die vor der Pandemie zumindest zeitweise im Homeoffice gearbeitet haben.

Um die Auswirkungen von einem Anstieg des Niveaus der im Homeoffice verbrachten Arbeitstage zu erfassen, sind weitere Analysen des Datensatzes nötig. Homeoffice wird oft als Massnahme zur Reduktion des Verkehrs empfohlen. Aufgrund der besonderen Situation während der Covid-19 Pandemie kann man aufgrund der aufgezeichneten Bewegungsdaten keine direkten Annahmen für die Auswirkungen auf den Verkehr nach der Pandemie ableiten. Neben den kurzfristigen Auswirkungen auf den Verkehr werden zudem langfristige Anpassungen an die neue Situation erwartet. Die durch Homeoffice eingesparte Fahrtzeit kann einen Anreiz zur Vergrößerung der Pendeldistanzen bieten. Es könnte eine Abwanderung der Einwohner von den Städten ins Umland entstehen. Diese Entwicklungen können erst längerfristig beobachtet und untersucht werden.

Danksagung

Ich möchte allen danken, die mich während der Erstellung der Projektarbeit «Who is working from home and when?» unterstützt haben.

Ein besonderer Dank geht an meine Betreuer Herrn Prof. Dr. Kay W. Axhausen und Frau Caroline Winkler vom Institut für Verkehrsplanung und Transportsysteme (IVT) der ETH Zürich. Für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit möchte ich mich herzlich bedanken.

Literatur

Albers, Sönke (Hg.) (2007) *Methodik der empirischen Forschung*, 2., überarb. und erw. Aufl., Gabler, Wiesbaden.

Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. (2018) *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 15., vollständig überarbeitete Auflage, Springer Gabler, Berlin, Heidelberg.

Bloom, N., Liang, J., Roberts, J. und Ying, Z. J. (2015) Does Working from Home Work? Evidence from a Chinese Experiment, *The Quarterly Journal of Economics*, **1**, 165–218.

Bundesamt für Gesundheit BAG (17.03.2020a) Coronavirus: Bundesrat erklärt die «ausserordentliche Lage» und verschärft die Massnahmen. Online verfügbar unter <https://www.bag.admin.ch/bag/de/home/das-bag/aktuell/medienmitteilungen.msg-id-78454.html>, zuletzt geprüft am 10.05.2021.

Bundesamt für Gesundheit BAG (27.05.2020b) Coronavirus: Bundesrat beschliesst weitgehende Lockerungen per 6. Juni, Bern. Online verfügbar unter <https://www.bag.admin.ch/bag/de/home/das-bag/aktuell/medienmitteilungen.msg-id-79268.html>, zuletzt geprüft am 10.05.2021.

Bundesamt für Gesundheit BAG (2021a) Covid-19 Schweiz: Informationen zur aktuellen Lage, Situationsbericht, Website, <https://www.covid19.admin.ch/de/epidemiologic/case>, zuletzt geprüft am 31.03.2021.

Bundesamt für Gesundheit BAG (13.01.2021b) Coronavirus: Bundesrat verlängert und verschärft Massnahmen, Bern. Online verfügbar unter <https://www.bag.admin.ch/bag/de/home/das-bag/aktuell/medienmitteilungen.msg-id-81967.html>.

Bundesamt für Statistik (2021) Teleheimarbeit, Website, <https://www.bfs.admin.ch/bfs/de/home/statistiken/kultur-medien-informationsgesellschaft-sport/informationsgesellschaft/gesamtindikatoren/volkswirtschaft/teleheimarbeit.html>, zuletzt geprüft am 01.06.2021.

Careerbuilder (2021) Mitarbeiterrechte im Homeoffice: Worauf müssen Arbeitsgeber achten?, Website, <https://arbeitsgeber.careerbuilder.de/blog/mitarbeiterrechte-im-homeoffice>.

Dempster, A. P., Laird, N. M. und Rubin, D. B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **1**, 1–22.

Doméning, P. (2016) Homeoffice-Arbeit als besondere Erscheinungsform im Einzelarbeitsverhältnis, *Schriften zum Schweizerischen Arbeitsrecht SSA*, **79**.

Eurostat (2018) Working from home in the EU, Website, <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20180620-1>, zuletzt geprüft am 31.05.2021.

Garnadt, N., Schnitzer, M. und Viete, S. (2020) Räumliche Flexibilisierung durch zunehmende Homeoffice-Nutzung, *Wirtschaftsdienst*, **9**, 661–666.

Giménez-Nadal, J. I., Molina, J. A. und Velilla, J. (2019) Work time and well-being for workers at home: evidence from the American Time Use Survey, *IJM*, **2**, 184–206.

Graaff, T. de und Rietveld, P. (2007) Substitution between working at home and out-of-home: The role of ICT and commuting costs, *Transportation Research Part A: Policy and Practice*, **2**, 142–160.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Philipps, T. et al. (2021) A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker), *Nature Human Behaviour*, **5**, 529–538.

Honaker, J., King, G. und Blackwell, M. (2011) Amelia II: A Program for Missing Data, *J. Stat. Soft.*, **7**.

Honaker, J., King, G. und Blackwell, M. (2019) Package 'Amelia': A Program for Missing Data, *Dokumentation für R-Paket*.

Konjunkturforschungsstelle KOF (2021) KOF Stringency Indices, Website, <https://kof.ethz.ch/prognosen-indikatoren/indikatoren/kof-stringency-index.html>, zuletzt geprüft am 17.05.2021.

Lachapelle, U., Tanguay, G. A. und Neumark-Gaudet, L. (2018) Telecommuting and sustainable travel: Reduction of overall travel time, increases in non-motorised travel and congestion relief?, *Urban Studies*, **10**, 2226–2244.

Mokhtarian, P. L. (2004) Reducing road congestion: a reality check—a comment, *Transport Policy*, **2**, 183–184.

Pendyala, R. M., Goulias, K. G. und Kitamura, R. (1991) Impact of telecommuting on spatial and temporal patterns of household travel, *Transportation*, **18**, 383–409.

Rüger, h., Stawarz, N., Skora, T. und Jaszlovsky, V. (2021) Verändertes Pendelverhalten durch mehr Homeoffice? Mögliche Auswirkungen infolge der Corona-Pandemie, *Bevölkerungsforschung Aktuell*, **1**, 3–7.

Spiess, M. (2010) Missing Data Techniken: Analyse von Daten mit fehlenden Werten, 1. Auflage, Lit, Münster.

Staatssekretariat für Wirtschaft SECO (2019) Arbeiten zu Hause: Homeoffice, Bern.

Statista (2021a) Schweiz - Anteil Erwerbstätiger im Homeoffice 2021, Website, <https://de.statista.com/statistik/daten/studie/1117561/umfrage/anteil-der-schweizer-erwerbstaetigen-im-homeoffice/#professional>, zuletzt geprüft am 01.06.2021.

Statista (2021b) Schweiz - Homeoffice-Wunsch nach Corona-Krise 2021, Website, <https://de.statista.com/statistik/daten/studie/1231249/umfrage/umfrage-in-der-schweiz-zu-homeoffice-wunsch-nach-der-corona-pandemie/>, zuletzt geprüft am 01.06.2021.

Temme, J. (2007) Discrete-Choice-Modelle, in Sönke Albers (Hg.) *Discrete-Choice-Modelle*, 2., überarb. und erw. Aufl., Gabler, Wiesbaden.

van Buuren, S. und Oudshoorn, C. G. M. (2000) Multivariate imputation by chained equations: MICE V1.0 users's manual, Netherlands Organization for Applied Scientific Research (TNO), Leiden.

van Buuren, S. und Groothuis-Oudshoorn, K. (2021) Package 'mice': Multivariate Imputation by Chained Equations, *Dokumentation für R-Paket*.

Wahl, A. und Urban, D. (2020) Verfahren der Multiplen Imputation bei Schätzung von Strukturgleichungsmodellen mit latenten Variablen. Ein systematischer Vergleich mittels Monte-Carlo-Simulationen, *Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart*, **50**.

Anhang

A 1 Anhang zur Imputation der Datensätze

A 1.1. Verwendete Variablen im vorbereiteten Datensatz

Tabelle 12 Übersicht Variablen für vorbereiteten Datensatz

Bezeichnung Variable	Erklärung	Typ (Methode mice)	Ausprägungen
participant ID	ID	ID	
age	Alter	int	
gender	Geschlecht	nominal, int	0 = Female 1 = Male
postcode_home	PLZ Wohnort	int	
education	Höchster Bildungsabschluss	nominal, int	1 = Mandatory 2 = Secondary 3 = Higher
income	Einkommen	ordinal, int (polr)	1 = up to 4000 CHF 2 = 4001 – 8000 3 = 8001 - 12000 4 = 12001-16000 5 = More than 16000
household_size	Haushaltsgrösse	ordinal, int (polr)	

p_employment	Haupttätigkeit	nominal, int (polyreg)	1 = Employed 2 = Self_employed 3 = Unemployed 4 = Apprentice 5 = Student 6 = Retired 7 = Other
workload_jobs_main	Arbeitspensum, Haupttätigkeit	int (pmm)	
workload_jobs_secondary	Arbeitspensum, Nebentätigkeit	int (pmm)	
postcode_jobs_main	PLZ Arbeitsstätte Haupttätigkeit	int	
own_vehicles_car	Besitzt eigenes Auto?	nominal, int (polyreg)	1 = Yes 2 = Able to borrow one 3 = No
activity1	Aktivität	ordinal, int (polr)	1-7
p_weekly_commuter	Wochenaufenthalter	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_commutemode	Verkehrsmittelwahl Wohnort/Arbeit	nominal, int (polyreg)	1 = Walk 2 = Bike 3 = eBike 4 = e-Scooter 5 = Car 6 = Moto 7 = Bus/Tram 8 = Metro 9 = Train 10 = Other

			1 = Employed 2 = Self_employed 3 = Unemployed 4 = Apprentice 5 = Student 6 = Retired 7 = Other 8 = Pupil 9 = At home 10 = Military 11 = looking for work 12 = invalid
p_employment.x	Tätigkeit	nominal, int (polyreg)	
p_workload.x	Beschäftigungsgrad	int (pmm)	
p_homeoffice_nocoron	Möglichkeit Homeoffice ohne COVID-19	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_homeoffice_nocdays	#Tage/Woche Homeoffice ohne COVID-19	int (polr)	
p_homeoffice_coron.x	Möglichkeit Homeoffice mit COVID-19	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_kurzarbeit.x	Aktuell in Kurzarbeit	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_kurzarbeit..x	Beschäftigungsgrad%	int (pmm)	
p_occup_forcedleave.x	Zwangsurlaub	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_regular.x	Haben Sie regelmässige Arbeitszeiten	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_shift.x	Schichtarbeit	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_lcoation	Hauptarbeitsort ja/nein	nominal, int (logreg)	0 = TRUE 1 = FALSE

hh_size_kk	Anzahl Kleinkinder in Haushalt	int (polr)	
hh_size_k	Anzahl Kinder in Haushalt	int (polr)	
hh_size_jug	Anzahl Jugendliche in Haushalt	int (polr)	
hh_size_erw	Anzahl Erwachsene in Haushalt	int (polr)	
day_outworking.x	# Tagen an Arbeitsplatz ausserhalb der Wohnung	int (polr)	
day_homeworking.x	#Tage Homeoffice	int (polr)	
p_employment.y	Tätigkeit	nominal, int (polyreg)	1 = Employed 2 = Self_employed 3 = Unemployed 4 = Apprentice 5 = Student 6 = Retired 7 = Other 8 = Pupil 9 = At home 10 = Military 11 = looking for work 12 = invalid
p_workload.y	Beschäftigungsgrad	int (pmm)	
p_homeoffice_coron.y	Möglichkeit Homeoffice während COVID-19	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_homeoffice_cdays	#Tage/Woche Homeoffice während COVID19	int , (polr)	
p_occup_kurzarbeit.y	Kurzarbeit	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_kurzarbeit..y	Beschäftigungsgrad	int (pmm)	

p_occup_forcedleave.y	Zwangsurlaub	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_regular.y	Regelmässige Arbeitszeiten	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_shift.y	Schichtarbeit	nominal, int (logreg)	0 = TRUE 1 = FALSE
p_occup_location	Hauptarbeitsort ja/nein	nominal, int (logreg)	0 = TRUE 1 = FALSE
day_outworking.y	# Tagen an Arbeitsplatz ausserhalb der Wohnung	int (polr)	
day_homeworking.y	#Tage Homeoffice	int (polr)	
activity2	Aktivität	ordinal, int (polr)	1-7

A 2 Übersicht elektronischer Anhang

Im Elektronischen Anhang sind alle verwendeten Ausgangsdatensätze, generierten Datensätze sowie die einzelnen Codes, die in R verwendet wurden enthalten.

Im Ordner «Data IVT» sind alle Ausgangsdatensätze der MOBIS und der MOBIS: COVID-19 Studie enthalten. Alle weiteren verwendeten Datensätze aus anderen Quellen sind im Ordner «Input» gespeichert.

Im Ordner «Codes» werden alle Codes gespeichert. Tabelle 13 enthält eine Übersicht der einzelnen Codes in aufeinander aufbauender Reihenfolge.

Alle generierten Datensätze sind im Ordner «Outputs» dokumentiert. In Tabelle 14 ist eine Übersicht der erstellten Datensätze zusammengefasst.

Im Ordner «Models» sind alle erstellten Entscheidungsmodelle abgespeichert. Das finale Entscheidungsmodell Logit ist bezeichnet als «logit10t». Das Mixed Logit-Modell, das ebenfalls mit den zehn imputierten Datensätzen gerechnet wurde, ist mit dem Namen «mixed3t» bezeichnet. Unter «Logit10_observed» ist das Logit-Modell, dass ausschliesslich mit den beobachteten Werten geschätzt wurde, dokumentiert.

Tabelle 13 Übersicht der erstellten Codes für das Projekt

Bezeichnung	Beschreibung
Preparation Imputation	Vorbereitung des Datensatzes für die Imputation
Imputation Amelia	Imputation mit dem Paket Amelia
Imputation Mice	Imputation mit dem Paket Mice
Analysis Imputation	Analyse der Ergebnisse der beiden Imputationsmethoden, Plots und Lagemasse
Preparation Choice Model	Vorbereitung des Datensatzes für den Aufbau des Entscheidungsmodells der imputierte Datensatz wird mit den aufgezeichneten Bewegungsdaten verknüpft die unabhängige Variable «choice» wird erstellt und beobachteten Tage werden zugeordnet
Preparation Choice Model.x	Entspricht dem Datensatz «Preparation Choice Model» Bereitet die einzelnen imputierten Datensätze für den Aufbau des finalen Entscheidungsmodells vor
Data Choice Model Analysis	Analysiert die Daten vor Aufbau des Entscheidungsmodells
Choice Model	Aufbau der unterschiedlichen Entscheidungsmodelle Logit-Modell Probit-Modell Mixed-Logit-Modell
Choice Model observed	Datensatz für Aufbau des Choice Modells mit nur beobachteten Daten, imputierte Werte werden nicht miteinbezogen
Choice Model Final	Schätzung der finalen Entscheidungsmodelle mit Hilfe der einzelnen zehn imputierten Datensätze

Tabelle 14 Übersicht erstellte Datensätze

Bezeichnung	Beschreibung
data_complete.csv	Datensatz enthält alle Daten aus den vier Erhebungen
data_complete_prepared.csv	Datensatz nach Datenbereinigung Datensatz ist vorbereitet für die Imputation
Unterverzeichnis Amelia	Enthält alle 10 imputierten Datensätze, die mit der Funktion Amelia erstellt wurden.
data_imp_amelia_final.csv	Gemittelter Datensatz aus den zehn verschiedenen imputierten Datensätzen, die mit der Funktion Amelia erstellt wurden
Unterverzeichnis Mice	Enthält alle 10 imputierten Datensätze, die mit der Funktion Mice erstellt wurden.
data_imp_mice_final.csv	Gemittelter Datensatz aus den zehn verschiedenen imputierten Datensätzen, die mit der Funktion Mice erstellt wurden
data_prep_average.csv	Vorbereiteter Datensatz für Aufbau und Schätzung des Entscheidungsmodells mit dem gemittelten imputierten Datensatz
data_prepx.csv	Vorbereiteter Datensatz für Aufbau und Schätzung des Entscheidungsmodells mit den einzelnen imputierten Datensätzen (1-10)
data_observed_prep.csv	Vorbereiteter Datensatz für Aufbau des Entscheidungsmodells mit ausschliesslich beobachteten Werten
data_prep_total.csv	Beinhaltet die Daten aus allen zehn Datensätzen data_prepx.csv



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Eigenständigkeitserklärung

Die unterzeichnete Eigenständigkeitserklärung ist Bestandteil jeder während des Studiums verfassten Semester-, Bachelor- und Master-Arbeit oder anderen Abschlussarbeit (auch der jeweils elektronischen Version).

Die Dozentinnen und Dozenten können auch für andere bei ihnen verfasste schriftliche Arbeiten eine Eigenständigkeitserklärung verlangen.

Ich bestätige, die vorliegende Arbeit selbständig und in eigenen Worten verfasst zu haben. Davon ausgenommen sind sprachliche und inhaltliche Korrekturvorschläge durch die Betreuer und Betreuerinnen der Arbeit.

Titel der Arbeit (in Druckschrift):

Who is working from home and when?

Verfasst von (in Druckschrift):

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich.

Name(n):

Martin

Vorname(n):

Anna-Sophie

Ich bestätige mit meiner Unterschrift:

- Ich habe keine im Merkblatt „[Zitier-Knigge](#)“ beschriebene Form des Plagiats begangen.
- Ich habe alle Methoden, Daten und Arbeitsabläufe wahrheitsgetreu dokumentiert.
- Ich habe keine Daten manipuliert.
- Ich habe alle Personen erwähnt, welche die Arbeit wesentlich unterstützt haben.

Ich nehme zur Kenntnis, dass die Arbeit mit elektronischen Hilfsmitteln auf Plagiate überprüft werden kann.

Ort, Datum

Zürich 7.06.2021

Unterschrift(en)

A. Martin

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich. Durch die Unterschriften bürgen sie gemeinsam für den gesamten Inhalt dieser schriftlichen Arbeit.