femlogit: Implementation und Anwendung der multinominalen logistischen Regression mit "fixed effects"

Pforr, Klaus

Veröffentlichungsversion / Published Version
Monographie / monograph

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Pforr, K. (2013). *femlogit: Implementation und Anwendung der multinominalen logistischen Regression mit "fixed effects".* (GESIS-Schriftenreihe, 11). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.21241/ssoar.37199

gesis
Leibniz-Institut
für Sozialwissenschaften

Schriftenreihe Band 11

femlogit

Implementation und Anwendung der multinominalen logistischen Regression mit „fixed effects“

Klaus Pforr

femlogit: Implementation und Anwendung der multinominalen logistischen Regression mit „fixed effects“

GESIS-Schriftenreihe
herausgegeben von GESIS – Leibniz-Institut für Sozialwissenschaften

Band 11

Klaus Pforr

femlogit: Implementation und Anwendung der multinominalen logistischen Regression mit „fixed effects"

Die vorliegende Arbeit wurde von der Fakultät für Sozialwissenschaften der Universität Mannheim als Dissertation zur Erlangung des akademischen Grades eines Doktors der Sozialwissenschaften (Dr. rer. soc.) angenommen.

Dekan: Prof. Dr. Michael Diehl
1. Gutachter: Prof. Dr. Josef Brüderl
2. Gutachter: PD Dr. Henning Best
Vorsitzender der Prüfungskommision: Prof. Dr. Frank Kalter

Tag der Disputation: 09. Oktober 2012

Klaus Pforr

femlogit

Implementation und Anwendung der multinominalen logistischen Regression mit „fixed effects“

GESIS – Leibniz-Institut für Sozialwissenschaften 2013

Bibliographische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISBN 978-3-86819-020-5
ISSN 1869-2869

Herausgeber,
Druck u. Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln, Tel.: 0221 / 476 94 - 0
info@gesis.org
Printed in Germany

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Vorbemerkung

Ich danke Josef Brüderl für die Betreuung und Unterstützung, aber besonders dafür, dass er mir die Möglichkeit und die Freiheit gegeben hat, diese Arbeit fertigzustellen. Weiterhin danke ich Jette Schröder für die Inspiration zu dieser Arbeit und für die Bereitstellung der Analysedaten zur Replikation. Ich danke Volker Ludwig und Jeffrey Pitblado für viele hilfreiche Hinweise. Ferner danke ich Ulrich Kohler für die Bereitstellung der Analysefiles, die eine Reanalyse seiner Ergebnisse ermöglichten. Henning Best danke ich für die Übernahme des Zweitgutachtens und hilfreiche Hinweise zu dieser Arbeit. Außerdem danke ich Reinhard Schunck und Renate Pforr für die Prüfung des Manuskripts. Schließlich danke ich Milanka Pantic für die Prüfung des Manuskripts und für die kontinuierliche Unterstützung und Motivation auch in der schwersten Zeit.

Die Arbeit entstand im Rahmen des durch die Deutsche Forschungsgemeinschaft (DFG) geförderten, am Mannheimer Zentrum für europäische Sozialforschung (MZES) angesiedelten Panelprojekts „Pairfam LV”.

Mannheim, Januar 2013 Klaus Pforr

1 Einleitung

Zur Identifikation kausaler Effekte haben sich in den letzten Jahren auch in der Soziologie fixed-effects-Modelle mehr und mehr durchgesetzt. Mit diesen statistischen Modellen kann man mit Längsschnittdaten den Effekt zeitlich variierender Variablen schätzen und dabei vollständig für die Effekte zeitlich konstanter Variablen kontrollieren, ohne diese zu beobachten. Bei Multilevelmodellen, wie z. B. bei Modellen über Kinder in Schulklassen, kann man, sofern die Datenlage ausreicht, Effekte auf Schülerebene untersuchen und für alle Effekte auf Klassenniveau kontrollieren, selbst wenn die potentiell Einfluss nehmenden Drittvariablen nicht oder nur schwer messbar sind.

Statistische Modelle diesen Typs sind mittlerweile für verschiedenste abhängige Variablen entwickelt worden und in den gängigen Statistiksoftwarepaketen implementiert. So findet man in nahezu allen geläufigen Paketen vorgefertigte Schätzroutinen für kontinuierliche abhängige Variablen, für Zähldaten, für dichotome kategoriale Variablen und für Ereignisdaten. Bei Anwendungen, bei denen die abhängige Variable eine multinomiale kategoriale Variable oder eine metrische Variable mit Zensierung oder Trunkierung ist, liegen jedoch bestenfalls nur Implementationen von gepoolten und random-effects-Modellen vor.

Für multinomiale kategoriale abhängige Variablen hat Chamberlain (1980) ein fixed-effects-Modell als statistisches Modell entwickelt. Dieses ist jedoch bisher in keinem gängigen Statistikpaket implementiert worden. Ziel der vorliegenden Arbeit ist es, diese Forschungslücke zu schließen, und das statistische Modell von Chamberlain in Stata umzusetzen.

Um die Anwendbarkeit der Umsetzung unter Beweis zu stellen, wird das implementierte Modell für zwei Replikationsanalysen angewendet. Die erste Anwendung ist eine Replikationsanalyse von Schröder (2010), die den Effekt der Fertilität auf das Erwerbstätigkeitsverhältnis von Frauen untersucht. Eine zweite Beispielanwendung ist die Replikation einer Analyse von Kohler (2002), der den Einfluss der sozialen Klassenlage auf die Parteipräferenz analysiert.

Im Folgenden soll die Relevanz des multinomialen logistischen Regressionsmodells mit „fixed-effects” genauer beleuchtet werden. Zuvor werden die in der Arbeit verwendeten Begriffe definiert. Mit *fixed-effects-Modellen* sind, der ökonometrischen Terminologie folgend, statistische Modelle der folgenden Art gemeint: Zunächst einmal betrachten wir Daten auf mindestens zwei Ebenen. Bei Paneldaten ist die übergeordnete Ebene typischerweise eine Person mit einzelnen Messzeitpunkten auf der untergeordneten Ebene. Bei Mehrebenenanwendungen betrachtet man häufig Staaten auf der übergeordneten Ebene und Personen innerhalb der Staaten auf der untergeordneten Ebene. Allgemein formuliert betrachtet man bei fixed-effects-Modellen statistische Modelle dieser Art: $y_{it} = f(\alpha_i, x_{it}, \beta, \epsilon_{it})$. Die abhängige Variable y_{it} hängt von der unabhängigen Variable x_{it} ab. Der Effekt dieser Variable wird über den Regressionskoeffizienten β vermittelt. Die abhängige und die unabhängige Variablen variieren über die übergeordnete Ebene der Personen $i = 1, \ldots, N$ und über die untergeordnete Ebene der Messzeitpunkte $t = 1, \ldots, T$. Daneben hängt die abhängige Variable von zwei unbeobachteten Fehler-

termen ab: Der zeitlich konstanten, unbeobachteten Heterogenität α_i und dem unabhängigen Fehlerterm ϵ_{it}. Fixed-effects-Modelle zeichnen sich nun dadurch aus, dass die unbeobachtete Heterogenität als *Konstante* verstanden werden kann. Diese Modellierung ist äquivalent damit, dass die unbeobachtete Heterogenität beliebig mit der unabhängigen Variable korreliert sein kann.[1]

Als Gegensatz zu fixed-effects-Modellen werden in der Regel die sogenannten *random-effects-Modelle* angesehen. Hier wird die unbeobachtete Heterogenität α_i so modelliert, dass sie nicht als Konstante verstanden werden kann, sondern sie wird explizit als Zufallsvariable verstanden. Um die Regressionskoeffizienten schätzen zu können, muss diese Zufallsvariable dann als unabhängig von den unabhängigen Variablen angenommen werden. Diese Annahme ist natürlich wesentlich restriktiver.

In der vorliegenden Arbeit wird bewusst nicht auf random-effects-Modelle eingegangen. Im Gegensatz dazu wird das multinomiale logistische Regressionsmodell mit fixed-effects dem entsprechenden Querschnittsmodell und dem *pooled-Modell*, dass dem Querschnittsmodell in der Übertragung auf Längsschnittdaten entspricht, gegenübergestellt. Einerseits widerspricht diese Gegenüberstellung den konventionellen Modellvergleichen. Andererseits betrachten wir den Vergleich mit dem pooled-Modell als sinnvoller, da beim random-effects-Modell grundsätzlich dieselben Annahmen getroffen werden wie beim pooled-Modell, darüber hinaus aber zusätzlich strikte Exogenität und serielle Unabhängigkeit der Fehlerterme angenommen wird (vgl. hierzu Wooldridge 2010, S. 610–625 und die Erläuterungen in Abschnitt 3.2.2). D. h. wenn man die Annahme, dass die unbeobachtete Heterogenität von den unabhängigen Variablen unabhängig ist, treffen will, benötigt man für die konsistente Schätzung nicht die zusätzlichen Annahmen der random-effects-Modelle, wobei diese natürlich die Effizienz der Schätzer vergrößern, sofern die Annahmen auch gelten.

In der vorliegenden Arbeit werden die fixed-effects-Modelle auch als Längsschnittmodelle und pooled-Modelle auch als Querschnittsmodelle bezeichnet. Ausnahmen von dieser Regel sind ausdrücklich gekennzeichnet. Weiterhin wird das dichotome logistische Regressionsmodell als Querschnitts- oder pooled-Modell als *logit-Modell* abgekürzt. Das multinomiale logistische Regressionsmodell als Querschnitts- oder pooled-Modell wird als *mlogit-Modell* abgekürzt. Die entsprechenden Modelle mit fixed-effects werden dann als *felogit-* und *femlogit-Modell* abgekürzt.

1.1 Relevanz des statistischen Modells

Im Mittelpunkt dieser Arbeit steht das femlogit-Modell. Im Folgenden soll ausgeführt werden, warum dieses Modell von großer Bedeutung für die Sozialwissenschaften ist. Die Darstellung erfolgt zweigleisig entlang der beiden Komponenten, aus denen sich das Modell zusammensetzt, also der multinomialen logistischen Regression einerseits und andererseits der fixed-effects-Komponente. Die Bedeutung des mlogit wird im Wesentlichen

1 Man beachte, dass die Bedeutung der fixed-effects-Modelle häufig missverständlich damit erklärt wird, dass die Heterogenität konstant über die Zeit, bzw. die untergeordnete Dimension t, ist. Dieser Umstand ist aber nur eine Folge der Schätzung, nicht aber der Grund für die Modellbezeichnung. Vgl. hierzu Wooldridge (2010, S. 286).

anhand von Verweisen auf Standardlehrbücher in ihrer Entwicklung seit den 1980er Jahren und der Entwicklung der Zitationszahlen dieser Modelle ausgeführt. Die Relevanz der fixed-effects-Modelle, die die zweite Komponente bilden, erfolgt ebenso durch die Darstellung, wie sich die Behandlung der Modelle in den gängigen Lehrbüchern der Ökonometrie seit den 1980er Jahren entwickelt haben, und der Beschreibung der Zitationsentwickung der fe-Modelle in den vergangenen fünfzig Jahren. Abschließend wird die Bedeutung des Modells mit einer Bedarfsanalyse der Artikel der wichtigsten sozialwissenschaftlichen Zeitschriften der letzten fünf Jahre dargestellt.

1.1.1 Rezeption in der Literatur

Zunächst soll die Bedeutung der ersten Komponente, der multinomialen logistischen Regression, erläutert werden. Hier werden zwei Wege beschritten: 1) Welchen Stellenwert findet das Modell in den statistischen Standardlehrbüchern? 2) Werden Modelle diesen Typs in Fachartikeln zitiert?

Die multinomiale logistische Regression ist ein Spezialfall der allgemeinen discrete-choice-Modelle. Die Grundlagen dieser Modelle sind in den späten 70er Jahren des zwanzigsten Jahrhunderts entwickelt worden.[2] Der Blick in die Literatur zeigt zwei Dinge.[3]

Erstens wird aus der Literatur erkennbar, dass die aktuellen Standardlehrbücher der Ökonometrie den Modellen einen großen Stellenwert einräumen. Greene (2000) widmet den discrete-choice-Modellen ein eigenes Kapitel über 85 Seiten, davon alleine dem mlogit-Modell 17 Seiten.[4] Cameron und Trivedi (2009) und Wooldridge (2010) befassen sich mit den discrete-choice-Modellen auf zwei Kapiteln über ca. 70 bzw. ca. 100 Seiten, davon auf ca. 28 bzw. ca. 12 Seiten mit dem mlogit-Modell.[5]

Zweitens weisen die Autoren der Standardlehrbücher explizit auf die große Bedeutung der discrete-choice-Modelle im Allgemeinen und des mlogit-Modells im Speziellen hin.[6] Schon Amemiya (1981, S. 1484) stellt in Bezug auf discrete-choice Modelle im Allgemeinen fest:"Economists deal with many variables which are either naturally discrete or recorded discretely [...] I believe that QR [qualitative response] models are so important in economics that every applied researcher should acquire at least a cursory knowledge of these facts."Konkret für das mlogit-Modell führt er aus:"[...] so far as the alternatives are dissimilar, it [multinomial logit] is the most useful multi-response model, as well as being the most frequently used one." (Amemiya 1981, S. 1517). Maddala (1983, S. 35) verweist auf die leichte Berechenbarkeit des mlogit-Modells: "[...] from the computational point of view the logistic [distribution of the error term, implying the mlogit model,] is the easiest to handle."Außerdem listet er eine Reihe von Anwendungsbeispielen in der Ökonomie auf (ebd., S. 41f). Ähnlich findet sich wieder bei Amemiya (1985, S. 267): "Qualitative response models [...] are regressions in which dependent (or endogenous) variables take discrete

2 Die Geschichte dieser Modelle ist ausführlich bei McFadden (2001) dargestellt.
3 Vgl. z. B. Ben-Akiva und Lerman (1994); Long (1997); Greene (2000); McFadden (2001); Train (2009).
4 Siehe Greene (2000, S. 857–876).
5 Siehe Cameron und Trivedi (2009, S. 490–528) und Wooldridge (2010, S. 561–667).
6 Die frühen Lehrbücher verwenden für discrete-choice-Modelle häufig den Begriff des „qualitative-choice"-Modells, teilweise aber auch austauschbar für das spezielle mlogit-Modell.

values. These models have numerous applications in economics because many behavioral responses are qualitative in nature." Train (1986, S. 15) schreibt bzgl. des mlogit-Modells:

> "By far the most widely used qualitatitve choice model is logit. Its popularity is due to the fact that the formula for logit choice probabilities is readily interpretable, particularly compared with other qualitative choice models, and the parameters of logit models are relatively inexpensive to estimate."

Schon in den Lehrbüchern der ersten Generation wird auf das breite Anwendungsfeld und die einfache Schätz- und Interpretierbarkeit des mlogit-Modells hingewiesen.

Auch die Lehrbücher der späteren Generationen bleiben sehr deutlich. Agresti (1990, S. 2) weist wieder auf die vielfältige Anwendbarkeit der discrete-choice-Modelle hin:

> "Though categorical scales are common in the social and biomedical sciences, they are by no means restricted to those areas. They occur frequently in the behavioural sciences, public health, ecology, education, and marketing. They even occur in highly quantative fields such as engineering sciences and industrial quality control."

Maier und Weiss (1990, S. 3) stellen im ersten deutschsprachigen Standardlehrbuch dar: „Viele Probleme, mit denen sich die Sozial- und Wirtschaftswissenschaften beschäftigen, sind das Ergebnis diskreter Entscheidungen. In allen diesen Fällen können diskrete Entscheidungsmodelle zur Anwendung kommen und wurden bereits oft angewendet." Bezüglich der Attraktivität der discrete-choice-Modelle führen sie weiter aus:

> „Der wesentliche Vorteil gegenüber vielen traditionellen Methoden liegt auch hier darin, dass die Modelle die gesamte auf der Individualebene vorliegende Information verwenden können und erst die Ergebnisse zu den gesuchten Markoaussagen aggregiert werden müssen. Damit kann die Heterogenität der Bevölkerung weitgehend berücksichtigt werden, was zu präziseren Aussagen zu planungs- und politikrelevanten Fragen führt, insbesondere bei der Prognose der Auswirkungen einer veränderten sozioökonomischen Zusammensetzung der Bevölkerung.
>
> Die möglichen Anwendungsgebiete diskreter Entscheidungsmodelle reichen allerdings weit über die Ökonomie und ihre Spezialdisziplinen hinaus. Vor allem im Bereich der Betriebswirtschaftslehre steht ein weites Feld von Anwendungsmöglichkeiten offen. [...] Da die den Modellen zugrundeliegende Entscheidungsstruktur recht allgemein ist, gibt es in vielen Wirtschaftsdisziplinen Anwendungsmöglichkeiten." (Maier und Weiss 1990, S. 6f.)

Maier und Weiss weisen also darauf hin, dass discrete-choice-Modelle direkt auf die einzelnen Entscheidungen von Individuen angewendet werden können. Beim bis dahin gängigen Vorgehen werden die Entscheidungen der Individuen zu Anteilen, Mengen oder Stückzahlen aggregiert und mit konventionellen linearen Modellen analysiert. Der starke Informationsverlust über die individuelle Heterogenität, der beim ersten Schritt auftritt, wird in Kauf genommen (vgl. ebd., S. 1–6).

Auch die Autoren der neueren, noch immer aktuellen Standardlehrbüchern weisen immer noch auf die Bedeutung der discrete-choice-Modelle und des mlogit-Models hin. Long (1997, S. 148f.) formuliert im ersten Standardlehrbuch für den soziologischen Markt

sehr pointiert: “Nominal outcomes are found in every area of the social sciences. [...] The multinomial logit model is the most frequently used model for nominal outcomes.” Bei Greene (2000, S. 858) findet sich: “[...] the logit model [...] has been widely used in many fields, including economics, market research, and transportation engineering.” Über die aktuelle Bedeutung des mlogit-Modells befindet Train (2009, S. 18):

> “Logit [...] is by far the most widely used discrete choice model. It is derived under the assumption that ϵ_{ni} is iid extreme value for i. [...] This assumption, while restrictive, provides a very convenient form for the choice probability. The popularity of the logit model is due to this convenience.”

Zusammenfassend lässt sich also feststellen, dass die Autoren der Standardlehrbücher dem mlogit-Modell im Speziellen und den discrete-choice-Modellen im Allgemeinen ungebrochen eine große Bedeutung zuordnen. Ein weiterer Beleg für den hohen Stellenwert dieser statistischen Modelle ist zu werten, dass deren „Entdecker” Daniel McFadden im Jahr 2000 der Nobelpreis für Ökonomie verliehen wurde “for his development of theory and methods for analyzing discrete choice” (Nobelprize.org 2000).[7] Zur Bedeutung des mlogit-Modells führt McFadden (2001, S. 354f.) in Bezugnahme auf die Preisverleihung aus:

> “Viewed as a statistical model for discrete response, the MNL [multinomial logit] model was a small and in retrospect obvious contribution to microeconomic analysis, although one that has turned out to have many applications. The reason my formulation of the MNL model has received more attention than others that were developed independently during the same decade seemes to be the direct connection that I provided to consumer theory, linking unobserved preference heterogeneity to a fully consistent description of the distribution of demands. [...] The obvious similarities between the travel demand problem and applications such as education and occupation choices, demand for consumer goods, and location choices have led to adoption of these methods in a variety of studies of choice behavior of both consumers and firms.”

Insgesamt zeigt die Übersicht über die Standardlehrbücher ein geschlossenes Bild einer ungebrochen großen Bedeutung des hier zentralen Modells. Abschließend soll nun ein Blick auf die Rezeption des Modells in wissenschaftlichen Artikeln geworfen werden. Wie viele Arbeiten zu diesen Modellen werden über die Zeit hinweg betrachtet publiziert? Dieser Frage soll mit einer Abfrage des Social Science Citation Index (SSCI) beantwortet werden. Mit Hilfe der Datenbank wird für jedes Jahr von 1970–2010 ausgezählt, wie viele Arbeiten in einem Jahr erschienen sind und den Begriff „discrete choice“ im Titel oder im Abstract enthalten oder als Schlagwort verwenden.[8] Um das Wachstum der wissenschaftlichen Produktion zu berücksichtigen, werden diese Zahlen ins Verhältnis zur Gesamtzahl

7 Die Entwicklung dieser Modelle lässt sich im Wesentlichen auf eine Arbeit zurückführen, nämlich McFadden (1974).

8 Die Abfrage wurde am 20. Mai 2011 durchgeführt. Es ist zu beachten, dass die alleinige Verwendung des Begriffs „discrete choice“ problematisch ist. Einerseits werden dadurch nicht die frühen Arbeiten berücksichtigt, die unter dem Schlagwort „qualitative choice“ erschienen sind. Andererseits werden dadurch auch insbesondere jüngere Arbeiten berücksichtigt, die sich mit Modellen beschäftigen, die sehr weit entfernt von den hier zentralen Modellen sind (vgl. Train 2009). Insgesamt erscheint das hier dargestellte Vorgehen aber

aller Arbeiten gesetzt, die im jeweiligen Jahr erschienen sind. Diese Zeitreihe des Anteils der Arbeiten, die sich mit „discrete choice“ befassen, ist in Abbildung 1.1 dargestellt.

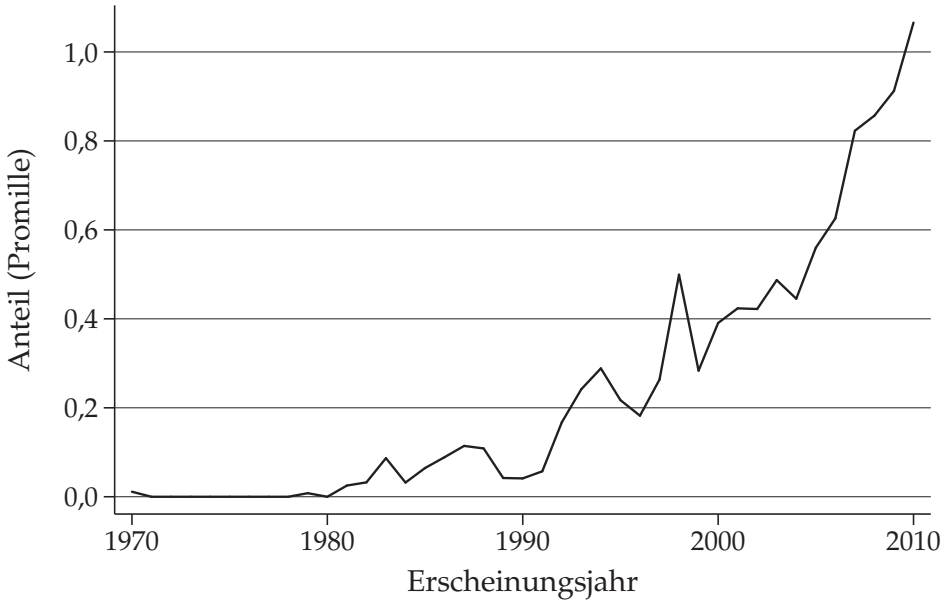


Abbildung 1.1: Publikationen zu „discrete choice” im SSCI

Es zeigt sich, dass der Anteil der entsprechenden Artikel noch bis Anfang der 90er Jahre mit ca. 0,1 Promille sehr gering ist, und dass der Anteil bis zum Jahrtausendwechsel auf knapp 0,5 Promille wächst. Zum Ausgang der Nullerjahre befasst sich jeder tausendste Artikel mit „discrete choice“. Insgesamt lässt sich also feststellen, dass die Bedeutung der discrete-choice-Modelle im Allgemeinen und des mlogit-Modells im Speziellen sehr groß ist.

Das in dieser Arbeit zentrale Modell besteht aber aus einer weiteren Komponente, indem es fixed-effects berücksichtigt. Im Weiteren soll die Bedeutung der fixed-effects-Modelle dargestellt werden, wobei das gleiche Vorgehen wie bei den obigen Ausführungen gewählt wird.

Der Blick in die Literatur zeigt drei Dinge. Erstens ist festzustellen, dass die fixed-effects-Modelle Anfang der 1960er entstanden sind, wobei die wesentlichen Beiträge in

vertretbar. Erstens ist die Anzahl der fälschlicherweise nicht berücksichtigten Artikel mit dem Schlagwort „qualitative choice” nach kursorischer Übersicht der Literatur verschwindend gering. Zweitens ist ein Rückgriff auf ein spezifischeres Schlagwort, wie z. B. „multinomiale logistische Regression”, nicht hilfreich, da der konkrete Modellname so gut wie nie verwendet wird. Insgesamt ist die Unterschätzung, die man bis Mitte der 1990er Jahre wegen der ersten Fehlerquelle erwarten könnte, gering. Wegen der zweiten Fehlerquelle ist aber von einer Überschätzung der Bedeutung des mlogit-Modells zumindest ab dem Jahrtausendwechsel auszugehen.

den 1970er Jahren entwickelt wurden.[9] In den gängigen Standardlehrbüchern finden sich zwar dezidierte Kapitel zu Panel-Modellen, jedoch sind die Ausführungen zu fixed-effects-Spezifikationen teilweise stark in den Büchern verteilt.[10]

Ähnlich wie bei den mlogit-Modellen finden sich auch hier durchgängige Hinweise auf die große Bedeutung von fixed-effects-Modellen von den ersten Lehrbüchern, die Mitte der 1980er Jahre erscheinen, bis zu den aktuellen Lehrbüchern. Schon Hsiao (1986) weist auf die wesentlichen Vorteile von fixed-effects-Modellen hin. Er verdeutlicht dies an einem Beispiel über den Einfluss der Gewerkschaftszugehörigkeit auf die Arbeitsproduktivität. Der stabil messbare Zusammenhang werde von vielen Autoren als Kausaleffekt interpretiert. Andere Autoren dagegen “believe that the observed union/nonunion differences are mainly due to differences between union and nonunion firms/workers prior to unionization or postunion sorting” (ebd., S. 2f.). Hsiao erklärt, dass man den vermuteten Selektionseffekt mit Paneldaten berücksichtigen kann:

> “A single cross-sectional data set usually cannot provide a direct choice between these two hypotheses, because the estimates are likely to reflect interindividual differences inherent in comparisons of *different* people or firms. However, if panel data are used, one can distinguish these two hypotheses by studying the wage differential for a worker moving from a nonunion firm to a union firm, vice versa.” (Hsiao 1986, S. 2f., Hervorhebung im Original)

Trotz dieses Verweises auf die Vorteile von Paneldaten für die Schätzung von Kausaleffekten relativiert Hsiao (1986, S. 41ff.) an anderer Stelle die Vorteile von fixed-effects Modellen:

> “In general, whether one wishes to consider the conditional likelihood function [i.e. fixed-effects models] or the marginal-likelihood function, [i.e. random-effects models] depends on the context of the data, the manner in which they are gathered, and the environment from which they came. [...] When inferences are going to be confined to the effects in the model, the effects are more appropriately considered fixed. When inferences will be made about a population of effects from which those in the data are considered to be a random sample, then the effects should be considered random.”

Gegenüber Hsiao betont Allison (1994, S. 181) nahezu uneingeschränkt die Vorteile von fixed-effects-Modellen: “[The] danger [of ‘unmeasured selectivity’] leads me to conclude that the change-score [i.e. ‘fixed-effects’] estimator is nearly always preferable for estimating the effects of events with nonexperimental data.” Diese Position wird von Baltagi (1995, S. 3ff.) geteilt:

> “There are several benefits from using panel data: [...] controlling for individual heterogeneity. [...] Panel data are able to control for these state and time-invariant variables for inclusion in the consumption equation. Omission of those variables leads to bias in the resulting estimates.”

9 Vgl. hierzu z. B. Halaby (2004, S. 508). Cameron und Trivedi (2009) nennen als erste Arbeiten Kuh (1959) und Hoch (1962). Diese Autoren vergleichen Zeitreihen- und Querschnittmethoden. Die erste Verwendung des fixed-effects-Begriff im Zusammenhang mit der Identifizierung von individueller Heterogenität findet sich vermutlich bei Mundlak (1961).

10 Siehe die relativ geordnete, aber oberflächliche Darstellung bei Greene (2000, S. 560–567,837–841) und die dagegen wesentlich deutlicheren, aber stark verteilten Ausführungen bei Cameron und Trivedi (2009, S. 715–803) und Wooldridge (2010, S. 300–315,619,762).

Maddala (2001, S. 576) vertritt ebenso die relativierende Position von Hsiao. Er betont die Nachteile von fixed-effects-Modellen gegenüber random-effects-Modellen. Bei fixed-effects-Modellen verliere man gegenüber random-effects-Modellen Information, da nur Fälle mit Variation auf der abhängigen Variablen genutzt werden. Weiterhin könne die unbeobachtete Heterogenität immer als Zufallsvariable modelliert werden und müsse nicht als fix angenommen werden. Ferner sei die Entscheidung für fixed- bzw. random-effects-Modellen grundsätzlich abhängig von der Perspektive, die man gegenüber den Daten einnimmt. Wolle man Aussagen über die Grundgesamtheit machen, müsse man ein random-effects-Modell verwenden. Wenn man ein fixed-effects-Modell verwendet, könne man nur Rückschlüsse über die vorliegende Stichprobe machen. Schließlich hätten fixed-effects-Modell gegenüber random-effects-Modellen den Nachteil, dass man keine Effekte für zeitlich konstante Variablen schätzen könne.[11]

Dem gegenüber verweist Lee (2002, S. 16, Notation angepasst) wiederum auf die Vorteile von Paneldaten:

> "Compared with cross-section data, panel data hold a number of advantages. First, whereas cross-section data have difficulty controlling for unobserved variables, panel data can control for them much better either by removing them or by providing more instruments: the ability to remove time-invariant unobservable α_i can be the single most important advantage of panel data."

Noch prägnanter formuliert es Arellano (2003, S. 8): "A major motivation for using panel data has been to control for possibly correlated, time-invariant heterogeneity without observing it."

Trotz der frühen starken Positionierung von Allison findet man erst ab diesem Zeitpunkt eindeutige Verweise auf die entschiedenen Vorteile der fixed-effects-Modelle für die Schätzung von Kausaleffekten. Halaby (2004, S. 516f.) führt in seinem Überblicksartikel aus:

> "The DID, fixed effects, and first difference estimators offer researchers the capacity to dispense with the random effects assumption and still obtain unbiased and consistent estimates of parameters when unit effects are arbitrarily correlated with measured explanatory variables. This is widely regarded as the primary advantage of panel data and the reason the effort to extend the benefits of fixed effect models beyond the static linear case is one of the central thrusts of econometric research on panel analysis over the past 15 years."

Allison (2009, S. 1) wiederholt seine frühere Positionierung und bringt den Kern der fixed-effects-Modellierung folgendermaßen auf den Punkt: "[...] fixed effects models [...] make it possible to control for variables that have not or cannot be measured. The basic idea is simple: Use each individual as his or her own control." Auch Cameron und Trivedi

11 Das Argument von Hsiao (1986) und Maddala (2001) gegen fixed-effects-Modelle, dass in späteren Arbeiten nicht mehr zu finden ist, ist insgesamt nicht nachvollziehbar. Allem Anschein sollen fixed-effects-Modelle nur für die betrachtete Stichprobe interpretierbar sein, da bei linearen Modellen die individuellen unbeobachteten Heterogenitäten eben nur für die Stichprobe schätzbar sind, während man dagegen bei den random-effects-Modellen qua Annahme eine Verteilung für die Population hat.

(2009, S. 715) lassen in ihrem Standardlehrbuch keinen Zweifel an den Vorteilen der fe-Modellierung:

> “The fixed effects model has the attraction of allowing one to use panel data to establish causation under weaker assumptions [...] than those needed to establish causation with cross-section data or with panel data models without fixed effects, such as pooled models and random effects models.”

Wooldridge führt in seinem Standardlehrbuch aus, dass er im Vorteil der fixed-effects-Modelle den eigentlichen Grund für die Erhebung von Paneldaten sieht: “In many applications the whole point of using panel data is to allow for α_i to be arbitrarily correlated with the X_{it}. A fixed effects analysis achieves this purpose explicitly” (Wooldridge 2010, S. 300, Notation angepasst). Diese Position ist insoweit beachtlich, als die Erhebung von Paneldaten mit wesentlich höheren Kosten verbunden ist (vgl. z. B. Diekmann 2007, S. 311). D. h. Wooldridge weist explizit darauf hin, dass der Vorteil für die Kausalanalyse einerseits mit höheren Erhebungskosten bezahlt werden muss, aber auch umgekehrt diese Erhebungskosten erst dadurch legitimiert werden, wenn man Paneldaten auch entsprechend mit fixed-effects-Modellen oder ähnlich adäquaten Verfahren ausnutzt. In ähnlicher Weise fasst Brüderl (2010, S. 975) diese Positionen zusammen:

> „Angesichts der Tatsache, dass bei den meisten sozialwissenschaftlichen Fragestellungen personenspezifische unbeobachtete Heterogenität vorhanden sein dürfte, ist die Antwort [auf die Frage, ob man das RE- oder FE-Modell verwenden soll,] eigentlich klar: Man sollte das FE-Modell verwenden, weil es den besonderen Vorzug von Paneldaten – die Möglichkeit des Within-Vergleichs – voll umsetzt, und deshalb bei Vorliegen von personenspezifischer unbeobachteter Heterogenität nicht verzerrt ist. Dagegen wird in dieser Situation der RE-Schätzer verzerrt sein, weil er auch den Between-Vergleich mit einbezieht [...].“

Zusammenfassend findet man also in der Literatur, insbesondere seit dem Jahrtausendwechsel, starke Hinweise auf die Bedeutung der fixed-effects-Modelle. Es stellt sich auch hier die Frage, inwieweit Arbeiten zu diesem Thema rezipiert werden. Wie oben soll dies wieder mit einer Abfrage des SSCI untersucht werden. Analog werden wieder die Arbeiten für jedes Jahr von 1970–2010 gezählt, die in einem Jahr erschienen sind, und das Stichwort „fixed-effects“ als Schlagwort oder im Titel oder im Abstract enthalten.[12] Diese Zahlen werden durch die Gesamtzahl aller Publikationen im jeweiligen Publikationsjahr geteilt. Die Anteile der Arbeiten, die sich mit fixed-effects-Modellen befassen, sind in Abbildung 1.2 dargestellt.

Es zeigt sich ähnlich wie bei der Entwicklung der Bedeutung der mlogit-Modelle, dass der Anteil der Artikel, die sich mit fixed-effects-Modellen befassen, bis Anfang der 1990er Jahre verschwindend gering ist. Ab diesem Zeitpunkt wächst der Anteil exponentiell. Um 2000 ist der Anteil ca. 0,5 Promille, ungefähr im Jahr 2004 bei 1,0 und 2010 sind 1,5 Promille aller Artikel zu fixed-effects-Modellen erschienen.

12 Die Abfrage erfolgte am 19. August 2011. Auch hier ist die alleinige Verwendung des Begriffs „fixed-effects“ angreifbar, nicht zuletzt da dieser Begriff in den 1970er Jahren noch nicht durchgängig verwendet wurde. Insgesamt erscheint auch hier das Vorgehen als vertretbar.

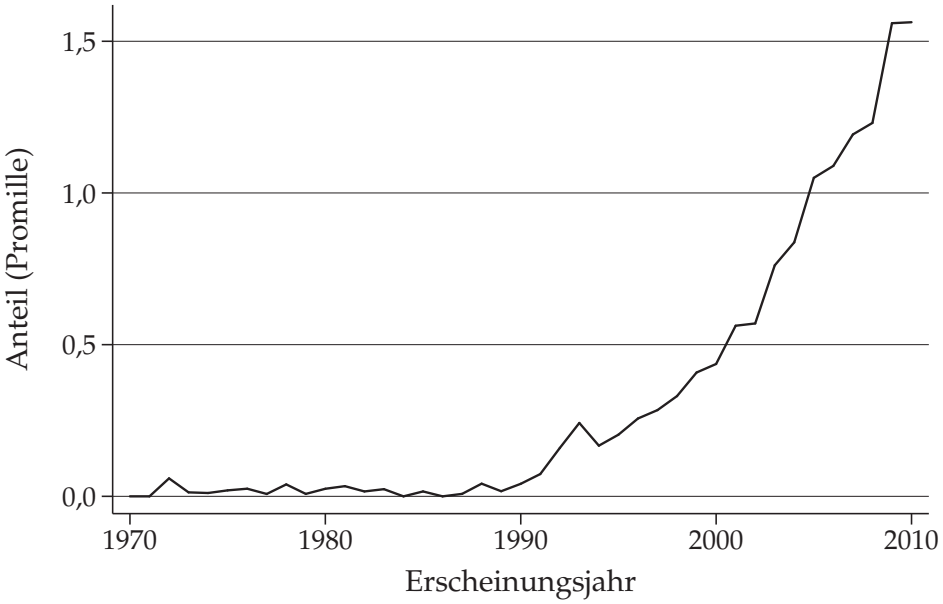


Abbildung 1.2: Publikationen zu „fixed-effects" im SSCI

Zusammenfassend findet man also in verschiedenen autoritativen Quellen Hinweise auf die große Bedeutung der beiden Komponenten des in dieser Arbeit implementierten Modells. Die große Bedeutung der mlogit- und fixed-effects-Modelle wird außerdem dadurch gestützt, dass die Publikationszahlen zu diesen Modelle seit den 1990er Jahren stark ansteigen.

1.1.2 Bedarf in der aktuellen Forschung

Die bisherigen Verfahren geben aber nur indirekte Hinweise auf die gegenwärtige Bedeutung des in dieser Arbeit im Zentrum stehenden statistischen Verfahrens. Daher soll die Bedeutung durch eine zusätzliche Bedarfsanalyse der aktuellen Literatur gezeigt werden. Hiermit soll untersucht werden, wie der Bedarf für das femlogit in der aktuellen sozialwissenschaftlichen Literatur ist. Um diese Frage zu beantworten, wurden jeweils die zwei wichtigsten Zeitschriften der Soziologie, Politikwissenschaft und Ökonomie die Ausgaben seit 2006 nach Artikeln durchgesehen, die für ihre Kernaussage durch die Verwendung des femlogit-Modells profitieren würden. Für die Soziologie wurden die Zeitschriften „American Sociological Review" (ASR) und „American Journal of Sociology" (AJS), für die Politikwissenschaft die Zeitschriften „American Political Science Review" (APSR) und „American Journal of Political Science" (AJPS) und für die Ökonomie die Zeitschriften „Quarterly Journal of Economics" (QJE) und „Journal of Political Economy"

(JPE) analysiert. Die Tabelle 1.1 zeigt die verwendeten Ausgaben der Zeitschriften und ihren „impact factor".[13]

Tabelle 1.1: Untersuchte Zeitschriften

Bereich	Zeitschrift	Ausgaben	Impact Factor	Rang
Soziologie	ASR	71(1)–76(4)	3,278	1
	AJS	111(4)–117(1)	2,588	3
Politikwissenschaft	APSR	100(1)–105(2)	3,693	1
	AJPS	50(1)–55(3)	3,358	2
Ökonomie	QJE	121(1)–126(2)	5,940	2
	JPE	114(1)–119(3)	4,065	6

Der Rang gibt den Rang des impact factors der Zeitschriften im jeweiligen Fachbereich wieder. Die Auswahl dieser Zeitschriften ergibt sich daraus, dass die höherrangigen Zeitschriften einen ungeeigneten Schwerpunkt haben. Für den Fachbereich Soziologie ist die Zeitschrift „Annual Review of Sociology" auf dem zweiten Rang. Da in dieser Zeitschrift aber nur Überblicksartikel veröffentlicht werden, in denen eigene Analysen die Ausnahme sind, ging diese nicht in die Analyse ein. Im Fachbereich Politikwissenschaft sind die beiden höchstrangigen Zeitschriften gut geeignet für die Auszählung der Artikel, die durch das femlogit-Modell profitieren würden. Die höchstrangige Zeitschrift im Fachbereich Ökonomie „Journal of Economic Literatur" hat ebenfalls einen Schwerpunkt auf theoretische Überblicksartikel, d. h. sie ist für diese Untersuchung ungeeignet. Auf Rang drei bis fünf sind die Zeitschriften „Technological and Economic Development of Economy", „Review of Financial Studies", „Journal of Finance" mit spezielleren finanzwissenschaftlichen Ausrichtungen, so dass auch diese für diese Untersuchung nicht berücksichtigt wurden. Wie in Tabelle 1.2 dargestellt, wurden in den betrachteten Ausgaben dieser Zeitschriften 1 373 Artikel veröffentlicht. Von diesen waren 996 empirische Artikel.

Die Zuordnung erfolgte nach den folgenden Kriterien: Ein Artikel wird als nicht empirisch ausgeschlossen, wenn (1) sich der Artikel auf eine rein theoretische Ausarbeitung beschränkt, wenn (2) der methodische Teil sich auf Simulationen beschränkt, die nicht durch multivariate Analysen geprüft werden und wenn (3) sich der methodisch-empirische Teil auf univariate, einfache bivariate oder einfache Zeitreihen-Analysen beschränkt.[14]

Von diesen 996 empirischen Artikeln fanden sich 42 Artikel, die von der Verwendung des in dieser Arbeit im Mittelpunkt stehenden Verfahrens profitieren würden.[15] Hier sind

13 Quelle: ISI Web of Science, Stand Mai 2011.

14 Schwer einzustufende Fälle sind häufig sogenannte Modellkalibrierungen in der Ökonomie. Hier werden für komplexe theoretische Modelle für manche Parameter Modellannahmen getroffen und andere Parameter werden mit verschiedenen Verfahren geschätzt. Da der Schwerpunkt in diesen Arbeiten in der Regel aber auf dem theoretischen Modell liegt, wurden diese in der Regel als nicht empirisch klassifiziert.

15 Diese Treffer sind: *ASR*: 71(4), 618–638; 72(5), 659–680; 75(2), 205–226; 75(2), 303–329; 74(2), 316–337; 74(6), 916–937; *AJS*: 111(6), 1910–1949; 112(4), 1095–1134; 114(4), 871–923; 116(3), 855–908; 116(6), 1982–2018; *APSR*: 100(1), 133–146; 102(3), 279–301; 102(4), 509–524; 103(2), 175–192; 103(2), 193–213;

Tabelle 1.2: Ergebnisse der Auszählung replizierbarer Artikel

Fachbereich	N Total	N Empirisch	Relevant	% von Empirisch	% von Total
Soziologie	430	322	11	3,42 %	2,56 %
Politikwissenschaft	540	404	21	5,20 %	3,89 %
Ökonomie	403	270	10	3,70 %	2,48 %
Insgesamt	1 373	996	42	4,20 %	3,06 %

die Auswahlkriterien natürlich essentiell für das Ergebnis. Ein Artikel wird im Wesentlichen als relevant betrachtet, wenn vier notwendige Bedingungen erfüllt sind. Die (1) Beobachtungseinheit muss mindestens zweistufig sein, d. h. dass es neben der „Individual"-Ebene eine zweite Ebene wie die Zeit oder eine strukturell übergeordnete Ebene wie z. B. Staaten gibt. Weiterhin muss die (2) abhängige Variable polytom kategorial sein. Diese Bedingung gilt auch als erfüllt, wenn zumindest in der methodischen Diskussion eine kategoriale Operationalisierung angesprochen wurde.[16] Ferner gibt es zumindest Hinweise für (3) Varianz mindestens einer zentralen Kovariaten innerhalb der übergeordneten Ebene über die untergeordnete Ebene.[17] Schließlich darf eine (4) fixed-effects-Modellierung nicht kategorisch zugunsten einer anderen Modellierung ausgeschlossen sein. In der Regel führen hier die Autoren aus, dass eine Random-Effects-Modellierung entscheidende Vorteile gegenüber einer fixed-effects-Modellierung hat. Wenn diese vier Bedingungen für die konkreten Daten und die konkreten Hypothesen in einem Artikel zutreffen, sollte dieser Artikel von der Verwendung des Modells profitieren.[18]

Die Auszählung ergibt also, dass 4,2 % aller seit 2006 erschienen empirischen Artikel der sechs wichtigsten sozialwissenschaftlichen Zeitschriften durch die Verwendung des in dieser Arbeit im Mittelpunkt stehenden statistischen Verfahrens profitieren würden. Auf den ersten Blick unterstreicht das die bereits dargestellte große Bedeutung des Modells. Dennoch muss auf verschiedene mögliche Fehlerquellen hingewiesen werden. Eine Möglichkeit der Unterschätzung besteht im publication bias. Gerade bei den hier

104(2), 243–267; 104(2), 307–323; 104(3), 503–518; 104(4), 625–643; *AJPS*: 50(1), 62–80; 50(2), 294–312; 51(1), 140–150; 51(1), 166–191; 51(3), 640–654; 52(2), 322–343; 53(3), 552–571; 54(1), 18–33; 54(2), 494–510; 55(2), 201–218; 55(2), 247–262; 55(2), 398–416; *QJE*: 121(1), 267–288; 122(2), 441–485; 122(4), 1603–1637; 123(1), 219–277; 124(3), 1057–1094; *JPE*: 115(3), 447–481; 116(4), 709–745; 118(2), 300–354; 118(3), 599–647; 118(6), 1110–1150.

16 Ein Grenzfall sind ordinale Modellierungen. Die notwendige Bedingung gilt hier als erfüllt, wenn erkennbar ist, dass die Autoren für einzelne Stufen der ordinalen Variablen unterschiedliche Effekte einzelner zentraler unabhängiger Variablen erwarten.

17 In der Regel ist diese Varianz entweder innerhalb der Individuen über die Zeit oder innerhalb von Staaten über die Individuen gegeben.

18 Es ist darauf hinzuweisen, dass die Hypothesen alleine in der Regel noch viel Interpretationsraum für die Auswahlentscheidung lassen. Daher wurde, auch um die Auswahl zu erleichtern, in erster Linie nur die konkreten gerechneten Modelle für die Auswahlentscheidung herangezogen.

ausgewählten Zeitschriften ist zu erwarten, dass ein Artikel nur dann erscheint, wenn er positive, signifikante Ergebnisse vorbringt und in sich logisch konsistent ist.[19] D. h. ein Artikel, der so geschrieben ist, dass er vordergründig den Mangel aufweist, dass das statistische Verfahren für die Hypothesen unzureichend ist, sollte wahrscheinlich eher abgelehnt werden. Artikel in den sechs betrachteten Zeitschriften sollten sich also tendenziell in ihren Kernaussagen auf Anwendungen beschränken, die mit den verfügbaren Methoden überzeugend gezeigt werden können. Mit anderen Worten sollten also Artikel in diesen Zeitschriften vergleichsweise „perfekter" sein als durchschnittliche Artikel, d. h. bei diesen sollte insgesamt tendenziell weniger Verbesserungsbedarf bestehen. Also folgt alleine aus dem publication bias, dass Artikel in den sechs wichtigsten Zeitschriften in ihrer jeweiligen Kernaussage auch vergleichsweise seltener von dem in dieser Arbeit behandelten Verfahren profitieren sollten als der durchschnittliche Artikel.

Ein weiterer Grund gegen eine zu optimistische Interpretation der oben ausgeführten Ergebnisse ist die Forschungskultur in den Teildisziplinen. In der Soziologie und der Politikwissenschaft findet man zwar einen vergleichsweise hohen Anteil empirischer Arbeiten. Außerdem finden sich hier im Vergleich gerade zur Ökonomie viele empirische Anwendungen, die nicht kontinuierliche abhängige Variablen in besonderer Weise berücksichtigen, wie z. B. discrete-choice-Modelle, aber auch Modelle für ordinale Daten oder Zähldaten. Jedoch hat sich in der Soziologie und der Politikwissenschaft im Vergleich zur Ökonomie die fixed-effects-Modellierung nicht durchgesetzt (vgl. z. B. Halaby 2004). Gerade in der Politikwissenschaft findet man zwar viele Arbeiten, die Panel- oder Multi-Level-Daten ausnutzen, aber hier ist die random-effects-Modellierung gängig. In der Ökonomie hingegen gibt es erstens eine Tendenz zur einfachen Interpretierbarkeit. Dies führt dazu, dass in der Ökonomie lineare Modellierungen wesentlich weiter verbreitet sind als in den anderen Disziplinen. Zweitens gibt es in der Ökonomie einen starken Trend hin zu fortgeschritteneren Kausalanalyse-Modellierungen, wie z. B. klassische Experimente, der Einsatz von Instrumentvariablen von exogenen Naturereignissen und regression-discontinuity-Designs. Diese Untersuchungsstrategien haben den Vorteil – natürlich nur bei Gültigkeit der zugrundeliegenden Annahmen – dass sie ebenso den Einfluss störender unbeobachtbarer Drittvariablen ausschließen können, und so eine Alternative zu fixed-effects-Modellierungen darstellen können (vgl. z. B. Morgan und Winship 2007).

Die gegenwärtigen Forschungsarbeiten bieten also viel Potential für die Anwendung des in dieser Arbeit zentralen Modells. Eine zu naive Betrachtung dieses Ergebnisses übersieht aber die Hindernisse, die sich aus der Forschungskultur in diesen Teildisziplinen ergeben. Insgesamt sind die Ergebnisse der Auszählung ein starker Hinweis für die große Relevanz des femlogit-Modells, da erstens zu erwarten ist, dass die hier dargestellte Auszählung den Bedarf durch die Auswahl der Zeitschrift eher unterschätzt. Zweitens kann die Forschungskultur alleine nicht als definitives Ausschlusskriterium betrachtet werden,

19 Die Annahmequote der Zeitschriften sind nach öffentlich zugänglichen Informationen und persönlichen Anfragen, sofern verfügbar, folgendermaßen: ASR 2010 6 %; AJS 2010 10 %; APSR 2010 9,1 %; AJPS 2010 11,4 %; JPE „5 %–8 % in den letzen Jahren". Die Annahmequote des QJE war trotz mehrere Anfragen nicht in Erfahrung zu bringen.

wenn auch zu erwarten ist, dass in manchen Feldern für die spezifischen Vorteile des femlogit-Modells stärker argumentiert werden muss.

Zusammenfassend ist festzustellen, dass es in den Standardlehrbüchern der Ökonometrie und der gegenwärtigen Forschungsliteratur zahlreiche Hinweise auf die große Relevanz des in dieser Arbeit im Mittelpunkt stehenden statistischen Verfahrens gibt. Im nächsten Abschnitt wird dargestellt, wie das Modell in der Forschung gegenwärtig angewendet wird.

1.2 Bisherige Anwendungen

Die ursprüngliche Ableitung des Modells findet sich bei Chamberlain (1980, S. 231). Diese sehr knappe Darstellung ist eine skizzierte Verallgemeinerung des bekannten binären logit-Modells mit "fixed-effects". Eine klare Darstellung des Modells in einem Lehrbuch findet sich bei Lee (2002, S. 143–148). In neueren Lehrbüchern wie Cameron und Trivedi (2009, S. 798) und Allison (2009, S. 44) finden sich nur oberflächliche Darstellungen. Allison weist an dieser Stelle explizit darauf hin, dass "there is no commercial software available", mit der das femlogit-Modell geschätzt werden kann.

Obwohl eine allgemeine Implementation des Modells nicht verfügbar ist, finden sich Anwendungen in der Literatur. Die ersten Anwendungen des femlogit-Modells findet man bei Börsch-Supan (1987, Kap. 10), Börsch-Supan (1990) und Börsch-Supan und Pollakowski (1990). In verschiedenen Abwandlungen verwendet Börsch-Supan das Modell zur Analyse der Nachfrage nach Immobilien. Bei diesen Analysen verwendet Börsch-Supan Personen-Paneldaten und berücksichtigt entsprechend fixed-effects innerhalb einer Person über die Zeit hinweg. Es werden vier Alternativen und zwischen drei und fünf Zeitpunkte untersucht. Bezüglich der Umsetzung stellt Börsch-Supan ausführlich dar, wie er durch Datentransformation das femlogit-Modell schätzen kann. Hierzu werden, wie in Abbildung 1.3 dargestellt, für jede Beobachtung alle Permutationen der Alternativenkombinationen über den Beobachtungszeitraum als Alternativenraum für die entsprechende Beobachtung angenommen.

Zeit	Abh. Var.
1	1
2	2
3	3

$\Rightarrow$

Lfd. Nr.	Permutation	Realisation
1	$(1,2,3)$	1
2	$(1,3,2)$	0
3	$(2,1,3)$	0
4	$(2,3,1)$	0
5	$(3,1,2)$	0
6	$(3,2,1)$	0

Abbildung 1.3: Datentransformation nach Börsch-Supan (1987)

Die Abbildung zeigt die Alternativenwahlen für eine beliebige Untersuchungseinheit. Zum ersten Messzeitpunkt hat die Einheit die Alternative 1 gewählt, zum Zeitpunkt 2 die Al-

ternative 2 und zum Zeitpunkt 3 die dritte Alternative. Für diese Zeitreihe $1, 2, 3$ ergeben sich sechs verschiedene Permutationen, die rechts in der Abbildung dargestellt sind. Wenn man nun die Permutationen als mögliche Alternativen betrachtet und die realisierte Zeitreihe als gewählte Alternative, dann kann man das femlogit-Modell mit "multinomial logit packages that allow for choice sets that vary by [individual]"[20] geschätzt werden. Das Problem bei dieser Lösung besteht darin, dass man für alle Beobachtungseinheiten der Stichprobe die Menge aller Permutationen bilden muss. Dadurch benötigt man bei vielen Anwendungen einen gewaltigen Datensatz, der auch von heute gebräuchlichen Rechnern nicht bearbeitet werden kann.

Weitere Anwendungen des femlogit-Modells finden sich bei Pitt und Rosenzweig (1990), Rosenzweig (1993) und Rosenzweig und Wolpin (1994). Rosenzweig wendet das Modell zusammen mit Kollegen auf verschiedene Fragestellungen an, wobei der jeweils gemeinsame Ansatz die Betrachtung von Transfers innerhalb von Haushalten ist. Pitt und Rosenzweig (1990) verwenden einen Mehrebenen-Ansatz und berücksichtigen hierbei haushaltsspezifische fixed-effects. Es werden vier Alternativen und innerhalb eines Haushalts jeweils drei Personen betrachtet. Rosenzweig (1993) verwendet Paneldaten über Haushalte, so dass haushaltsspezifische fixed-effects über die Zeit hinweg betrachtet werden. Bei dieser Arbeit werden drei Alternativen und drei Zeitpunkte untersucht. Bei Rosenzweig und Wolpin (1994) werden ebenso Paneldaten über Haushalte untersucht, wobei wieder entsprechend haushaltsspezifische fixed-effects über die Zeit berücksichtigt werden. Hier werden sechs Alternativen untersucht; die Anzahl der betrachteten Zeitpunkte ist nicht genau nachvollziehbar. Die explizite Schätzung wird in den drei Arbeiten nur sehr oberflächlich dargestellt. Allem Anschein nach wird auch in diesen Arbeiten dieselbe Datentransformations-Technik wie bei Börsch-Supan verwendet.

Auf diese Arbeiten folgen mehrere einzelne Arbeiten, die das femlogit-Modell anwenden: Pradhan (1998) wendet das Modell auf die Analyse von Bildungsentscheidungen an. Die fixed-effects-Modellierung ergibt sich durch eine Mehrebenenstruktur in den Daten, wobei innerhalb eines Sampleclusters mehrere Haushalte als eigentliche Beobachtungseinheiten vorliegen. Es werden drei Alternativen und innerhalb eines Clusters ca. 16 Haushalte betrachtet. Es wird knapp erwähnt, dass für die Umsetzung die Datentransformation-Technik von Börsch-Supan verwendet wird. Royalty und Solomon (1999) verwenden das Modell zur Analyse der Nachfrage nach bestimmten Formen privater Krankenversicherungspakete. Die Autoren verwenden Paneldaten über Individuen und berücksichtigen wie Börsch-Supan fixed-effects innerhalb einer Person. Es werden vier Alternativen und drei Zeitpunkte betrachtet. Sie führen explizit aus, dass sie zur Analyse das gleiche Verfahren wie Börsch-Supan und Pollakowski (1990) verwenden. Hoffman und Foster (2000) wenden das Modell auf die Untersuchung von Fertilität an. Die Autoren verwenden Personen-Paneldaten mit Mehrebenen-Struktur. Bei den Analysen berücksichtigen sie sowohl personen-spezifische, als auch Mehrebenen-fixed-effects. Sie betrachten drei Alternativen. Das Mehrebenen-Panel-Design betrachtet Personenjahre als Einheiten der unteren Ebene und Bundesstaaten als Einheiten der oberen Ebene. Die Anzahl der Personenjahre

20 Siehe Börsch-Supan (1987, S. 83), vgl. aber auch die ungenauere und optimistischere Formulierung bei Börsch-Supan (1990, S. 71) und Börsch-Supan und Pollakowski (1990, S. 136).

pro Bundesstaat geht aus den Ausführungen nicht hervor. Die Umsetzung des Modells wird nicht genau nachvollziehbar dargestellt. In direkter Bezugnahme auf Royalty und Solomon verwenden Scanlon, Chernew, McLaughlin, und Solon (2002) das Modell inhaltlich im Wesentlichen auf den gleichen Untersuchungsgegenstand. Auch hier liegt wieder ein Personenpanel vor, so dass die Autoren bei dem Individuen-fixed-effects berücksichtigen können. Soweit es die undeutliche Darstellung des Designs erkennen lässt, werden 69 Alternativen und zwei Zeitpunkte untersucht. Auch bei der Darstellung der konkreten Umsetzung bleiben die Autoren sehr oberflächlich. Der Verweis auf Royalty und Solomon (1999) deutet aber darauf hin, dass auch in dieser Arbeit das Verfahren wie bei Börsch-Supan verwendet wird. Ähnlich wie bei den Arbeiten von Börsch-Supan untersucht Andrew (2004) die Nachfrage nach Immobilien. Er betrachtet Personenpaneldaten mit Individuen-fixed-effects. Es werden vier Alternativen und fünf Zeitpunkte untersucht. Die genaue Implementation wird nicht ausgeführt. Eine Anwendung in der Seelogistik findet sich bei Malchow und Kanafani (2004). In dieser Arbeit werden Mehrebenendaten analysiert, wobei Beförderer-spezifische fixed-effects über mehrere Lieferungen hinweg berücksichtigt werden. Es werden acht Umschlaghäfen als Alternativen berücksichtigt; die Anzahl der Lieferungen pro Beförderer ist nicht nachvollziehbar. Auch die genaue Umsetzungstechnik wird nicht erläutert. Schließlich findet sich eine Anwendung auf die Wahlentscheidung bei Lind (2007). Er verwendet Personenpaneldaten unter Berücksichtigung von Personen-fixed-effects über die Zeit. Es werden vier Alternativen und zwei Zeitpunkte untersucht. Die genaue Umsetzung des mlogit-Modells wird nicht dargestellt.

Bei acht von den zwölf bisherigen Anwendungen kann sicher davon ausgegangen werden, dass sie die Datentransformationstechnik von Börsch-Supan (1987) verwenden.[21] Dieses Verfahren hat zur Folge, dass die Anwendungen vereinfacht werden müssen, um den Datensatz handhabbar zu halten. Hierfür werden, sofern die Anzahl der Alternativen und Zeitpunkte nicht ohnehin schon klein ist, bei einzelnen Anwendungen entweder eine Zufallsstichprobe aus den Beobachtungen oder aus den Zeitreihen gezogen (vgl. Cameron und Trivedi 2009, S. 797). Die in dieser Arbeit dargestellte Implementation des femlogit-Modells kann dieses Problem lösen, so dass grundsätzlich alle diese Anwendungen ohne Einschränkungen bearbeitet werden können. Im nächsten Kapitel wird der statistische Hintergrund des Modells genauer dargestellt.

1.3 Aufbau

Die Arbeit ist in zwei Teile gegliedert. Der erste Teil befasst sich ausschließlich mit dem statistischen femlogit-Modell und seiner Implementation in Stata. Hier wird in Kapitel 3 zunächst auf rein theoretischer Ebene das statistische Modell der fixed-effects- und dazu vergleichend das der pooled-Modelle vorgestellt und erläutert. In diesem Kapitel wird zunächst eine einheitliche Notation eingeführt. Dann werden im zweiten Unterkapitel in zwei Teilen der statistische Hintergrund zu den pooled- und zu den fixed-effects-Modellen

21 Bei den Arbeiten von Rosenzweig ist zwar erkennbar, dass die gleiche Technik verwendet wird, es ist aber unklar, ob die Autoren die Technik selbst entwickelt haben, oder von den früheren Arbeiten von Börsch-Supan übernommen haben.

ausführlich dargestellt. Abschließend werden verschiedene Erweiterungen und Probleme dieser Modelle diskutiert. Danach wird in Kapitel 4 die Umsetzung des femlogit-Modells in Stata dargestellt. Hier wird zunächst das Maximum-Likelihood-Verfahren theoretisch erklärt. Im zweiten Schritt wird die praktische Umsetzung des femlogit-Modells in Stata erläutert. Abschließend wird in Abschnitt 4.3 ein Performanztest der Implementation durchgeführt, der zeigt, dass die Implementation erwartungsgemäß konsistente Schätzer liefert.

Der zweite Teil der Arbeit befasst sich ausschließlich mit der Anwendung des femlogit-Modells auf echte Fragestellungen. Hierfür wird in Kapitel 7 die Anwendung auf eine Replikation von Analysen aus der Arbeit von Schröder (2010) gezeigt. In Kapitel 8 wird das femlogit-Modell für eine Replikation eines Teils der Arbeit von Kohler (2002) dargestellt. In beiden Unterabschnitten wird zunächst eine kurze Einführung in die theoretische Fragestellung und ein knapper Überblick über den Forschungsstand der jeweiligen Arbeit gegeben. Danach werden die eigentlichen Analysen der Originalarbeiten, die Replikationen und die erweiterten Analysen unter Verwendung des femlogit-Modells gezeigt und diskutiert.

Teil I

Implementation des multinomialen logistischen Regressionsmodells mit „fixed effects"

2 Einleitung

Der erste Teil der Arbeit befasst sich mit dem femlogit-Modell. Zuerst wird das statistische Modell erläutert. Die Darstellung legt ihren Schwerpunkt auf die Einbettung des Modells in die direkt verwandten Modelle. Weiterhin geht die Darstellung von den bestehenden Ausführungen aus den ökonometrischen Standardwerken aus. Die Ausführungen verfolgen im Wesentlichen drei Zielstellungen. Erstens wird die Verwandtschaft zum felogit- und zum mlogit-Modell ausgeführt. Zweitens wird der discrete-choice-Hintergrund des Modells mit der Herleitung aus dem Zufallsnutzenmodell dargestellt. Im Zuge dessen wird die Modellierung von alternativen-spezifischen und „generischen“ Variablen erläutert. Drittens wird für die Darstellung eine einheitliche Notation und ein einheitlicher Rahmen verwendet, so dass die verwandten Modelle gemeinsam dargestellt werden können.

Nach der Beschreibung des theoretischen Hintergrunds wird die eigentliche Implementation ausgeführt. Hierfür werden zuerst allgemein dargestellt, wie Maximum-Likelihood-Schätzer in Stata implementiert werden. Anschließend wird die allgemeine Implementationsstrategie ausgeführt. Weiterhin werden die speziell für die Implementation notwendigen mathematischen Schritte erläutert, also die ersten beiden Ableitungen der Likelihoodfunktion. Danach wird die Implementation Schritt für Schritt im Detail dargestellt. Ferner wird die Implementation mit simulierten Daten, bei denen die Ergebnisse a priori bekannt sind, getestet, und Sensitivitätschecks der Simulation durchgeführt, dargestellt und diskutiert. Abschließend werden verschiedene Implementationsalternativen diskutiert.

3 Statistik

Im Weiteren soll nun das zentrale statistische Modell femlogit dargestellt werden. Das Modell wird als allgemeines Modell über mehrere Alternativen gezeigt. Die binäre logistische Regression wird als Spezialfall abgebildet. Das fixed-effects-Modell wird dem gepoolten Querschnittsmodell, also dem mlogit-Modell, gegenüber gestellt. So sollen die Ähnlichkeiten und die Unterschiede der beiden Modelle verdeutlicht werden.

Die Darstellung verfolgt im Wesentlichen drei Zielstellungen. Erstens soll die Verwandtschaft des femlogit-Modells zur binären logistischen Regression mit fixed-effects und zur multinomialen logistischen Regression im Querschnitt aus klar und deutlich gezeigt werden. Zweitens soll der Zufallsnutzen-Hintergrund des Modells veranschaulicht werden. Drittens wird für die Darstellung eine einheitliche Notation und einheitlicher Rahmen verwendet, so dass die verwandten Modelle gemeinsam abgebildet werden können.

3.1 Notation

Im Folgenden wird eine einheitliche Notation eingeführt, die bei der Darstellung der Querschnitts- und Längsschnittsmodellen durchgängig verwendet wird. Die hier verwendete Notation weicht von der Notation der Standardwerke Long (1997); Greene (2000); Cameron und Trivedi (2009); Wooldridge (2010) ab und folgt der Notation von Lee (2002), ergänzt diese jedoch. Dies ist hauptsächlich darin begründet, dass nur Lee die multinomiale logistische Regression mit fixed-effects ausführlich darstellt. Die Notation in Chamberlain (1980) ist zu ungenau für eine verständliche, integrierte Darstellung der beiden Querschnitts- und der beiden Längsschnittmodelle. Zunächst wird die Notation der beobachteten, danach die der unbeobachteten Variablen dargestellt. Zum leichteren Verständnis wird die Notation der beobachteten Variablen getrennt für die Querschnitts- und Längsschnittmodelle dargestellt.

Bei Querschnittsmodellen betrachtet man eine unabhängige und identisch verteilte Stichprobe von N Personen. Die Personen entscheiden über J Alternativen. Die Alternative $B \in \{1, \ldots, J\}$ ist die Referenzkategorie. Die Entscheidung über die J Alternativen für Person i wird mit y_i bezeichnet.

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}.$$

Parallel zu den kategorialen Entscheidungen gibt es für jede Alternative eine latente Neigung y_j^*, diese zu wählen.

Weiterhin werden M unabhängige Variablen x_{im} gemessen, die über die Personen hinweg variieren können, d. h. x_{im} bezeichnet die Ausprägung der m-ten unabhängigen

Variable für Person i. Die M unabhängigen Variablen werden als Zeilenvektor $X_i = (x_{i1}, \ldots, x_{iM})$ und als Matrix

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \ldots & x_{NM} \end{pmatrix}$$

zusammengefasst dargestellt. Der Vektor der unabhängigen Variablen X_i enthält keine Konstante.

Bei Längsschnittsmodellen betrachtet man ebenso eine unabhängige und identisch verteilte Stichprobe von N Personen. Für die Person i mit $i = 1, \ldots, N$ liegen T_i Messzeitpunkte vor. Die Personen entscheiden wieder über J Alternativen; die Alternative $B \in \{1, \ldots, J\}$ ist wieder die Referenzkategorie. Die Zeitreihe der Entscheidungen über die J Alternativen für Person i wird mit y_i bezeichnet. Die Entscheidung für Person i zum Zeitpunkt t wird mit y_{it} bezeichnet. Die Zeitreihen über alle Personen hinweg werden als gestapelte Vektoren notiert:

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT_i} \end{pmatrix}, \; y = \left(y_{11}, \ldots, y_{1T_1}, y_{21}, \ldots, y_{2T_2}, \ldots, y_{N1}, \ldots, y_{NT_N}\right)'.$$

Wie im Querschnitt gibt es auch hier für jede Alternative eine latente Variable y^*_{itj}, die Neigung, eine bestimmte Alternative zu wählen, ausdrückt.

Weiterhin werden M unabhängige Variablen x_{itm} gemessen, die über die Personen und Messzeitpunkte hinweg variieren können, d. h. x_{itm} bezeichnet die Ausprägung der m-ten unabhängigen Variable für Person i zum Zeitpunkt t. Die m unabhängigen Variablen werden hier als Zeilenvektor X_{it} bzw. Matrix X_i oder X zusammengefasst dargestellt. X_{it} bezeichnet den Vektor $(x_{it1}, \ldots, x_{itM})$. Die Matrix der unabhängigen Variablen X_i enthält auch im Längsschnitt keine Konstante. Die Matrizen entsprechen diesen Ausdrücken:

$$X_i = \begin{pmatrix} x_{i11} & \ldots & x_{i1M} \\ \vdots & \ddots & \vdots \\ x_{iT_i1} & \ldots & x_{iT_iM} \end{pmatrix},$$

$$X = \begin{pmatrix} x_{111} & \ldots & x_{11M} \\ \vdots & & \vdots \\ x_{1T_11} & \ldots & x_{11M} \\ \vdots & & \vdots \\ x_{N11} & \ldots & x_{N1M} \\ \vdots & & \vdots \\ x_{NT_N1} & \ldots & x_{NT_NM} \end{pmatrix}.$$

Insgesamt ergibt sich also die Querschnittsnotation als Spezialfall der Längsschnittnotation, da hier $T_i = 1$ für alle i Personen gilt. In diesem Fall werden y_{it} als y_{i1}, x_{itm} als x_{i1m} und X_{it} als X_{i1} notiert. Die explizite Darstellung des t-Subskripts ist hinfällig, so dass y_{i1}, x_{i1m} und X_{i1} verkürzt als y_i, x_{im} und X_i dargestellt werden.

Die unbekannten Modellparameter werden folgendermaßen notiert. Die zeitlich konstante, unbeobachtete Heterogenität für Person i für die Alternative j wird mit α_{ij} bezeichnet. Im Längsschnitt ist die unbeobachtete Heterogenität eine Zufallsvariable. Im Querschnitt degeneriert die Heterogenität zur alternativen-spezifischen Konstanten α_j, die geschätzt wird. Insoweit kann sie im Querschnitt nicht mehr als Heterogenität betrachtet werden, da sie über die Beobachtungseinheiten hinweg konstant ist. Der Regressionskoeffizient für die m-te unabhängige Variable für die Alternative j wird mit β_{jm} bezeichnet. Die Regressionskoeffizienten für die Alternative j werden als Zeilenvektor $\beta_j = (\beta_{j1}, \ldots, \beta_{jM})$ zusammengefasst. Diese Vektoren werden wiederum im Zeilenvektor $\beta = (\beta_{11}, \ldots, \beta_{JM})$ zusammengefasst. Im Querschnitt ist α_j analog zu β ein zu schätzender Regressionskoeffizient. An einzelnen Stellen wird im Querschnitt der Vektor der zu schätzenden Regressionskoeffizienten als Vektor (α, β) notiert.

Bei den Regressionskoeffizienten wird allgemein zwischen dem wahren Wert und dem Schätzwert unterschieden. Der wahre Regressionskoeffizient wird ohne weitere Zusätze mit β notiert. Der entsprechende Schätzwert wird mit der „Dach”-Schreibweise $\widehat{\beta}$ notiert. Die entsprechende Notation gilt im Querschnitt für α. Der idiosynkratische Fehlerterm wird bei den Querschnittsmodellen mit ϵ_{ij}, bei den Längsschnittmodellen mit ϵ_{itj} bezeichnet.[1]

Abschließend ist noch auf die Unterscheidung zwischen der Grundgesamtheit und der Stichprobe hinzuweisen. Zur Darstellung des statistischen Modells wird analytisch in Bezugnahme auf Wooldridge (2010) zwischen der Betrachtungsebene der theoretischen Grundgesamtheit und der Stichprobe unterschieden. Das statistische Modell bezieht sich auf Zufallsvariablen in der theoretischen Grundgesamtheit. Analytisch getrennt davon ist die Stichprobe eine Realisation dieser gemeinsam verteilten Zufallsvariablen. D. h. diese Unterscheidung wird in der Darstellung so vollzogen, dass das statistische Modell in der Grundgesamtheit ohne Beobachtungsindizes dargestellt wird, sofern dies nicht irreführend ist. Dagegen wird die konkrete Stichprobe immer mit dem Verweis auf Beobachtungsindizes notiert.[2]

3.2 Modelle

Im Folgenden werden die binäre und die multinomiale logistische Regression im Querschnitt und im Längsschnitt mit fixed-effects ausgehend von den notwendigen Annahmen erläutert. Die Darstellung gibt im Wesentlichen die Ausführungen von Chamberlain

1 Mit dem idiosynkratischen Fehlerterm ist der Fehler gemeint, der im Querschnitt einer Beobachtungseinheit i und einer Alternative j, bzw. im Längsschnitt einer Beobachtungseinheit i, einer Alternative j und einem Messzeitpunkt t zugeordnet wird.

2 Der Vorteil dieser Unterscheidung beruht im Wesentlichen darauf, dass die Variablen in der Grundgesamtheit als Zufallsvariablen verstanden werden. Dies erlaubt die konsequente Betrachtung möglicher Korrelationen zwischen den abhängigen und unabhängigen Variablen und den Fehlertermen. Vgl. ausführlich Wooldridge (2010, S. 3–11).

(1980), Greene (2000), Cameron und Trivedi (2009) und Wooldridge (2010) wieder. Diese werden in der einheitlichen, zuvor eingeführten Notation wiedergegeben. Für die Darstellung wird der Bezugsrahmen von Wooldridge (2010, S. 3–11) gewählt, d. h. die statistischen Modelle beziehen sich auf Zufallsvariablen in der theoretischen Grundgesamtheit. Weiterhin erfolgt die Ausführung grundsätzlich strukturell getrennt für die Querschnitts- und die Längsschnittmodelle. Dies widerspricht zwar einerseits der gesetzten Zielvorgabe einer einheitlichen Darstellung. Andererseits würde eine solche Darstellung der fixed-effects-Modelle und der Querschnitts-Modelle einen falschen Eindruck über den Zusammenhang zwischen den Modellfamilien vermitteln. Die fixed-effects-Modellierung ist eine von mehreren Möglichkeiten, das Problem der unbeobachteten Heterogenität tendenziell abzuschwächen. Bei Querschnitts-Modellen muss die unbeobachtete Heterogenität per Annahme ausgeschlossen werden. Wie im Weiteren noch ausführlich dargestellt wird, kann bei der fixed-effects-Modellierung zwar auf die Annahme verzichtet werden, dass die zeitlich konstante Heterogenität entweder voll beobachtbar ist oder keine Rolle spielt. Zur Identifikation der Regressionskoeffizienten müssen aber andere Annahmen getroffen werden, die vom eigentlichen Problem der unbeobachteten Heterogenität unabhängig sind. Sowohl die Querschnitts- als auch die Längsschnittmodelle werden aus dem Zufallsnutzen-Ansatz nach McFadden (1974) hergeleitet dargestellt, da so der Bezug zu den Fehlertermen deutlicher wird.

3.2.1 Querschnitts- und Pooled-modelle

Im Folgenden werden die Querschnittsmodelle logit und mlogit beschrieben. Die Darstellung des logit-Modells folgt den Ausführungen von Greene (2000, S. 811–825), Cameron und Trivedi (2009, S. 466–470) und Wooldridge (2010, S. 469–482, 565–569), die Darstellung des mlogit-Modells bezieht sich auf Greene (2000, S. 857–865), Cameron und Trivedi (2009, S. 490–506) und Wooldridge (2010, S. 643–651). Die beiden Modelle werden wegen ihren starken Verwandtschaft in einem Modell vorgestellt, und die Unterschiede an entsprechender Stelle diskutiert.

Annahmen

Annahme 1 (Lineares parametrisches Modell)**:**
Den Ausgangspunkt bildet das folgende statistische Modell, dass für jede Beobachtungseinheit in der betrachteten Population gilt. Eine Beobachtungseinheit entscheidet über eine fixe Anzahl von J Alternativen. Für alle $j \in \{1, \ldots, J\}$ Alternativen gilt nun:

$$y_j^* = \alpha_j + X\beta_j' + \epsilon_j.$$

Im Kontext der Zufallsnutzenmodellierung stellt diese Gleichung den Nutzen der Alternative j dar. In diesem Modell sind für alle $j \in \{1, \ldots, J\}$ Alternativen die Ausdrücke y_j^*, x_m und ϵ_j Zufallsvariablen. Dagegen sind α_j und alle β_{jm} konstante Parameter. Für die Querschnittsmodelle gelten nun folgende Annahmen:

Annahme 2 (Verteilung der Fehlerterme):
Für alle Alternativen ist der idiosynkratische Fehlerterm ϵ_j Standard-Gumbel-verteilt, d. h. formal:[3]

$$\forall j \in \{1, \ldots, J\} : \epsilon_j \sim \text{Gumbel mit } \mathrm{E}(\epsilon_j) = \gamma \text{ und } \mathrm{Var}(\epsilon_j) = \frac{\pi^2}{6}.$$

Die in der Verteilungsannahme enthaltene Festlegung der Varianz auf $\pi^2/6$ ist für die Identifikation der Regressionskoeffizienten notwendig. Diese Festlegung ist die Ursache für das Problem der „neglected heterogeneity” (vgl. Wooldridge 2010, S. 582ff.). Dadurch wird die Interpretation der Effekte wesentlich erschwert (vgl. Allison 1999; Mood 2010). Das Problem der „neglected heterogeneity” ist in einem eigenen Abschnitt auf Seite 62 erläutert.

Annahme 3 (Verknüpfung zu beobachteter Entscheidung):
Der Zusammenhang zur beobachteten Entscheidung wird durch eine Link-Gleichung gebildet:

$$\forall j \in \{1, \ldots, J\} : \Pr(y = j) = \Pr(y_j^* > \max_{k \neq j} y_k^*).$$

Diese Annahme bezieht ihre Bedeutung aus der Verteilungsannahme über die Fehlerverteilung. Es ist aber zu beachten, dass die Annahme über die Fehlerverteilung grundsätzlich auch anders getroffen werden kann und von der beobachteten Entscheidung unabhängig ist. Deshalb wird diese Annahme hier getrennt dargestellt.

Annahme 4 (Kontemporäre Exogenität):
Zur Identifikation der Koeffizienten muss angenommen werden, dass für jede Alternative die Fehlerterme unabhängig von den beobachteten unabhängigen Variablen sind, d. h. formal:

$$\forall j \in \{1, \ldots, J\} : f_{\epsilon_j | X} = f_{\epsilon_j}.$$

Die Bezeichnung „kontemporäre Exogenität” ist von Wooldridge (2010, S. 164, 609) und Cameron und Trivedi (2009, S. 748f., 781) übernommen, die damit „pooled”-Modelle charakterisieren. Der Zusammenhang zu den fixed-effects-Modellen wird in Abschnitt 3.3 ausführlicher diskutiert.

Annahme 5 (Unabhängigkeit der Fehlerterme über die Alternativen):
Die idiosynkratischen Fehlerterme sind über die Alternativen hinweg paarweise unabhängig, d. h. formal:

$$\forall j, k \in \{1, \ldots, J\} : \epsilon_j \perp \epsilon_k.$$

Diese Annahme ist die wesentliche Ursache für die Eigenschaft der Unabhängigkeit von irrelevanten Alternativen (IIA) (vgl. Ben-Akiva und Lerman 1994, S. 109, Greene 2000,

3 Die Gumbel-Verteilung wird auch als Extremal-I-Verteilung bezeichnet. Die allgemeine Verteilungsfunktion ist $F_X(x) = \exp(-\exp(-(x-\mu)/\sigma))$ mit $\mathrm{E}(X) = \mu + \sigma\gamma$ und $\mathrm{Var}(X) = (\pi^2/6)\sigma^2$. γ ist keine Variable, sondern die Euler-Mascheroni-Konstante. Für weitere Details vgl. Maier und Weiss (1990, S. 135ff.) und Johnson, Kotz, und Balakrishnan (1994, Kap. 22).

S. 865). Die IIA-Eigenschaft ist für $J = 2$ natürlich hinfällig. Sie wird in der Literatur im Allgemeinen als streng und in der Regel als verletzt angesehen. Entsprechend wurden für Querschnittsdaten statistische Modelle entwickelt, die diese Annahme abschwächen bzw. ganz auf sie verzichten (vgl. Train 2009, S. 54ff.). Da aber keine Modelle für Längsschnittdaten vorliegen, die fixed-effects berücksichtigen, wird in dieser Arbeit auf eine weitergehende Darstellung dieser Modelle verzichtet.

Annahme 6 (Referenzkategorie):
Zur Identifikation muss eine Alternative B als Referenzkategorie gewählt werden, für die die unbeobachtete Heterogenität und die Regressionsparameter null gesetzt werden, d. h. formal:

$$\alpha_B = \beta_{B1} = \cdots = \beta_{BM} = 0.$$

Annahme 7 (Lineare Unabhängigkeit der Kovariatenmatrix):
Zur Identifikation muss weiterhin angenommen werden, dass die unabhängigen Variablen linear unabhängig sind, d. h. die Matrix X nicht perfekt kollinear ist, d. h. formal:

$$\text{Rang}(X'X) = M.$$

Diese lineare Unabhängigkeit ist die wesentliche Bedingung für die Invertierbarkeit der Hesse-Matrix, die für den Maximum-Likelihood-Schätzer, der weiter unten ausgeführt wird, notwendig ist.

Annahme 8 (Einfache Zufallsstichprobe):
Wie im Weiteren ausgeführt wird, werden die logit- und mlogit-Modelle mit dem Maximum-Likelihood-Verfahren geschätzt. Bis auf die bereits dargestellten Annahmen wird hierfür angenommen, dass aus der Population eine einfache Zufallsstichprobe der abhängigen und der unabhängigen Variablen $(y_i, X_i)_{i=1,\ldots,N}$ von N Beobachtungseinheiten gezogen wird. Die einfache Zufallsziehung impliziert, dass die Stichprobe unabhängig und identisch verteilt ist. Dass die Stichprobe unabhängig und identisch verteilt ist, bedeutet, dass für jedes beliebige Paar von zwei Beobachtungseinheiten aus der Stichprobe die gemeinsamen Dichtefunktionen über die abhängige und die unabhängigen Variablen paarweise gleich und paarweise unabhängig sind. D. h. formal:

$$\begin{aligned}
&\forall i, j \in \{1, \ldots, N\} : \\
&f_{(y_i, X_i)}(\cdot) = f_{(y_j, X_j)}(\cdot), \\
&f_{(y_i, X_i)}(\cdot) \perp f_{(y_j, X_j)}(\cdot).
\end{aligned}$$

Es ist anzumerken, dass diese Annahme von den anderen Annahmen unabhängig ist, d. h. die Modelle können grundsätzlich auch geschätzt werden, wenn eine komplexe Stichprobe gezogen wird. Hier müssen dann die Auswahlwahrscheinlichkeiten entsprechend berücksichtigt werden. Diese Verfahren werden in dieser Arbeit nicht weiter ausgeführt.[4] Die Implikation dieser Annahmen ist für die Auswahlwahrscheinlichkeit irrelevant, und

4 Eine aktuelle, ausführliche Darstellung findet sich z. B. bei Levy und Lemeshow (2008), siehe aber auch die Klassiker Kish (1965) und Cochran (1977) und die aktuelle, klare und konzise Ausführung von Lynn (2005).

kommt erst bei der eigentlichen Maximum-Likelihood-Schätzung zum Tragen. Zur Übersichtlichkeit wird diese Annahme gemeinsam mit den anderen Annahmen dargestellt.

Auswahlwahrscheinlichkeiten

Aus diesen Annahmen lassen sich nun die Auswahlwahrscheinlichkeiten der einzelnen Alternativen ableiten. Nach Annahme 3 ist die Wahrscheinlichkeit, dass die Alternative j gewählt wird, die Wahrscheinlichkeit, dass der Nutzen y_j^* dieser Alternative größer ist als der aller anderen Alternativen:

$$\Pr(y = j) = \Pr(y_j^* \geq \max_{k \neq j} y_k^*) = \Pr(0 \geq \max_{k \neq j} y_k^* - y_j^*).$$

Die Umformung dieses Ausdrucks erfolgt schrittweise von innen nach aussen. Aus der Annahme 2 über die Fehlerverteilung folgt aus den Reproduktionseigenschaften der Gumbelverteilung[5] für die Verteilung des Nutzens der Alternative j:

$$y_j^* \sim \text{Gumbel mit } \mathrm{E}(y_j^*) = \gamma + \alpha_j + X\beta_j' \text{ und } \mathrm{Var}(y_j^*) = \frac{\pi^2}{6}.$$

D. h. wir haben nun die Verteilung der Ausdrücke y_j^* und y_k^* in der obigen Gleichung bestimmt. Weiterhin lässt sich aus den Eigenschaften der Gumbelverteilung die Verteilung des Maximum der Nutzenterme aller Alternativen bestimmen:

$$* = \max_{k \neq j} y_k^* \sim \text{Gumbel} \quad \text{mit}$$

$$\mathrm{E}(*) = \gamma + \ln \sum_{k \neq j} \exp(\alpha_k + X\beta_k') \quad \text{und}$$

$$\mathrm{Var}(*) = \frac{\pi^2}{6}.$$

Aus den Verteilungen für diese beiden Ausdrücke folgt nun – wiederum aus den Eigenschaften der Gumbelverteilung, dass die Verteilung der Differenz der beiden Ausdrücke standard-logistisch verteilt ist. Beide Schlussfolgerungen benötigen die Annahme 5, dass die Fehler unabhängig über die Alternativen sind.

$$* = \max_{k \neq j} y_k^* - y_j^* \sim \text{Logistic} \quad \text{d. h.}$$

$$F_*(x) = \frac{1}{1 + \exp(-(x - ((\ln \sum_{k \neq j} \exp(\alpha_k + X\beta_k')) - (\alpha_j + X\beta_j'))))}$$

Dieser Ausdruck lässt sich dann weiter vereinfachen. Zuerst werden die Klammern in der Exponentialfunktion aufgelöst.

$$= \frac{1}{1 + \exp(\ln \sum_{k \neq j} \exp(\alpha_k + X\beta_k') - \alpha_j + X\beta_j' - x)}$$

5 Vgl. Maier und Weiss (1990, S. 135ff.).

Danach wird die Differenz in der Exponentialfunktion in Quotienten aufgelöst und der entstehende Doppelbruch durch Erweitern aufgelöst.

$$= \frac{1}{1 + \frac{\sum_{k \neq j} \exp(\alpha_k + X\beta'_k)}{\exp(\alpha_j + X\beta'_j + x)}}$$

$$= \frac{\exp(\alpha_j + X\beta'_j + x)}{\exp(\alpha_j + X\beta'_j + x) + \sum_{k \neq j} \exp(\alpha_k + X\beta'_k)}$$

Die Summe im Nenner lässt sich dann zu einem Ausdruck zusammenziehen.

$$= \frac{\exp(\alpha_j + X\beta'_j + x)}{\sum_{k=1}^{J} \exp(\alpha_k + X\beta'_k)} = F_{\max_{k \neq j} y_k^* - y_j^*}(x).$$

Mit dieser abgeleiteten Verteilung haben wir nun eine Darstellung für die ursprüngliche gesuchte Auswahlwahrscheinlichkeit für eine beliebige Alternative j gefunden:

$$\Pr(y = j) = \Pr(y_j^* \geq \max_{k \neq j} y_k^*) = \Pr(0 \geq \max_{k \neq j} y_k^* - y_j^*) =$$

$$= F_{\max_{k \neq j} y_k^* - y_j^*}(0) = \frac{\exp(\alpha_j + X\beta'_j)}{\sum_{k=1}^{J} \exp(\alpha_k + X\beta'_k)}.$$

Nun muss nur noch der besondere Fall der Referenzkategorie berücksichtigt werden, der, wie in Annahme 6 dargestellt wurde, für die Identifikation der Regressionskoeffizienten eingeführt werden muss:

$$\begin{aligned} \Pr(y = j) &= \frac{\exp(\alpha_j + X\beta'_j)}{1 + \sum_{k \neq B} \exp(\alpha_k + X\beta'_k)} \quad \text{für } j \neq B, \\ \Pr(y = B) &= \frac{1}{1 + \sum_{k \neq B} \exp(\alpha_k + X\beta'_k)}. \end{aligned} \tag{3.1}$$

Diese Gleichung gibt die Auswahlwahrscheinlichkeiten der Alternativen sowohl für die binäre logistische als auch für die multinomiale logistische Regression an. Man erhält die entsprechenden Gleichungen für das logit-Modell, in dem man $J = 2$ wählt.

Maximum-Likelihood-Schätzung

Die logit- und mlogit-Modelle, für die wir nun die Auswahlwahrscheinlichkeiten abgeleitet haben, werden mit dem maximum-likelihood-Verfahren geschätzt. Hierfür wird nun der Übergang vom Populationsmodell zum Stichprobenmodell vollzogen, indem aus der Population eine einfache Zufallsstichprobe der abhängigen und der unabhängigen Variablen $(y_i, X_i)_{i=1,\ldots,N}$ von N Beobachtungseinheiten gezogen wird. Wie in Annahme 8 ausgeführt wurde, heißt das, dass die Stichprobe unabhängig und identisch verteilt ist.

Weiterhin ist die Likelihood-Funktion $\ell_i(\alpha, \beta)$ für die Beobachtungseinheit i definiert als die bedingte Dichte $f_{y_i|X_i,\alpha,\beta}$. Diese diskrete Dichtefunktion ist definiert durch:[6]

$$\ell_i(\alpha, \beta) = f_{y_i|X_i,\alpha,\beta} = \prod_{j=1}^{J} \Pr(y_i = j|X_i, \alpha, \beta)^{\delta_{y_i,j}}. \tag{3.2}$$

Mit der Annahme, dass die Stichprobe unabhängig und identisch verteilt gezogen ist, lässt sich nun der Erwartungswert der logarithmierten Likelihood-Funktion der gesamten Stichprobe angeben:

$$\mathrm{L}(\alpha, \beta) = \mathrm{E}(\ln \ell_i(\alpha, \beta)) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{y_i,j} \ln \Pr(y_i = j|X_i, \alpha, \beta). \tag{3.3}$$

Das Maximum-Likelihood-Schätzverfahren liefert damit einen Schätzer für die Regressionskoeffizienten. Der geschätzte Vektor der Regressionskoeffizienten $\widehat{(\alpha, \beta)}_{\mathrm{ML}}$ ist das Maximum der Likelihood-Funktion über alle möglichen Koeffizientenwerte. Dieser Schätzer ist konsistent, d. h. er konvergiert in Wahrscheinlichkeit gegen den wahren Koeffizientenvektor (α, β). Weiterhin ist der Schätzer asymptotisch normalverteilt, d. h. er konvergiert in Verteilung gegen die Normalverteilung, wobei die Varianz durch den Erwartungswert der zweiten partiellen Ableitungen der logarithmierten Likelihood-Funktion über die Stichprobe gegeben ist. D. h. formal:

$$\widehat{(\alpha, \beta)}_{\mathrm{ML}} = \max_{(\alpha,\beta)} \mathrm{L}(\alpha, \beta), \tag{3.4a}$$

$$\widehat{(\alpha, \beta)}_{\mathrm{ML}} \xrightarrow{p} (\alpha, \beta), \tag{3.4b}$$

$$\widehat{(\alpha, \beta)}_{\mathrm{ML}} \xrightarrow{d} \mathcal{N}\Big((\alpha, \beta), \frac{-\mathrm{E}(\mathrm{H}_i(\alpha, \beta))^{-1}}{N}\Big). \tag{3.4c}$$

Hier muss hinzugefügt werden, dass mit $\mathrm{E}(\mathrm{H}_i(\alpha, \beta))$ der Erwartungswert der Hesse-Matrix über die Stichprobe dargestellt ist. Die Hesse-Matrix ist die Matrix der zweiten partiellen Ableitungen der Likelihoodfunktion nach dem Vektor der Regressionskoeffizienten. Die formale Darstellung dieses Ausdrucks ist also:

$$\mathrm{E}(\mathrm{H}_i(\alpha, \beta)) = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2 \ln \ell_i(\alpha, \beta)}{\partial(\alpha, \beta)' \partial(\alpha, \beta)} \Bigg|_{\widehat{(\alpha,\beta)}_{\mathrm{ML}}}.$$

Die wesentliche Bedingung für die Invertierbarkeit des Erwartungswerts der Hesse-Matrix über die Stichprobe ist die Annahme 7, d. h. die lineare Unabhängigkeit der Kovariatenmatrix.

Interpretation

Die geschätzten Regressionskoeffizienten können bei den Querschnitts-Modellen mit verschiedenen Techniken interpretiert werden. In ökonometrischen Lehrbüchern findet man

6 Mit $\delta_{a,b}$ ist die Kronecker-Delta-Funktion bezeichnet. Diese Funktion ist folgendermaßen definiert: $\delta_{a,b} = 1$, wenn $a = b$, und $\delta_{a,b} = 0$, wenn $a \neq b$.

im Wesentlichen zwei Interpretationsformen: den marginalen Effekt (ME) und den „discrete-change"-Effekt (DE).[7] Die Betonung liegt, wie im Folgenden noch ausführlicher diskutiert wird, auf dem Effekt auf die Auswahlwahrscheinlichkeit, und nicht auf dem Effekt auf die latente Variable. Von hauptsächlichem Interesse ist, wie die Änderung einer bestimmten unabhängigen Variablen X_m die Auswahlwahrscheinlichkeit einer bestimmten Alternativen $\Pr(y = j)$ beeinflusst. Für den Effekt von kontinuierlichen Variablen wird die Interpretation des ME empfohlen. Der ME der Variable X_m auf die Auswahlwahrscheinlichkeit der Alternative j ist folgendermaßen definiert:

$$\text{(3.5)} \qquad \begin{aligned} \text{ME}_{X_m,j} &= \left.\frac{\partial \Pr(y=j)}{\partial X_m}\right|_X \\ &= \Pr(y=j|X)\left(\widehat{\beta}_{jm} - \sum_{k\neq B} \widehat{\beta}_{km} \Pr(y=k|X)\right). \end{aligned}$$

Für zwei Alternativen ist dieser komplexe Ausdruck wesentlich einfacher. Wenn man zwei Alternativen $0 = B$ und 1 betrachtet, ist der ME der Variable X_m auf die Alternative 1:

$$\widehat{\beta}_m \Pr(y=1|X)\big(1 - \Pr(y=1|X)\big).$$

Bei diskreten unabhängigen Variablen wird die Interpretation des DE empfohlen. Allgemein ist der DE einer Variablen X_m auf die Auswahlwahrscheinlichkeit der Alternative j definiert als die Differenz der Wahrscheinlichkeit bei Veränderung von X_m um eine Einheit, d. h. formal:

$$\text{(3.6)} \qquad \begin{aligned} \text{DE}_{X_m,j} = \Pr(y=j|X_1,\ldots,X_{m-1},X_m+1,X_{m+1},\ldots,X_M) \\ - \Pr(y=j|X). \end{aligned}$$

Der Ausdruck für den DE kann – auch für nur zwei Alternativen – nicht weiter vereinfacht werden. Sowohl der ME als auch der DE einer Variable X_m auf die Auswahlwahrscheinlichkeit der Alternative j ist natürlich vom Regressionskoeffizienten $\widehat{\beta}_{jm}$ abhängig, darüber hinaus aber auch abhängig vom Wert der Variablen X_m und von den Ausprägungen der anderen Variablen X_1, ..., X_M. D. h. Effekte auf die Auswahl lassen sich nur für bestimmte Ausprägungen *aller* unabhängigen Variablen angeben.

In den ökonometrischen Standardwerken wird als Lösung für dieses Problem die Interpretation gemittelter Effekte vorgeschlagen.[8] Hier gibt es wieder zwei Möglichkeiten: Erstens kann man den ME und den DE für die Mittelwerte aller unabhängigen Variablen betrachten. Diese Effekte werden als „marginal effects at the mean" (MEM) bzw. analog „discrete-change-effect at the mean" (DEM) bezeichnet. Formal dargestellt wird der MEM einer Variablen X_m auf die Auswahlwahrscheinlichkeit der Alternative j so:

$$\text{(3.7)} \qquad \text{MEM}_{X_m,j} = \left.\frac{\partial \Pr(y=j)}{\partial X_m}\right|_{\bar{X}}$$

7 Vgl. hierzu Long (1997, S. 61–79, 164–170), Cameron und Trivedi (2009, S. 470, 501–503), Wooldridge (2010, S. 566f., 644).

8 Vgl. hierzu Long (1997, S. 74–78), Cameron und Trivedi (2009, S. 467, 492–494), Wooldridge (2010, S. 575–582), Best und Wolf (2010, S. 839f.).

$$= \Pr(y = j|\bar{X})\left(\widehat{\beta}_{jm} - \sum_{k \neq B} \widehat{\beta}_{km} \Pr(y = k|\bar{X})\right).$$

Für den Spezialfall mit nur zwei Alternativen lässt sich dieser Ausdruck wieder etwas vereinfachen. Wir betrachten wieder die Alternativen $0 = B$ und 1. Dann ist der MEM der Variable X_m auf die Auswahlwahrscheinlichkeit 1:

$$\frac{\partial \Pr(y = 1)}{\partial X_m}\bigg|_{\bar{X}} = \widehat{\beta}_m \Pr(y = 1|\bar{X})\big(1 - \Pr(y = 1|\bar{X})\big).$$

Der DEM einer Variable X_m auf die Auswahlwahrscheinlichkeit der Alternative j ist so definiert:

$$\begin{aligned}(3.8)\qquad \mathrm{DEM}_{X_m,j} &= \Pr(y = j|\bar{X}_1, \ldots, \bar{X}_{m-1}, \bar{X}_m + 1, \bar{X}_{m+1}, \ldots, \bar{X}_M) \\ &\quad - \Pr(y = j|\bar{X}).\end{aligned}$$

Dieser Ausdruck lässt sich wieder nicht weiter vereinfachen. Eine weitere Möglichkeit der Aggregation ist die Interpretation der gemittelten ME und DE. Bei dieser Interpretation werden für jede Beobachtungseinheit die Ausprägungen der unabhängigen Variablen verwendet, die die entsprechenden Beobachtungseinheit hat. Die über die Beobachtungseinheiten unterschiedlichen DE werden dann über die Stichprobe gemittelt, d. h. formal ist der sogenannte „average marginal effect" (AME) einer Variable X_m auf die Auswahlwahrscheinlichkeit der Alternative j so definiert:

$$\begin{aligned}(3.9)\qquad \mathrm{AME}_{X_m,j} &= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \Pr(y = j)}{\partial X_m}\bigg|_{X_i} \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\Pr(y = j|X_i)\left(\widehat{\beta}_{jm} - \sum_{k \neq B} \widehat{\beta}_{km} \Pr(y = k|X_i)\right)\right).\end{aligned}$$

Für mehr als zwei Alternativen lässt sich der Ausdruck nicht weiter vereinfachen, aber für nur zwei Alternativen erhält man wieder einen einfacheren Ausdruck. Wieder sind die Alternativen $0 = B$ und 1. Dann ist der AME der Variable X_m auf die Auswahlwahrscheinlichkeit 1:

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \Pr(y = 1)}{\partial X_m}\bigg|_{\bar{X}_i} = \widehat{\beta}_{jm} \frac{1}{N}\sum_{i=1}^{N}\Big(\Pr(y = 1|X_i)\big(1 - \Pr(y = 1|X_i)\big)\Big).$$

Der gemittelte DE ist analog wie der gemittelte ME definiert. Der „average discrete-change-effect" der Variable X_m auf die Auswahlwahrscheinlichkeit ist der über die Stichprobe gemittelte DE:

$$\begin{aligned}(3.10)\qquad \mathrm{ADE}_{X_m,j} &= \frac{1}{N}\sum_{i=1}^{N}\big(\Pr(y = j|X_{i1}, \ldots, X_{im-1}, X_{im}+1, X_{im+1}, \ldots, X_{iM}) \\ &\quad - \Pr(y = j|X_i)\big)\end{aligned}$$

Neben diesen Interpretationsformen, die man vorwiegend in der Ökonomie findet, sind in der Soziologie und Politikwissenschaften noch andere Interpretationsstrategien verbreitet.[9] Generell wird in diesen Disziplinen der discrete-change-Effekt dem marginalen Effekt vorgezogen, da der ME eben nur für infinitesimal kleine Veränderungen der betrachten unabhängigen Variable die richtige Differenz angibt. Wie man in den Gleichungen 3.5 und 3.6 und in Abbildung 3.1 sieht, unterscheiden sich die beiden Effekte, was aus der nicht linearen Verknüpfung zwischen der latenten Variable y_j^* und der Entscheidung y_j folgt.

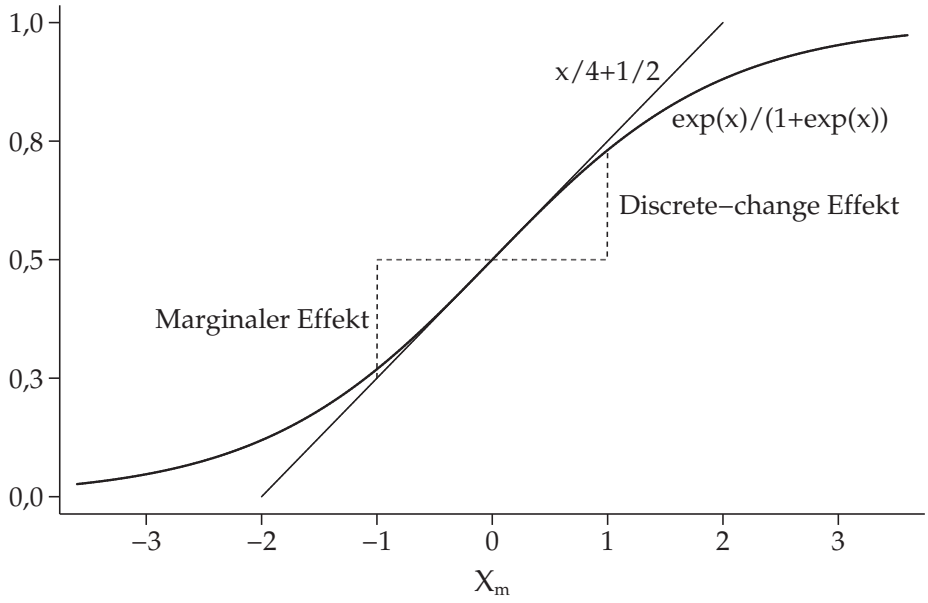


Abbildung 3.1: Discrete-change- und Marginaler Effekt

In der Abbildung 3.1 ist die einfache logistische Funktion $\exp(x)/(1+\exp(x))$ über die unabhängige Variable X_m abgebildet. Die Abbildung zeigt, wie die Auswahlscheinlichkeit $\Pr(y = 1|X)$ für zwei Alternativen von den Werten der unabhängigen Variablen X_m abhängt. In dieser Abbildung sind der DE und der ME an der Stelle $X_m = 0$ graphisch dargestellt. Der DE lässt sich so ablesen: Man geht von der Stelle $X_m = 0$, für die die Auswahlwahrscheinlichkeit hier $1/2$ ist, eine Einheit nach rechts und liest zwischen den beiden Punkten die Differenz der Auswahlwahrscheinlichkeiten ab. Diese ist in diesem Beispiel:

$$\frac{\exp(1)}{1+\exp(1)} - \frac{1}{2} \approx 0,231.$$

9 Vgl. hierzu Long (1997, S. 61–82, 164–178), Brüderl (2000, S. 631–633), Best und Wolf (2010, S. 831–833, 837–840).

Der marginale Effekt wird genauso abgelesen. Man geht von der Stelle $X_m = 0$ eine Einheit nach rechts und liest die Differenz der Punkte auf der Tangente ab. Zur einfacheren Lesbarkeit ist der Effekt in der Abbildung „negativ" dargestellt, d. h. dargestellt ist die Differenz $(\Pr(y = j|X - 1) - \Pr(y = j|X))/-1$. Dies ist beim marginalen Effekt immer möglich. Einfacher lässt sich der ME natürlich als die Steigung der Tangente an der Stelle $X_m = 0$ ablesen. Wie in der Abbildung dargestellt, ist die Tangenten-Funktion an der Stelle $X_m = 0$ $y = x/4 + 1/2$, d. h. der ME an dieser Stelle ist $1/4$, und damit etwas höher als der discrete-change-Effekt.

In der auf die Soziologie ausgerichteten Literatur wird häufig kritisiert, dass der ME keine „echte" Differenz von Wahrscheinlichkeiten ist, da er eben nur für infinitesimal kleine Differenzen der unabhängigen Variablen gilt. In dieser Lesart wird dem DE der Vorzug gegeben, da es sich hier immer um Differenzen zwischen zwei vorhergesagten Wahrscheinlichkeiten handelt. Aus dieser Perspektive heraus werden dann verschiedene Interpretationen abgeleitet, die sich aus dem Vergleich von vorhergesagten Wahrscheinlichkeiten für bestimmte Merkmalskombinationen ergeben. Der DE, bei dem die betrachtete unabhängige Variable um eine Einheit verändert wird und alle anderen Variablen an einer bestimmten Kombination fixiert werden, wird verschiedenartig verallgemeinert. Erstens werden andere Differenzen auf der unabhängigen Variable betrachtet, d. h. z. B. Differenzen zwischen dem Minimum und dem Maximum in der Stichprobe oder der Änderung der Wahrscheinlichkeit bei Änderung um eine Standardabweichung. Zweitens werden bestimmte inhaltliche bedeutsame Merkmalskombinationen über die unabhängigen Variablen gebildet, und die entsprechenden Auswahlwahrscheinlichkeiten miteinander verglichen. Die dritte, anschaulichste und am häufigsten verwendete Technik ist der „conditional effects plot", bei dem die vorhergesagte Wahrscheinlichkeit für eine bestimmte Alternative in Abhängigkeit für eine bestimmte unabhängige Variable graphisch dargestellt wird. Die anderen unabhängigen Variablen werden in der Regel entweder auf den Mittelwert in der Stichprobe oder auf inhaltlich bedeutsame Werte fixiert.

Neben diesen Interpretationsformen findet man vor allem in der Soziologie und der Politikwissenschaft zwei andere Strategien, die das Problem der Abhängigkeit des DE und ME von den Ausprägungen aller unabhängigen Variablen umgehen. Beim DE und ME wird die absolute und marginale Änderung der Auswahlwahrscheinlichkeit bei Änderung einer unabhängigen Variable betrachtet. Wenn man statt Wahrscheinlichkeiten „Odds" betrachtet, verschwindet die Abhängigkeit von allen unabhängigen Variablen. „Odds", geläufig bei professionellen Sport- und Pferdewetten, sind Verhältnisse von Wahrscheinlichkeiten. So betrachtet man z. B. bei Sportwetten meist die zwei Alternativen „Sieg" oder „Niederlage". Hier werden statt der Wahrscheinlichkeiten eines Sieges (z. B. $\Pr(\text{„Sieg"}) = 1/2$) die Odds, d. h. das Verhältnis der Sieg-Wahrscheinlichkeit und der Wahrscheinlichkeit einer Niederlage (1:1) berichtet. Bei mehreren Alternativen betrachtet man mit den Odds das Verhältnis von Auswahlwahrscheinlichkeiten von zwei beliebigen Alternativen j und k. Formal stehen Odds also folgendermaßen in Beziehung zu den Wahrscheinlichkeiten:

$$\text{Odds}_{jk}(X) = \frac{\Pr(y = j|X)}{\Pr(y = k|X)} \tag{3.11}$$

Man setzt Gleichung 3.1 ein und erhält:

$$= \frac{\frac{\exp(\widehat{\alpha}_j + X\widehat{\beta}'_j)}{1+\sum_{l\neq B}\exp(\widehat{\alpha}_l + X\widehat{\beta}'_l)}}{\frac{\exp(\widehat{\alpha}_k + X\widehat{\beta}'_k)}{1+\sum_{l\neq B}\exp(\widehat{\alpha}_l + X\widehat{\beta}'_l)}} = \frac{\exp(\widehat{\alpha}_j + X\widehat{\beta}'_j)}{\exp(\widehat{\alpha}_k + X\widehat{\beta}'_k)}$$
$$= \exp((\widehat{\alpha}_j - \widehat{\alpha}_k) + X(\widehat{\beta}_j - \widehat{\beta}_k)').$$

Zur Interpretation eines Effektes einer unabhängigen Variablen X_m betrachtet man die Änderung des Odds_{jk}, wenn sich X_m ändert. Es ist zu beachten, dass im Gegensatz zu ME und DE keine Differenzen, sondern Verhältnisse von Odds, sogenannte „Odds Ratios" (OR), betrachtet werden. D. h. formal ist der „Odds-Effekt", bzw. das Odds Ratio der Variable X_m auf das Verhältnis der Auswahlwahrscheinlichkeiten der Alternativen j zu k so definiert:

$$\text{OR}_{jk}(X_m) = \frac{\text{Odds}_{jk}(X_1, \ldots, X_{m-1}, X_m + 1, X_{m+1}, \ldots, X_M)}{\text{Odds}_{jk}(X)}$$

Man setzt die vereinfachten Ausdrücke aus Gleichung 3.11 für die Odds ein und erhält:

$$= \frac{\exp((\widehat{\alpha}_j - \widehat{\alpha}_k) + (X_m + 1)(\widehat{\beta}_{jm} - \widehat{\beta}_{km})' + X_{-m}(\widehat{\beta}_{j-m} - \widehat{\beta}_{k-m})')}{\exp((\widehat{\alpha}_j - \widehat{\alpha}_k) + X(\widehat{\beta}_j - \widehat{\beta}_k)')}$$

In der exponentierten Linearkombination im Zähler lässt sich die Differenz $\widehat{\beta}_{jm} - \widehat{\beta}_{km}$ heraus multiplizieren:

$$= \frac{\exp(\widehat{\beta}_{jm} - \widehat{\beta}_{km})\exp((\widehat{\alpha}_j - \widehat{\alpha}_k) + X(\widehat{\beta}_j - \widehat{\beta}_k)')}{\exp((\widehat{\alpha}_j - \widehat{\alpha}_k) + X(\widehat{\beta}_j - \widehat{\beta}_k)')}$$

Durch Kürzen erhält man dann den einfachen Ausdruck:

(3.12) $$\text{OR}_{jk}(X_m) = \exp(\widehat{\beta}_{jm} - \widehat{\beta}_{km}).$$

Diese Ableitung zeigt, dass das OR einer unabhängigen Variablen X_m im Gegensatz zum ME und DE von den Ausprägungen der Variablen X_m und auch aller anderen Variablen X_{-m} unabhängig ist. Für alle Werte steigt das Verhältnis der Auswahlwahrscheinlichkeiten der Alternativen j zu k um den Faktor $\exp(\widehat{\beta}_{jm} - \widehat{\beta}_{km})$, wenn die Variable X_m um eine Einheit steigt. In der Regel wird das Odds Ratio einer Alternative j zur Referenzkategorie B interpretiert. Da nach Annahme 6 die Regressionskoeffizienten für die Referenzkategorie null sind, wird so der Ausdruck noch weiter vereinfacht:

$$\text{OR}_{jB}(X_m) = \text{OR}_j(X_m) = \exp(\widehat{\beta}_{jm} - 0) = \exp(\widehat{\beta}_{jm}).$$

Verwandt mit dem Odds Radio ist der „Logit-Effekt". Auch dieser Effekt hat den Vorteil, dass er von den Ausprägungen der unabhängigen Variablen unabhängig ist. Beim Logit-Effekt betrachtet man den Einfluss der Änderung einer unabhängigen Variable X_m auf das logarithmierte Verhältnis der Auswahlwahrscheinlichkeiten der Alternativen j zu k. D. h. man betrachtet also den Effekt auf das logarithmierte Odds Ratio, was verkürzt als

Logits bezeichnet wird. Formal ist der Logit-Effekt der Variable X_m für das Verhältnis der Alternativen j zu k so definiert:

$$\text{Logit}_{jk}(X_m) = \ln\left(\frac{\Pr(y=j|X+1)}{\Pr(y=k|X+1)} \Big/ \frac{\Pr(y=j|X)}{\Pr(y=k|X)}\right) \tag{3.13}$$

Aus den Ableitungen 3.11 und 3.12 folgt:

$$= \ln(\exp(\widehat{\beta}_{jm} - \widehat{\beta}_{km})) = \widehat{\beta}_{jm} - \widehat{\beta}_{km}.$$

D. h. der Logit-Effekt der Variable X_m auf das Verhältnis der Alternativen j zu k ist die Differenz der entsprechenden Regressionskoeffizienten. Der Ausdruck vereinfacht sich analog wie oben, wenn man das Verhältnis zur Referenzkategorie B betrachtet, so dass der Logit-Effekt der Variable X_m auf das Verhältnis der Alternativen j zu B einfach der Regressionskoeffizient $\widehat{\beta}_{jm}$ ist. Es ist darauf hinzuweisen, dass die Interpretation der Regressionskoeffizienten bei den Logit-Effekten zumindest bei zwei Alternativen nicht nur auf die logarithmierten Odds Ratios, sondern eingeschränkt auch direkt auf die Auswahlwahrscheinlichkeiten möglich ist. Wie auf Seite 30 dargestellt, sind die ME und DE bei zwei Alternativen zwar abhängig von allen unabhängigen Variablen, aber proportional zum Regressionskoeffizienten. D. h. bei zwei Alternativen ist für die Regressionskoeffizienten eine Vorzeicheninterpretation auf die Änderung der Auswahlwahrscheinlichkeiten möglich. Ein positiver Regressionskoeffizient $\widehat{\beta}_{jm}$ bedeutet, dass, wenn die unabhängige Variable X_m steigt, auch die Auswahlwahrscheinlichkeit der Alternative j steigt. Dies ist aber nur bei zwei Alternativen möglich.[10]

Zusammenfassend gibt es also insgesamt acht formale Interpretationen der logit- und mlogit-Modelle: Einen Ausgangspunkt bilden der marginale Effekt ME und der discrete-change Effekt DE. Daraus werden jeweils der „marginal effect at the mean" MEM und der „average marginal effect" AME und die entsprechenden Analoga für den discrete-change Effekt DEM und ADE abgeleitet. Daneben gibt es den Odds Ratio und den Logit-Effekt. Neben diesen formalen Formen findet man die graphische Interpretation mit „conditional effects plots".

Der ME und DE und die daraus abgeleiteten Formen haben gegenüber dem Odds Ratio und Logit-Effekt den Vorteil, dass man hier direkt die Auswahlwahrscheinlichkeiten, d. h. die beobachtbaren Entscheidungen interpretiert. Der Nachteil gegenüber dem Odds Ratio und Logit-Effekt besteht darin, dass der ME und DE von den Ausprägungen aller unabhängigen Variablen abhängig ist. Im Vergleich zum Logit-Effekt hat das Odds Ratio den Vorteil, dass es als Faktor als Veränderung der Odds interpretierbar ist. Im Gegensatz dazu ist der Logit-Effekt zwar wie bei der linearen Regression sehr einfach additiv interpretierbar. Dafür ist aber das logarithmierte Odds Ratio nicht mehr anschaulich. Bei mehr als zwei Alternativen kann eine leichtfertige Interpretation der Odds Ratios und Logit-Effekte in die Irre führen. Bei einer positiven Differenz der Regressionskoeffizienten $\widehat{\beta}_{jm} - \widehat{\beta}_{km}$ steigt zwar das Verhältnis zwischen den Auswahlwahrscheinlichkeiten zwischen den Al-

10 Man beachte, dass das Problem nicht nur für die Logit-Effekte, sondern auch für die Odds-Effekte gilt.

ternativen j zu k bei steigender unabhängiger Variable X_m. Die Wahrscheinlichkeiten selbst können aber sinken.

Die Empfehlungen der besten Interpretation gehen in der Literatur auseinander. In den Standard-Ökonometrie-Lehrbüchern von Cameron und Trivedi (2009) und Wooldridge (2010) findet man fast ausschließlich Verweise auf den marginalen Effekt und den AME. Best und Wolf (2010) raten explizit von Odds Ratio Interpretationen ab, und empfehlen die graphische Interpretation und die Interpretation des AME. Bei diesen drei Quellen findet man im Bezug auf den AME auch den Hinweis auf den Vorteil, dass damit das bei Annahme 2 angesprochene Problem der „neglected heterogeneity" behoben werden kann. Im Gegensatz zu den drei genannten Quellen findet man bei Long (1997) eine ausführliche Darstellung der Interpretation des Odds Ratio.

Insgesamt ist der Empfehlung von Best und Wolf (2010) recht zu geben, dass die Schätzer bei logit- und mlogit-Modellen graphisch und mit dem AME interpretiert werden sollten. Mit diesen Techniken zielt man erstens direkt auf die leicht zugänglichen Wahrscheinlichkeiten statt auf die nur indirekt nachvollziehbaren Odds. Zweitens kann eine Logit-Effekt bzw. Odds Ratio Interpretation bei mehr als zwei Alternativen in die Irre führen. Der Schluss von positiven Odds Ratio Effekten auf einen Anstieg der Auswahlwahrscheinlichkeit ist nicht generell gültig. Der wichtigste Grund ist aber, dass man mit der Interpretation von „conditional effect plots" und der AME das Problem der „neglected heterogeneity" behoben hat, welches gerade bei Querschnitts-Modellen so gut wie immer auftritt. Sowohl die Odds Ratios als auch Logit-Effekte sind, wie in Annahme 2 dargestellt, direkt von der willkürlich gewählten Varianz der Fehlerterme abhängig.

3.2.2 Fixed-effects-modelle

Im Folgenden werden die Längsschnittmodelle felogit und femlogit vorgestellt. Die Darstellung des felogit-Modells orientiert sich an Chamberlain (1980), Greene (2000, S. 837–841), Lee (2002, S. 84–87, 92–94), Cameron und Trivedi (2009, S. 779–781, 795–797) und Wooldridge (2010, S. 494–497, 610–611, 620–625). Die Beschreibung des femlogit lehnt sich an den Ausführungen von Chamberlain (1980) und Lee (2002, S. 143–148) an. Analog zu den Querschnittsmodellen sind auch diese beiden Modelle wegen ihrer starken Verwandtschaft in einem Modell abgebildet. Die geringen Unterschiede werden wieder an entsprechender Stelle angesprochen.

Die Darstellung der Längsschnittmodelle folgt derselben Struktur wie bei den Querschnittsmodellen. Zuerst werden die Annahmen ausführlich gezeigt, dann die Auswahlwahrscheinlichkeiten aus dem Zufallsnutzen-Ansatz hergeleitet, und abschliessend der Maximum-Likelihood-Schätzer erläutert. Die Annahmen werden wie auch bei den Querschnittsmodellen vollständig und einheitlich so veranschaulicht, dass die gesamten, für die Schätzung notwendigen Annahmen deutlich werden. Dafür wird bewusst in Kauf genommen, dass sich die Ausführungen an dieser Stelle gegenüber den Erläuterungen zu den Querschnittsmodellen wiederholen. Dagegen wird die Ableitung aus dem Zufallsnutzen-Ansatz sehr knapp gehalten. Bei der Darstellung der Maximum-Likelihood-Schätzung liegt das Gewicht auf der Erläuterung des besonderen Lösungsansatzes von

Chamberlain (1980) für fixed-effects-Modelle, und der dadurch zusätzlichen notwendigen Verkomplizierung der für die Schätzung notwendigen Dichtefunktion.

Annahmen

Annahme 9 (Lineares parametrisches Modell)**:**
Für jede Beobachtungseinheit in der betrachteten Population wird angenommen, dass das folgende statistische Modell gilt. Eine Beobachtungseinheit entscheidet über eine fixe Anzahl von J Alternativen. Für jede Beobachtungseinheit liegt eine zufällige, endliche Anzahl von T Messzeitpunkten vor.[11] Für alle $j \in \{1, \ldots, J\}$ Alternativen und alle $t \in \{1, \ldots, T\}$ Messzeitpunkte gilt nun:

$$y^*_{tj} = \alpha_j + X_t \beta'_j + \epsilon_{tj}.$$

Diese Gleichung stellt im Zufallsnutzen-Kontext den Nutzen der Alternative j zum Zeitpunkt t dar. Hier sind für alle $j \in \{1, \ldots, J\}$ Alternativen und alle $t \in \{1, \ldots, T\}$ Messzeitpunkte die Ausdrücke y^*_{tj}, α_j, x_{tm} und ϵ_{tj} Zufallsvariablen. Dagegen sind alle β_{jm} konstante Parameter. Man beachte den Unterschied zu den Querschnittsmodellen, bei denen α_j wie β_{jm} ein konstanter, zu schätzender Parameter ist.

Annahme 10 (Verteilung der Fehlerterme)**:**
Für alle Alternativen und alle Messzeitpunkte ist der idiosynkratische Fehlerterm ϵ_{tj} Standard-Gumbel-verteilt, d. h. formal:

$$\forall t \in \{1, \ldots, T\}, \forall j \in \{1, \ldots, J\} :$$
$$\epsilon_{tj} \sim \text{Gumbel mit } \mathrm{E}(\epsilon_{tj}) = \gamma \text{ und } \mathrm{Var}(\epsilon_{tj}) = \frac{\pi^2}{6}.$$

Wie bei den Querschnittsmodellen ist die Festlegung der Varianz auf $\pi^2/6$ eine Annahme, die für die Identifikation der Regressionskoeffizienten notwendig ist. Auch hier ist diese Festlegung die Ursache für das Problem der „neglected heterogeneity", durch die die Effekte wesentlich schwerer interpretierbar sind. Es muss aber darauf hingewiesen werden, dass das Problem der „neglected heterogeneity" im Vergleich zu den Querschnittsmodellen geringer ist. Im Gegensatz zu den Querschnittsmodellen sind hier die Einflüsse aller zeitlich konstanten unabhängigen Variablen vollständig berücksichtigt. Dieser Unterschied wird im Abschnitt 3.3 ausführlicher diskutiert.

Annahme 11 (Verknüpfung zu beobachteter Entscheidung)**:**
Der Zusammenhang zur beobachteten Entscheidung wird für alle Messzeitpunkte durch eine Link-Gleichung gebildet:

$$\forall t \in \{1, \ldots, T\}, \forall j \in \{1, \ldots, J\} : \Pr(y_t = j) = \Pr(y^*_{tj} > \max_{k \neq j} y^*_{tk})$$

11 Für die weiteren Ausführungen ist zu beachten, dass die Anzahl der Messzeitpunkte T_i im Gegensatz zur Stichprobengröße N als Stichprobenrealisation fix ist, aber über die Beobachtungseinheiten variiert, und in der asymptotischen Herleitung der Schätzer nicht gegen Unendlich konvergiert.

Diese Annahme bezieht ihre Bedeutung aus der Verteilungsannahme über die Fehlerverteilung, dass diese erst die Abbildung der Dichte über die latente Variable y^* auf die Entscheidung y zulässt. Da die Verteilung grundsätzlich anders angenommen werden kann und von der Verknüpfung auf die Entscheidung unabhängig ist, wird diese Annahme hier getrennt dargestellt.

Annahme 12 (Strikte Exogenität)**:**
Zur Identifikation der Koeffizienten muss angenommen werden, dass für jede Alternative und jeden Messzeitpunkt die Fehlerterme unabhängig von den beobachteten unabhängigen Variablen zu allen Messzeitpunkten und unabhängig von der unbeobachteten Heterogenität sind, d. h. formal:

$$\forall t \in \{1, \ldots, T\}, \forall j \in \{1, \ldots, J\} : f_{\epsilon_{tj}|X_1,\ldots,X_T,\alpha_j} = f_{\epsilon_{tj}}.$$

Diese Annahme ist das Gegenstück zur „kontemporären Exogenität" der Querschnittsmodelle. Während bei diesen lediglich angenommen werden muss, dass der Fehlerterm nur innerhalb eines Messzeitpunktes exogen ist, muss bei den Fixed-Effects-Modellen zusätzlich angenommen werden, dass der Fehlerterm über alle Messzeitpunkte hinweg exogen ist. Diese Annahme impliziert, dass der Einfluss zeitlich verzögerter abhängiger Variablen[12] ausgeschlossen wird. Andererseits erzwingt diese Annahme, dass im Vektor der unabhängigen Variablen alle zeitverzögerten und zeitlich nachgelagerten Variablen enthalten sind, von denen ein Einfluss auf die abhängige Variable ausgeht (vgl. Wooldridge 2010, S. 494f., 610f.). Diese Annahme ist die wesentliche Grundlage dafür, dass bei Fixed-Effects-Modellen keine Annahmen über die gemeinsame Verteilung der unbeobachteten Heterogenität und den unabhängigen Variablen $f_{\alpha,X}$ getroffen werden muss, und dennoch eine konsistente Schätzung der Regressionsparameter möglich ist.

Annahme 13 (Unabhängigkeit der Fehlerterme über die Alternativen)**:**
Für alle Messzeitpunkte sind die idiosynkratischen Fehlerterme über die Alternativen hinweg paarweise unabhängig, d. h. formal:

$$\forall t \in \{1, \ldots, T\}, \forall j, k \in \{1, \ldots, J\} : \epsilon_{tj} \perp \epsilon_{tk}.$$

Wie bei den Querschnittsmodellen ist diese Annahme der Hauptgrund für die Eigenschaft der Unabhängigkeit von irrelevanten Alternativen (IIA), die bei mehr als zwei Alternativen zum Tragen kommt (vgl. Ben-Akiva und Lerman 1994, S. 109, Greene 2000, S. 865).

Annahme 14 (Serielle Unabhängigkeit)**:**
Für alle Alternativen sind die idiosynkratischen Fehlerterme über die Messzeitpunkte hinweg paarweise unabhängig, d. h. formal:

$$\forall s, t \in \{1, \ldots, T\}, \forall j \in \{1, \ldots, J\} : \epsilon_{sj} \perp \epsilon_{tj}.$$

Mit dieser Annahme wird Panelkorrelation ausgeschlossen. Diese Annahme wird in der Literatur als sehr stark und in der Regel verletzt erachtet. Zur Abhilfe wird in der Literatur empfohlen, panel-robuste Standardfehler zu schätzen (Cameron und Trivedi 2009,

12 Diese werden im Englischen „lagged dependent variables" genannt.

S. 788–791). Es ist anzumerken, dass das Problem grundsätzlich auch bei Querschnittsmodellen auftritt, wenn man diese Modelle auf gepoolte Daten anwendet. Hier wird diese Annahme in der Regel nicht explizit getroffen, sondern direkt auf panel-robuste Standardfehler verwiesen. Die Schätzung panel-robuster Standardfehler wird in Abschnitt 3.2.3 ausführlich diskutiert.

Annahme 15 (Referenzkategorie):
Zur Identifikation muss auch hier eine Alternative B als Referenzkategorie gewählt werden, für die die unbeobachtete Heterogenität und die Regressionsparameter null gesetzt werden. Im Gegensatz zu den Querschnittsmodellen ist hier die unbeobachtete Heterogenität α_B eine Zufallsvariable, die auf eine Konstante gesetzt werden muss, und kein zu schätzender Parameter. Formal heißt das:

$$\alpha_B = \beta_{B1} = \cdots = \beta_{BM} = 0.$$

Annahme 16 (Lineare Unabhängigkeit der Kovariatenmatrix):
Eine der wesentlichen Voraussetzungen für die Identifikation der Schätzer ist die Invertierbarkeit des Erwartungswertes der Hesse-Matrix. Bei den Querschnittsmodellen muss dafür die Kovariatenmatrix linear unabhängig sein. Bei Längsschnittmodellen gilt in ähnlicher Weise, dass die unabhängigen Variablen über die Messzeitpunkte gemittelt nicht perfekt kollinear sein dürfen, d. h. formal:

$$\text{Rang}\left((X - \bar{X})'(X - \bar{X})\right) = M, \quad \text{mit} \quad \bar{X} = \left(\frac{1}{T}\sum_{t=1}^{T} x_{tm}\right).$$

Die Matrix $\bar{X}$ ist also die Matrix der Mittelwerte der Einträge der Matrix X über die Messzeitpunkte T. Man beachte, dass $X - \bar{X}$ für jede Beobachtungseinheit i folgende Matrix ist:

$$\forall i \in \{1, \ldots, N\} : X - \bar{X} = \begin{pmatrix} x_{i11} - \sum_{t=1}^{T_i} x_{it1} & \cdots & x_{i1M} - \sum_{t=1}^{T_i} x_{itM} \\ \vdots & \ddots & \vdots \\ x_{iT_i1} - \sum_{t=1}^{T_i} x_{it1} & \cdots & x_{iT_iM} - \sum_{t=1}^{T_i} x_{itM} \end{pmatrix}.$$

Annahme 17 (Einfache Zufallsstichprobe):
Auch die Längsschnittmodelle werden mit dem Maximum-Likelihood-Verfahren geschätzt. Wieder wird angenommen, dass aus der Population eine einfache Zufallsstichprobe der abhängigen und der unabhängigen Variablen $(y_i, X_i)_{i=1,\ldots,N}$ von N Beobachtungseinheiten gezogen wird. Man beachte, dass für jede Beobachtungseinheit eine Messreihe der abhängigen und unabhängigen Variablen über die Messzeit T_i gezogen wird. Die Anzahl der Messzeitpunkte T_i ist selbst Ergebnis einer Zufallsstichprobe, wobei T unabhängig von y, X und α ist. Die einfache Zufallsziehung impliziert wieder, dass die Stichprobe $((y_{i1}, \ldots, y_{iT_i}), (X_{i1}, \ldots, X_{iT_i}))_{i=1,\ldots,N}$ unabhängig und identisch verteilt ist, d. h. formal:

$$\forall i, j \in \{1, \ldots, N\} :$$
$$f_{(y_i, X_i)}(\cdot) = f_{(y_j, X_j)}(\cdot),$$

$$f_{(y_i,X_i)}(\cdot) \perp f_{(y_j,X_j)}(\cdot).$$

Die Annahme ist wie bei den Querschnittsmodellen nicht von zentraler Bedeutung für die Schätzung der Regressionskoeffizienten. Bei einer komplexen Stichprobe muss nur für die Auswahlwahrscheinlichkeiten der Beobachtungseinheiten entsprechend korrigiert werden (vgl. Kish 1965; Cochran 1977; Lynn 2005; Levy und Lemeshow 2008). Es ist zu beachten, dass mit dieser Annahme grundsätzlich sowohl „unbalanced panel"-Daten als auch „balanced panel"-Daten berücksichtigt werden können. Der erste Fall liegt bei Gültigkeit der hier dargestellten Annahme vor. Der zweite Fall wird dadurch realisiert, dass die Anzahl der Messzeitpunkte für alle Beobachtungseinheiten gleich gesetzt werden. Weiterhin wird mit dieser Annahme ausgeschlossen, dass bei „unbalanced panel"-Daten selektiver Ausfall vorliegt. Das Problem selektiver Stichproben und von selektivem Attrition wird in Abschnitt 3.2.3 ausführlich diskutiert.

Auswahlwahrscheinlichkeiten

Da die Auswahlwahrscheinlichkeiten der einzelnen Alternativen analog wie bei den Querschnittsmodellen abgeleitet werden, wird die Darstellung hier bewusst sehr kurz gehalten. Wie bei den Querschnittsmodellen ist nach Annahme 11 die Wahrscheinlichkeit, dass die Alternative j gewählt wird, die Wahrscheinlichkeit, dass der Nutzen y_j^* dieser Alternative größer ist als der aller anderen Alternativen. Dies gilt nun für jeden Messzeitpunkt:

$$\forall t \in \{1,\ldots,T\} : \Pr(y_t = j) = \Pr(y_{tj}^* \geq \max_{k \neq j} y_{tk}^*) = \Pr(0 \geq \max_{k \neq j} y_{tk}^* - y_{tj}^*).$$

Aus dieser Gleichung werden unter Berücksichtigung der oben dargestellten Annahmen die Auswahlwahrscheinlichkeiten für jede Alternative für jeden Messzeitpunkt – wie bei den Querschnittsmodellen ausführlich dargestellt wurde – abgeleitet:

$$\begin{aligned} &\forall t \in \{1,\ldots,T\} : \\ &\Pr(y_t = j) = \frac{\exp(\alpha_j + X_t\beta_j')}{1 + \sum_{k \neq B} \exp(\alpha_k + X_t\beta_k')} \quad \text{für } j \neq B, \\ &\Pr(y_t = B) = \frac{1}{1 + \sum_{k \neq B} \exp(\alpha_k + X_t\beta_k')}. \end{aligned} \tag{3.14}$$

Diese Gleichung gibt die Auswahlwahrscheinlichkeiten der Alternativen sowohl für das felogit als auch das femlogit-Modell an. Man erhält die entsprechenden Gleichungen für das felogit-Modell, in dem man $J = 2$ wählt. Diese Auswahlwahrscheinlichkeit ist dann die Grundlage für die Maximum-Likelihood, die im Folgenden ausführlich dargestellt wird.

Maximum-Likelihood-Schätzung

Wie bei den Querschnittsmodellen werden auch hier die Regressionskoeffizienten mit dem Maximum-Likelihood-Verfahren geschätzt. Aus der Population wird eine einfache Zufallsstichprobe $(y_i, X_i)_{i=1,\ldots,N}$ von N Beobachtungseinheiten gezogen. Wie in An-

nahme 17 ausgeführt, wird im Gegensatz zu den Querschnittsmodellen für jede dieser Beobachtungseinheiten eine Zeitreihe von T_i Messzeitpunkten gezogen, wobei T_i eine Realisation einer unabhängigen Zufallsvariablen ist. Aus der einfachen Zufallsstichprobe folgt, dass die Stichprobe über die Beobachtungseinheiten hinweg unabhängig und identisch verteilt ist. Definitionsgemäß ist die Likelihood-Funktion $\ell_i(\beta)$ für die Beobachtungseinheit i die bedingte Dichte $f_{y_i|X_i,\alpha,\beta}$. Diese diskrete Dichtefunktion ergibt sich aus der Gleichung 3.14 über die Auswahlwahrscheinlichkeiten zu jedem Messzeitpunkt für jede Alternative:

$$\widetilde{\ell}_i(\alpha,\beta) = f_{y_i|X_i,\alpha,\beta} = \prod_{t=1}^{T_i}\prod_{j=1}^{J} \Pr(y_{it}=j|X_i,\alpha,\beta)^{\delta_{y_{it},j}}. \tag{3.15}$$

Grundsätzlich lässt sich auch bei den Längsschnittmodellen mit der Annahme, dass die Stichprobe unabhängig und identisch verteilt gezogen ist, der Erwartungswert dieser logarithmierten Likelihood-Funktion für die gesamte Stichprobe angeben:

$$\widetilde{\mathrm{L}}(\alpha,\beta) = \mathrm{E}(\ln \widetilde{\ell}_i(\alpha,\beta)) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{j=1}^{J} \delta_{y_{it},j} \ln \Pr(y_{it}=j|X_i,\alpha,\beta).$$

Im Gegensatz zu den Querschnittsmodellen kann aber mit dieser Gleichung nicht wie bei den Querschnittsmodellen mit dem Maximum-Likelihood-Verfahren die Regressionskoeffizienten geschätzt werden, da α eine unbeobachtete Zufallsvariable ist, die potentiell mit den unabhängigen Variablen X korreliert ist. Bei linearen fixed-effects-Modellen wird dieses Problem im Wesentlichen durch zwei Ansätze gelöst. Man geht von den gegebenen abhängigen und unabhängigen Variablen y und X auf die Differenzen zu den Mittelwerten über die Zeit innerhalb einer Beobachtungseinheit über. D. h. statt y und X wird $y-\bar{y}$ und $X-\bar{X}$ betrachtet. Auf diese Weise fällt die zeitlich konstante Heterogenität heraus. Ein zweiter Ansatz besteht darin, die individuell spezifischen Heterogenitäten α_i explizit zu schätzen, indem in das statistische Modell Dummies für die Beobachtungseinheiten einbezogen werden.[13] Bei nicht linearen Modellen wie die in dieser Arbeit betrachteten Modelle kommen beide Lösungsansätze nicht in Frage. Eine Lösung dieses Problems findet sich bei Chamberlain (1980). Er nutzt eine suffiziente Statistik für α_i, mit deren Hilfe ein konsistenter Schätzer für die Regressionskoeffizienten abgeleitet werden kann. Da weder bei Chamberlain noch in den anderen Quellen, in denen zumindest das felogit-Modell erläutert wird (Greene 2000; Lee 2002; Cameron und Trivedi 2009; Wooldridge 2010), eine ausführliche Darstellung dieser Ableitung zu finden ist, wird dies im Folgenden nachgeholt.[14]

Wie bereits angedeutet wurde, nutzt Chamberlain (1980) für seine Lösung eine suffiziente Statistik für die unbeobachtete Heterogenität α. Zum besseren Verständnis soll das Konzept einer suffizienten Statistik erläutert werden.[15] Zunächst einmal ist eine Statistik

13 Zu Identifikation muss eine Beobachtungseinheit als Referenzkategorie betrachtet werden, d. h. es werden $N-1$ Dummies berücksichtigt.

14 Die Darstellung baut auf Cameron und Trivedi (2009, S. 798f.) auf.

15 Vgl. hierzu ausführlicher Maier und Weiss (1990, S. 280f.) und Cameron und Trivedi (2009, S. 782).

k_{ij} in diesem Sinne eine Funktion über Datenpunkte – in der hier dargestellten Anwendung über die Messreihe $(y_{i1}, \ldots, y_{iT})$. Eine Statistik k_{ij} ist suffizient für einen Parameter α_{ij}, wenn die Verteilung der Datenpunkte, auf der die Statistik beruht, bedingt für diese Statistik unabhängig von dem entsprechenden Parameter ist. Die Dichtefunktion der Datenpunkte bedingt für die suffiziente Statistik hängt also nicht mehr von dem Parameter ab, d. h. formal in Bezug auf hier dargestellte Anwendung:

$$f_{y_i|k_i,\alpha_i} = f_{y_i|k_i}.$$

Aus dieser bedingten Dichtefunktion lässt sich gleichzeitig ein konsistenter Schätzer für die Regressionskoeffizienten bilden. Inhaltlich lässt sich die Suffizienz so interpretieren, dass in der suffizienten Statistik die gesamte Information bzgl. des entsprechenden Parameters steckt, die in der Stichprobe enthalten ist. Die Verteilung der Stichprobe bedingt für diese Statistik enthält also keine zusätzliche Information über den Parameter mehr.

Chamberlain (1980) nutzt in seinem Lösungsansatz, dass die Häufigkeit, mit der eine Alternative j über die Messzeit hinweg gewählt wurde, $\sum_{t=1}^{T_i} \delta_{y_{it},j}$ eine suffiziente Statistik für α_{ij} ist. D. h. die Wahrscheinlichkeit der Realisation einer bestimmten Zeitreihe y_i, bedingt für diese suffiziente Statistik, ist unabhängig von der unbeobachteten Heterogenität α_{ij}. Bevor dies detailliert ausgeführt wird, muss zunächst weitere Notation eingeführt und erläutert werden. Wir betrachten eine beliebige Beobachtungseinheit i. Für diese Person haben wir den Vektor der unbeobachteten Heterogenitäten $(\alpha_{i1}, \ldots, \alpha_{iJ})$ bereits definiert. Gemäß Annahme 15 haben wir für die Alternative B die entsprechende Heterogenität $\alpha_{iB} = 0$ gesetzt. D. h. relevant ist nur der eingeschränkte Vektor $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{iB-1}, \alpha_{iB+1}, \ldots, \alpha_{iJ})$. Für diesen Vektor sei nun der Vektor der entsprechenden suffizienten Statistiken gegeben durch $k_i = (k_{i1}, \ldots, k_{iB-1}, k_{iB+1}, \ldots, k_{iJ})$. Wie oben ist jede suffiziente Statistik definiert durch $k_{ij} = \sum_{t=1}^{T_i} \delta_{y_{it},j}$. Weiterhin sei nun der Vektor d_i definiert als $(d_{i1}, \ldots, d_{iT_i})$. d_i entspricht in seiner Struktur der abhängigen Variable y_i. Für eine suffiziente Statistik k_{ij} ist d_i der Vektor, für den gilt: $\sum_{t=1}^{T_i} \delta_{d_{it},j} = k_{ij}$. Inhaltlich ist d_i also ein Vektor, in dem die Alternative j genauso oft gewählt wird wie bei der abhängigen Variable y_i. Ferner sei mit Δ_i die Menge aller Vektoren d_i bezeichnet, d. h. formal:

$$\Delta_i = \Big\{ (d_{i1}, \ldots, d_{iT_i})' | \forall j = 1, \ldots, J, j \neq B : \sum_{t=1}^{T_i} \delta_{d_{it},j} = k_{ij} \Big\}.$$

Inhaltlich lässt sich Δ_i als die Menge aller Permutation d_i der unabhängigen Variable y_i über die Messzeit interpretieren. Bei allen Permutationen ist die Häufigkeit, mit der eine bestimmte Alternative j gewählt wird, gleich, aber ihre Reihenfolgen, zu welchem Messzeitpunkt die entsprechende Alternative gewählt wurde, sind unterschiedlich.

Abschließend soll diese Notation an einem Beispiel veranschaulicht werden. In Anlehnung an Abbildung 1.3 betrachte man eine Beobachtungseinheit mit $T_i = 3$, $J = 3$ und $y_i = (1, 2, 3)$. Das heisst, jede der drei Alternativen wird einmal gewählt. Wir betrachten die erste Alternative als Referenzkategorie: $B = 1$. Damit ist die suffiziente Statistik für die zweite Alternative $k_{i2} = 1$ und entsprechend $k_{i3} = 1$. Die Permutationsmatrix Δ_i ist

dann die Menge aller möglichen Zeitreihen mit drei Messzeitpunkten und drei Alternativen, bei denen die zweite und dritte Alternative jeweils genau einmal gewählt werden: $\{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$.

Mit dieser zusätzlichen Notation wird nun dargestellt, dass die Dichtefunktion $f_{y_i|X_i,\alpha_i,\beta}$, bedingt für die suffiziente Statistik k_i, unabhängig von der unbeobachteten Heterogenität ist. Man betrachte eine Beobachtungseinheit i mit y_i und X_i, wie bereits definiert, und den unbeobachteten Heterogenitäten α_i. Die bedingte Dichte, von der wir ausgehen, lässt sich nach dem Satz von Bayes folgendermaßen umformen:[16]

$$f_{y_i|k_i,\alpha_i} = \frac{f_{(y_i \wedge k_i)|\alpha_i}}{f_{k_i}}.$$

Da bei gegebenem y_i auch k_i vollständig determiniert ist, fällt dieser Term aus der Gleichung heraus. D. h. es gilt:

$$\frac{f_{(y_i \wedge k_i)|\alpha_i}}{f_{k_i}} = \frac{f_{y_i|\alpha_i}}{f_{k_i}}.$$

Die Dichtefunktion im Nenner gibt die Wahrscheinlichkeiten an, dass eine Zeitreihe realisiert wird, die der suffizienten Statistik k_i genügt. D. h. im Nenner steht die Summe über alle Permutation von y_i:

$$\frac{f_{y_i|\alpha_i}}{f_{k_i}} = \frac{f_{y_i|\alpha_i}}{\sum_{d_i \in \Delta_i} f_{d_i}}.$$

Die Dichte über d_i bezeichnet die Wahrscheinlichkeit, dass eine bestimmte Zeitreihe realisiert wird. Aus den Definitionen von y_i und d_i und Gleichung 3.15 für die unbedingte Dichtefunktion der Zeitreihe folgt:

$$\frac{f_{y_i|\alpha_i}}{\sum_{d_i \in \Delta_i} f_{d_i}} = \frac{\prod_{t=1}^{T_i} \prod_{j=1}^{J} \Pr(y_{it} = j)^{\delta_{y_{it},j}}}{\sum_{d_i \in \Delta_i} \prod_{t=1}^{T_i} \prod_{j=1}^{J} \Pr(d_{it} = j)^{\delta_{d_{it},j}}}$$

Wenn man nun hier die Auswahlwahrscheinlichkeiten für jede Alternative zu jedem Messzeitpunkt nach Gleichung 3.14 einsetzt, erhält man:

$$= \frac{\prod_{t=1}^{T_i} \prod_{j=1}^{J} \left(\frac{\exp(\alpha_{ij} + X_{it}\beta_j')}{1 + \sum_{k=1,k\neq B}^{J} \exp(\alpha_{ik} + X_{it}\beta_k')} \right)^{\delta_{y_{it},j}}}{\sum_{d_i \in \Delta_i} \left(\prod_{t=1}^{T_i} \prod_{j=1}^{J} \left(\frac{\exp(\alpha_{ij} + X_{it}\beta_j')}{1 + \sum_{k=1,k\neq B}^{J} \exp(\alpha_{ik} + X_{it}\beta_k')} \right)^{\delta_{d_{it},j}} \right)}$$

Man beachte, dass bei beiden inneren Zählern in den beiden Brüchen die Referenzkategorie aus Platzgründen nicht berücksichtigt ist.

Um zu zeigen, dass k_i ein suffizienter Schätzer für die unbeachteten Heterogenitäten α_i ist, muss dieser Ausdruck so umgeformt werden, dass α_i aus diesem Ausdruck verschwindet. Zunächst lassen sich die Nenner der beiden Brüche aus dem Ausdruck heraus kürzen. Der Term $1 + \sum_{k=1,k\neq B}^{J} \exp(\alpha_{ik} + X_{it}\beta_k')$ ist unabhängig von den Alternativen

16 Vgl. Bronstein, Semendjajew, Musiol, und Mühlig (2001, S. 771–773).

j. D. h. unabhängig davon, welche Alternative zum Zeitpunkt t gewählt wird, wird der von j abhängige Zähler genau einmal durch den unabhängigen Nenner dividiert:

$$= \frac{\prod_{t=1}^{T_i} \frac{\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{y_{it},j}}}{1+\sum_{k=1,k\neq B}^{J}\exp(\alpha_{ik}+X_{it}\beta'_k)}}{\sum_{d_i\in\Delta_i}\prod_{t=1}^{T_i}\frac{\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{d_{it},j}}}{1+\sum_{k=1,k\neq B}^{J}\exp(\alpha_{ik}+X_{it}\beta'_k)}}$$

Das Produkt für die Messzeitpunkte lässt sich nun in die Zähler und Nenner der beiden Brüche verschieben, so dass man unabhängige Faktoren erhält:

$$= \frac{\frac{\prod_{t=1}^{T_i}\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\prod_{t=1}^{T_i}(1+\sum_{k=1,k\neq B}^{J}\exp(\alpha_{ik}+X_{it}\beta'_k))}}{\sum_{d_i\in\Delta_i}\frac{\prod_{t=1}^{T_i}\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{d_{it},j}}}{\prod_{t=1}^{T_i}(1+\sum_{k=1,k\neq B}^{J}\exp(\alpha_{ik}+X_{it}\beta'_k))}}$$

Die Summe im Nenner läuft über Brüche mit jeweils dem gleichen Nenner, der von den einzelnen Summandenindizes d_i unabhängig ist. D. h. der Nenner lässt sich aus der Summe im Nenner ausmultiplizieren, so dass man dies erhält:

$$= \frac{\frac{\prod_{t=1}^{T_i}\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\prod_{t=1}^{T_i}(1+\sum_{k=1,k\neq B}^{J}\exp(\alpha_{ik}+X_{it}\beta'_k))}}{\frac{\sum_{d_i\in\Delta_i}\prod_{t=1}^{T_i}\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{d_{it},j}}}{\prod_{t=1}^{T_i}(1+\sum_{k=1,k\neq B}^{J}\exp(\alpha_{ik}+X_{it}\beta'_k))}}$$

Hier lassen sich nun die beiden Nenner der beiden Brüche kürzen, so dass sich der Doppelbruch auflöst.[17] Der Ausdruck lässt sich also zu diesem Term vereinfachen:

$$= \frac{\prod_{t=1}^{T_i}\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\prod_{t=1}^{T_i}\prod_{j=1}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{d_{it},j}}}$$

Nun können wir die zuvor aus Platzgründen eingeführte Vereinfachung, die Referenzkategorie in der Alternativenmenge zu ignorieren, aufgeben und den Bruch der ehemaligen Zähler des Doppelbruchs vollständig fassen. Da die Referenzkategorie in den ehemaligen Zählern nur eine Multiplikation mit dem Faktor Eins zu den Brüchen im Zähler und Nenner beitragen, verändert sich gegenüber dem vorherigen Ausdruck nur der Laufindex im zweiten Produkt über die Alternativen:

$$= \frac{\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}+X_{it}\beta'_j))^{\delta_{d_{it},j}}}$$

17 Die Gefahr, dass diese Nenner für bestimmte Parameterkombinationen Null sein könnte, ist nicht gegeben, da die Exponentialfunktion immer positiv ist, und somit eine Summe von positiven Zahlen und Eins immer größer Null ist.

Danach kann man das eigentliche Ziel, die unbeobachtete Heterogenität aus diesem Ausdruck zu kürzen, angehen. Zuerst wird hierfür die Linearkombination $\alpha_{ij} + X_{it}\beta'_j$ aufgetrennt, um so getrennte exponentierte Faktoren zu erhalten:

$$= \frac{\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\delta_{y_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\left(\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\delta_{d_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{d_{it},j}}\right)}$$

Dann wird beim ersten Faktor im Zähler und im Nenner über α die Produktreihenfolge vertauscht. D. h. die Produkte laufen erst über die Alternativen, und dann für jede einzelne Alternative über die Messzeitpunkte:

$$= \frac{\prod_{j=1,j\neq B}^{J}\prod_{t=1}^{T_i}(\exp(\alpha_{ij}))^{\delta_{y_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\left(\prod_{j=1,j\neq B}^{J}\prod_{t=1}^{T_i}(\exp(\alpha_{ij}))^{\delta_{d_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{d_{it},j}}\right)}$$

Wir betrachten nun Produkte über $(\exp(\alpha_{ij}))^{\delta_{y_{it},j}}$ bzw. $(\exp(\alpha_{ij}))^{\delta_{d_{it},j}}$. Für jede einzelne beliebige Alternative j ist der Term α_{ij} über die Messzeitpunkte t gleich. D. h. das Produkt über die Messzeitpunkte innerhalb einer bestimmten Alternative lässt sich als Potenz schreiben:

$$= \frac{\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\sum_{t=1}^{T_i}\delta_{y_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\left(\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\sum_{t=1}^{T_i}\delta_{d_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{d_{it},j}}\right)}$$

Der oben dargestellten Notationsdefinition von Δ_i und d_i folgend ist $\sum_{t=1}^{T_i}\delta_{y_{it},j} = k_{ij} = \sum_{t=1}^{T_i}\delta_{d_{it},j}$. Dies lässt sich in den Ausdruck einsetzen, so dass man erhält:

$$= \frac{\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\sum_{t=1}^{T_i}\delta_{y_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\left(\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\sum_{t=1}^{T_i}\delta_{y_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{d_{it},j}}\right)}$$

Hier ist nun der erste Faktor im Nenner von den Laufindizes d_i unabhängig, d. h. der Faktor kann aus der Summe über Δ_i heraus multipliziert werden:

$$= \frac{\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\sum_{t=1}^{T_i}\delta_{y_{it},j}}\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\left(\prod_{j=1,j\neq B}^{J}(\exp(\alpha_{ij}))^{\sum_{t=1}^{T_i}\delta_{y_{it},j}}\right)\sum_{d_i\in\Delta_i}\left(\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{d_{it},j}}\right)}$$

Bei diesem Bruch kann man jeweils den ersten Faktor im Zähler und Nenner kürzen, so dass der Ausdruck von den unbeobachteten Heterogenitäten α_i unabhängig ist.

$$= \frac{\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{y_{it},j}}}{\sum_{d_i\in\Delta_i}\left(\prod_{t=1}^{T_i}\prod_{j=1,j\neq B}^{J}(\exp(X_{it}\beta'_j))^{\delta_{d_{it},j}}\right)}.$$

Mit dieser Ableitung haben wir also gezeigt, dass die bedingte Dichte $f(y_i|k_i,\alpha_i)$ unabhängig von unbeobachteten Heterogenität α_i ist, d. h. k_i ist eine suffiziente Statistik für α_i.

Der entscheidende Vorteil ist, dass uns diese Ableitung neben dem Beweis über die Suffizienz gleichzeitig den Ausgangspunkt für einen konsistenten Maximum-Likelihood-Schätzer der Regressionskoeffizienten β liefert. Dies ist der wesentliche Beitrag von Chamberlain (1980), der erst eine fixed-effects-Modellierung für diskrete abhängige Variablen erlaubt.

Für die Entwicklung eines Maximum-Likelihood-Schätzers, der das Problem der unbeobachteten Heterogenität umgeht, betrachteten wir die bedingte Dichtefunktion als Likelihood-Funktion. Für die Ableitung des Schätzers wird der Ausdruck noch numerisch besser umgeformt, d. h. die Produkte werden in die Exponentialfunktion gezogen, und so zu Summen umgeformt. Dabei werden die Exponenten $\delta_{y_{it},j}$, bzw. $\delta_{d_{it},j}$ den Potentierungsregeln folgend zu Faktoren. Die Likelihood-Funktion für die in dieser Arbeit betrachteten Längsschnittmodelle ist also:

$$\ell_i(\beta) = f_{y_i|k_i,\alpha_i} = \frac{\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it},j}X_{it}\beta_j')}{\sum_{d_i\in\Delta_i}\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{d_{it},j}X_{it}\beta_j')}. \tag{3.16}$$

Wie bei den Querschnittsmodellen folgt aus der Annahme 17, dass die Stichprobe unabhängig und identisch verteilt gezogen ist, für den Erwartungswert der logarithmierten Likelihood-Funktion für die gesamte Stichprobe:

$$\begin{aligned}
\mathrm{L}(\beta) &= \mathrm{E}(\ln \ell_i(\beta)) \\
&= \frac{1}{N}\sum_{i=1}^{N}\ln\frac{\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it},j}X_{it}\beta_j')}{\sum_{d_i\in\Delta_i}\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{d_{it},j}X_{it}\beta_j')} \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i}\sum_{\substack{j=1,\\ j\neq B}}^{J}\delta_{y_{it},j}X_{it}\beta_j' - \ln\sum_{d_i\in\Delta_i}\exp\left(\sum_{t=1}^{T_i}\sum_{\substack{j=1,\\ j\neq B}}^{J}\delta_{d_{it},j}X_{it}\beta_j'\right)\right).
\end{aligned} \tag{3.17}$$

Wie bei den Querschnittsmodellen ist der Schätzer $\widehat{\beta}_{\mathrm{ML}}$ das Maximum der Likelihood-Funktion über alle möglichen Koeffizientenwerte. Dieser Schätzer ist konsistent und asymptotisch normalverteilt. Die Varianz ergibt sich wieder aus dem Erwartungswert der Hesse-Matrix. D. h. formal:

$$\widehat{\beta}_{\mathrm{ML}} = \max_{\beta}\mathrm{L}(\beta), \tag{3.18a}$$

$$\widehat{\beta}_{\mathrm{ML}} \xrightarrow{p} \beta, \tag{3.18b}$$

$$\widehat{\beta}_{\mathrm{ML}} \xrightarrow{d} \mathcal{N}\left(\beta, \frac{-\mathrm{E}(\mathrm{H}_i(\beta))^{-1}}{N}\right). \tag{3.18c}$$

Der Ausdruck $\mathrm{E}(\mathrm{H}(\beta))$ ist der Erwartungswert der Hesse-Matrix über die Stichprobe. Die Hesse-Matrix ist wieder die Matrix der zweiten partiellen Ableitungen der Likelihoodfunktion nach dem Vektor der Regressionskoeffizienten, d. h. formal:

$$\mathrm{E}(\mathrm{H}_i(\beta)) = \frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial^2\ln\ell_i(\beta)}{\partial\beta'\partial\beta}\right|_{\widehat{\beta}_{\mathrm{ML}}}.$$

Die Likelihood-Funktion in Gleichung 3.17 und der Schätzer in Gleichung 3.18a gelten für den allgemeinen Fall des femlogit-Modells für beliebig viele Alternativen J. Entsprechend erhält man den Spezialfall felogit mit nur zwei Alternativen, indem man $J = 2$ setzt.

Anmerkung zur abhängigen Variablen

Bei den hier dargestellten femlogit- und felogit-Schätzern wird vorausgesetzt, dass bei abhängigen Variablen Varianz über die Messzeitpunkte vorliegt. Für Beobachtungseinheiten, bei denen sich die abhängige Variable nicht ändert, kann kein Schätzer identifiziert werden. Außerdem muss es wie bei den Querschnittsmodellen für jede Alternative j mindestens eine Beobachtungseinheit geben, die diese Alternative gewählt hat.[18]

Beobachtungseinheiten, bei denen keine Varianz auf der abhängigen Variablen über die Messzeitpunkte vorliegt, sind für die Regressionskoeffizienten keiner Alternative informativ. Dazu betrachte man formal eine beliebige Beobachtungseinheit i mit der realisierten Zeitreihe der abhängigen Variablen $y_i = (y_{i1}, \ldots, y_{iT_i})$. Für die betrachtete Beobachtungseinheit liege nun keine Varianz auf der abhängigen Variablen über die Messzeitpunkte vor, d. h. $y_{i1} = y_{i2} = \cdots = y_{iT_i}$. Wir betrachten nun für diese Beobachtungseinheit die Realisationswahrscheinlichkeit einer beliebigen Zeitreihe $\tilde{y}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{iT_i})$, die nach Gleichung 3.16 definiert ist als:

$$f_{y_i|k_i,\alpha_i}(\tilde{y}_i) = \Pr(y_i = \tilde{y}_i) = \frac{\exp(\sum_{t=1}^{T_i} \sum_{j=1,j\neq B}^{J} \delta_{\tilde{y}_{it},j} X_{it}\beta'_j)}{\sum_{\tilde{d}_i \in \Delta_i} \exp(\sum_{t=1}^{T_i} \sum_{j=1,j\neq B}^{J} \delta_{\tilde{d}_{it},j} X_{it}\beta'_j)}.$$

Da keine Varianz auf der abhängigen Variablen über die Messzeit vorliegt, hat für die Beobachtungseinheit die Menge der Permutationen Δ_i nur ein Element, nämlich die realisierte Zeitreihe y_i. Daraus folgt, dass die Wahrscheinlichkeit für eine beliebige Zeitreihe zu einer Indikatorfunktion wird:

$$\Pr(y_i = \tilde{y}_i) = \begin{cases} 1 & \tilde{y}_i = y_i \\ 0 & \tilde{y}_i \neq y_i \end{cases}.$$

Wenn eine bestimmte Alternative h von keiner Beobachtungseinheit zu keinem Messzeitpunkt gewählt wird, taucht die Alternative auch für keine Beobachtungseinheit in der zugehörigen Permutationsmenge Δ_i auf, d. h. für alle Beobachtungseinheiten gilt: $\forall i \in \{1, \ldots, N\} : k_{ih} = 0$. D. h. die bedingte Realisationswahrscheinlichkeit nach Gleichung 3.16 für Zeitreihen mit $k_{ih} \neq 0$ ist immer null, und nur für Zeitreihen mit $k_{ih} = 0$ zumindest potentiell verschieden von null. Damit taucht der Vektor der Regressionskoeffizienten β_h nicht in der logarithmierte Likelihoodfunkion für die Gesamtstichprobe in Gleichung 3.16 auf. D. h. jeder beliebige Wert für β_h maximiert die logarithmierte Likelihoodfunkion, was wiederum bedeutet, dass β_h nicht eindeutig identifiziert werden kann.

Dieser Aspekt ist Ursache dafür, dass fixed-effects-Modelle stark dafür kritisiert werden, dass wertvolle Information in den Daten nicht genutzt wird. Dies wird in Abschnitt 3.3 ausführlicher diskutiert.

18 Es müssen aber nicht alle Beobachtungseinheiten alle Alternativen gewählt haben.

Interpretation

Die folgenden Ausführungen bauen im Wesentlichen auf denselben Grundlagen auf wie die Interpretationsformen bei den Querschnittsmodellen, wie oben ab Seite 29 dargestellt. Um Wiederholungen zu vermeiden, werden hier nur die Dinge erläutert, die für die fixed-effects-Modelle besonders sind. Bei den fixed-effects-Modellen sind die möglichen Interpretationsformen stark eingeschränkt. Dies ist dadurch begründet, dass bei allen Interpretationstechniken, die direkt auf die Auswahlwahrscheinlichkeiten einer Alternativen j zielen, alle Komponenten der Linearkombination $\alpha_{ij} + X_{it}\beta'_j$ bekannt sein müssen. Das Problem liegt nun darin, dass die unbeobachtete Heterogenität α_i nicht gemessen und nicht geschätzt werden kann. Weiterhin baut die Schätzung der fixed-effects-Schätzer darauf auf, dass keine Verteilungsannahmen über α_i getroffen werden, die man zumindest für eine Schätzung der Linearkombination nutzen könnte.

So bleiben im Wesentlichen nur zwei gültige Interpretationsformen übrig: Erstens kann man zumindest bei zwei Alternativen aus dem Regressionskoeffizienten direkt auf die Richtung des Effektes schließen, d. h. beim felogit-Modell ist wie im Querschnitt eine Vorzeicheninterpretation der Regressionskoeffizienten auf die Richtung der Änderung der Auswahlwahrscheinlichkeiten möglich. Zweitens kann man auch beim allgemeinen femlogit-Modell die Odd Ratios interpretieren, wobei hier dieselben Probleme auftreten wie im Querschnitt. Erstens ist der Odds Ratio Effekt unanschaulich, zweitens führt eine leichtfertige Interpretation bei mehr als zwei Alternativen leicht in die Irre, und drittens ist er direkt von der willkürlich gewählten Varianz des Fehlerterms, wie in Annahme 10 dargestellt, abhängig, d. h. das Problem der „neglected heterogeneity" tritt auch hier auf. Es ist aber zu beachten, dass dieses Problem gegenüber den Querschnittsmodellen geringer ist, da mit der unbeobachteten Heterogenität α_i zumindest alle zeitlich invarianten Faktoren berücksichtigt werden. Diese beiden Interpretationsformen werden mehr oder weniger explizit bei Allison (2009, S. 33) und Wooldridge (2010, S. 622) empfohlen.

Neben diesen sicheren Interpretationsformen findet man bei Cameron und Trivedi (2009, S. 797) bzw. Cameron und Trivedi (2010, S. 626–630) einen Vorschlag für eine dritte Interpretationsstrategie. Sie schlagen vor, sich bei der Interpretation von der unbedingten Wahrscheinlichkeit, wie sie in Gleichung 3.14 dargestellt ist, zu lösen, und stattdessen Effekte auf die bedingte Wahrscheinlichkeit in Gleichung 3.16 zu interpretieren. D. h. man interpretiert den DE, bzw. ME einer unabhängigen Variablen X_m auf die Wahrscheinlichkeit, dass eine bestimmte Zeitreihe von Alternativenwahlen gewählt wird, bedingt auf die für jede Beobachtungseinheit i spezifische suffiziente Statistik k_i. D. h. konkret für den DE der unabhängigen Variablen X_m auf die realisierte Zeitreihe y_i von Alternativenwahlen:

$$
\begin{aligned}
(3.19) \quad DE_{X_m,y_i} &= \Pr(y_i|X_i+1,k_i,\alpha_i) - \Pr(y_i|X_i,k_i,\alpha_i) \\
&= \frac{\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it},j}(X_{it}+1)\widehat{\beta}'_j)}{\sum_{d_i\in\Delta_i}\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{d_{it},j}(X_{it}+1)\widehat{\beta}'_j)} \\
&\quad - \frac{\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it},j}X_{it}\widehat{\beta}'_j)}{\sum_{d_i\in\Delta_i}\exp(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{d_{it},j}X_{it}\widehat{\beta}'_j)}.
\end{aligned}
$$

Aus der inhaltlichen Perspektive betrachtet, interpretiert man *nicht* den Effekt einer unabhängigen Variablen auf die Auswahlwahrscheinlichkeit einer Alternative, sondern man interpretiert den Effekt, den eine unabhängige Variable auf die Realisation einer bestimmten Zeitreihe der abhängigen Variablen hat, bedingt auf die allgemeine Neigung zu den einzelnen Alternativen. Im praktischen Fall würde man eine bestimmte, für die Stichprobe „durchschnittliche” Zeitreihe auswählen. Dieser Zeitreihe entspricht dann eine bestimmte „allgemeine Neigung” für jede Alternative. Wenn man nun eine unabhängige Variable zu einem bestimmten Zeitpunkt ändert, kann man die Änderung der Realisationswahrscheinlichkeit der gewählten Zeitreihe berechnen.

Bei dieser Technik ist ebenso eine marginale Effektinterpretation möglich, und für beide Formen sind dann entsprechend die DEM-, ADE-, MEM- und AME-Interpretationen möglich (siehe S. 29ff.). Hierbei bleibt der Vorteil der gemittelten Effekte, dass diese das Problem der „neglected heterogeneity” stark abschwächen, bestehen. Die von Cameron und Trivedi vorgeschlagene Interpretationstechnik ist selbst für den einfachen Fall mit nur zwei Alternativen in der Literatur bisher nicht angewendet worden.

Diese Interpretationsstrategie ist aber letztlich für realistische Anwendungen nicht sinnvoll anwendbar, da bei realen Anwendungen mit mehr als drei Zeitpunkten der Gegenstand der Interpretation völlig unanschaulich ist. Wie gesagt, interpretiert man den Effekt einer unabhängigen Variablen auf die Realisation einer bestimmten Zeitreihe bedingt auf die respektive Neigung. Wenn man den Nenner als die Menge der Permutationen der gewählten Zeitreihe betrachtet, ist diese Menge bei realistischen Anwendungen in der Regel unanschaulich komplex. Wenn man den Nenner als Neigung für eine bestimmte Realisationen der abhängigen Variablen betrachtet, bedeutet das, dass man implizit die unbeobachtete Heterogenität mit der suffizienten Statistik ersetzt. Dadurch tritt aber das Problem auf, dass man über die „Schätzung” der unbeobachteten Heterogenität durch die suffizienten Statistik hinaus Annahmen über die Korrelationen zwischen der Neigung und den unabhängigen Variablen treffen muss. Für so eine Annahme hat man aber im Gegensatz zur suffizienten Statistik keinerlei Hinweise.

Einen vierten Interpretationsvorschlag findet man bei Schröder (2010). Schröder schlägt vor, für die Heterogenität einen bestimmten Wert so zu wählen, dass für prototypische Messzeitpunkte einer Beobachtungseinheit sinnvolle Realisationswahrscheinlichkeiten vorhergesagt werden. Davon ausgehend lässt sich dann prinzipiell das gesamte Arsenal der Querschnittsinterpretationen auch hier anwenden. Jedoch besteht auch hier wie beim Vorschlag von Cameron und Trivedi das Problem, dass man zusätzliche Werte für die Korrelationen zwischen der Heterogenität und den unabhängigen Variablen bestimmen muss, wofür keine hinreichenden Hinweise vorliegen. Ohne diese Korrelationen können konkret für die vorhergesagten Wahrscheinlichkeiten keine Konfidenzintervalle berechnet werden. Allgemeiner heißt das, dass man keine Hinweise dafür hat, wie bedeutsam ein Interpretationsergebnis ist.

Insgesamt überzeugt damit weder der Vorschlag von Cameron und Trivedi noch der Vorschlag von Schröder (2010), da beide Strategien in Ermangelung konkreter Werte für die Korrelation zwischen den unabhängigen Variablen und der Heterogenität willkürliche Ergebnisse liefern. So bleiben bei den fixed-effects-Modellen zwei Interpretationen

übrig: Die auf zwei Alternativen eingeschränkte Vorzeicheninterpretation und die Odds-Ratio- Interpretation mit den bekannten Problemen. Die Handlungsempfehlung fällt damit durch den Mangel an Alternativen eindeutig aus. Bei den fixed-effects-Modellen ist man neben der sehr einfachen Vorzeicheninterpretation einzig und allein auf die Odds-Ratio-Interpretation beschränkt.

3.2.3 Erweiterungen

Ausgehend von den bisherigen Darstellungen der Querschnitts- und fixed-effects-Modelle werden im Weiteren drei Erweiterungen der bisher dargestellten Standard-Modelle diskutiert: (1) Selektiver Ausfall der Messzeitpunkte, (2) panel-robuste Standardfehler, und (3) Alternativen-spezifische Kovariaten.

Selektive Attrition

Bei den Ausführungen zu den fixed-effects-Modellen haben wir bei Annahme 9 und Annahme 17 das Problem gestreift, dass sich die Anzahl der Messzeitpunkte über die Beobachtungseinheiten hinweg unterscheiden können. Solche Daten werden in der Literatur als „unbalanced-panel"-Daten bezeichnet. Im Gegensatz dazu stehen „balanced-panel"-Daten, bei denen für jede Beobachtungseinheit gleich viele Messzeitpunkte vorliegen.[19] Ursache für „unbalanced-panel"-Daten sind im Wesentlichen drei Dinge: Erstens können für bestimmte Beobachtungseinheiten für bestimmte Zeitpunkte keinen Messungen qua Design vorgenommen werden. Dies tritt häufig bei rotierenden Panels und bei Auffrischungen auf. Hier liegen für bestimmte zufällige Beobachtungseinheiten bestimmte Messzeitpunkte nicht vor. Zweitens können für einzelne Beobachtungseinheiten die Messreihen vorzeitig enden. Entscheidend ist, dass die Beendigung der Messung nicht in Händen des Forschers liegt, sondern die Beobachtungseinheiten sich selbst aus der Messung heraus selektieren. Dieser Fall wird typischerweise als *selektive Attrition* bezeichnet. Die dritte Ursache liegt in nicht zufälligem Ausfall einzelner Messzeitpunkte für einzelne Beobachtungseinheiten.

Zunächst ist klarzustellen, dass es sich hier nicht um ein Problem des selektiven Stichprobenausfalls im engeren Sinne handelt. In der in dieser Arbeit verwendeten Notation bedeutet selektiver Stichprobenausfall, dass die Annahme 17 dahingehend verletzt ist, dass die Stichprobe der *Beobachtungseinheiten* nicht einfach ist, d. h. die Ziehungswahrscheinlichkeiten der einzelnen Einheiten nicht gleich sind oder nicht unabhängig sind. Dieses Problem ist grundsätzlich durch adäquate Sample-Gewichte und Berücksichtigung etwaiger Klumpungen bei den Standardfehlern behebbar.[20] Davon unabhängig ist die „Ziehung" der Messzeitpunkte für jede Beobachtungseinheit.

Bevor wir die Ausführungen dazu weiter vertiefen, muss die Notation etwas verallgemeinert werden. In den Annahmen 9 und 17 wurde die Möglichkeit von unterschiedlich langen Messreihen für jede Beobachtungseinheit so eingeführt, dass die Anzahl der Messzeitpunkte T_i eine unabhängige Zufallsvariable ist. Um insbesondere den Fall von

19 Vgl. Greene (2000, S. 566f.), Lee (2002, S. 17), Cameron und Trivedi (2009, S. 739), Wooldridge (2010, S. 284f., 827–845).

20 Vgl. hierzu Lynn (2005); Levy und Lemeshow (2008).

zwischenzeitlichen Messausfällen vollständig zu berücksichtigen, muss die Notation folgendermaßen erweitert werden. Wir betrachten über die Menge aller Beobachtungseinheiten den ersten Messzeitpunkt $t = 1$ und letzten Messzeitpunkt $t = T$, für den es eine Beobachtung mit einer Messung gibt. D. h. diese Messzeitpunkte markieren somit den frühesten und spätesten Messzeitpunkt im gemeinsamen Kalender über alle Beobachtungseinheiten. Ausgehend vom Populationsmodell wird dann eine Stichprobe von N Beobachtungseinheiten gezogen. Für jede Beobachtungseinheit i liegt eine bestimmte Messreihe $(y_{it}, X_{it}, \alpha_i)$ vor. Ob für einen bestimmten Zeitpunkt t über den gemeinsamen Kalender für eine bestimmte Beobachtungseinheit i eine Messung vorliegt, wird durch einen Selektionsvektor $s_i = (s_{i1}, \ldots, s_{iT})'$ notiert, wobei $s_{it} = 1$ ist, wenn für den Zeitpunkt t eine Messung vorliegt. Formal notieren wir die Stichprobe dann so: $\{(y_i, X_i, \alpha_i, s_i)\}_{i=1,\ldots,N}$. In dieser Notation ist dann T_i die Anzahl aller Messungen für die Beobachtungseinheit i: $\sum_{t=1}^{T_i} s_{it} = T_i$. Man beachte, dass in dieser Notation klar nachvollziehbar ist, ob für eine bestimmtes Paar von Beobachtungseinheiten i und j bestimmte Messzeitpunkte vergleichbar sind, da die Messungen eindeutig in einem gemeinsamen Kalender vorliegen.

Für die Schätzung der Regressionskoeffizienten werden „unbalanced-panel”-Daten nur unter bestimmten Umständen problematisch. Die entscheidende Annahme, die diesen Umstand zum Problem werden lässt oder nicht, ist, dass der idiosynkratische Fehler von der Existenz einer Messung unabhängig ist (vgl. Cameron und Trivedi 2009, S. 739). Wenn diese Annahme gilt, gelten dieselben Aussagen über die Schätzer, wie sie in den obigen Darstellungen ausgeführt wurden. Wenn die Annahme verletzt ist, sind auch die Schätzer verzerrt. Hier gibt es im Wesentlichen zwei typische Ursachen: Erstens kann der Ausfall einer Messung von zwischenzeitlichen Ereignisse beeinflusst sein, z. B. der Ausfall nach schweren Krisenereignissen wie Scheidungen oder Todesfällen in der Familie. Wenn diese Ereignisse selbst Gegenstand der Untersuchung sind oder mit ihnen korreliert sind, sind die Effektschätzer verzerrt. Zweitens kann die wiederholte Messung systematisch das Messergebnis verändern, d. h. es liegt „panel conditioning” vor.

Der fixed-effects-Schätzer für das felogit-Modell und femlogit-Modell setzt, wie in den Annahmen 9 und 17 dargestellt, voraus, dass die Anzahl der Messzeitpunkte unabhängig ist von allen anderen Variablen (y_i, X_i, α_i). Dies erlaubt natürlich eine konsistente Schätzung, wenn ein „balanced-panel” vorliegt. Weiterhin gibt es auch kein Problem, wenn der Messzeitpunkteausfall zufällig entsteht wie bei rotierenden Panels. Nach Wooldridge (2010, S. 830) ist eine konsistente Schätzung mit dem in dieser Arbeit dargestellten fixed-effects-Schätzer möglich, wenn der Messzeitpunkteausfall s_{it} vollständig durch die beobachteten unabhängigen Variablen X_{it} und die zeitlich unveränderliche unbeobachtete Heterogenität α_i vorausgesagt werden kann.[21] Für darüber hinausgehende Fälle ist eine aufwändigere Behandlung der Messzeitpunkteausfälle notwendig, wie sie bei Cameron und Trivedi (2009, S. 801) und Wooldridge (2010, S. 837–845) knapp dargestellt sind. Diese Verfahren werden in den dieser Arbeit dargestellten Modellen nicht implementiert, und werden daher nicht weiter ausgeführt.

21 Vgl. die ähnliche Darstellung bei Cameron und Trivedi (2009, S. 739).

Panel-robuste Standardfehler

Bei den fixed-effects-Modellen wurde darauf verwiesen, dass bei Verletzung der Annahme 14 über die serielle Unabhängigkeit der Fehlerterme über die Messzeitpunkte „panel-robuste" Standardfehler zu schätzen sind. Hier soll kurz die Bedeutung dieser Schätzer allgemein und für die in dieser Arbeit betrachteten Modelle dargestellt werden. Zunächst einmal muss klargestellt werden, dass sich robuste Standardfehler immer nur auf asymptotische Schätzer und nicht auf sogenannte „small-sample"-Schätzer[22] beziehen. So werden entsprechend small-sample-Schätzer wie beim OLS-Modell in asymptotische Schätzer überführt, wenn man für diese robuste Standardfehler schätzen will. Der Einfachheit halber beschränkt sich die Darstellung auf die Schätzung der asymptotischen Standardfehler der vier in dieser Arbeit betrachteten Modelle logit, mlogit, felogit und femlogit.

Wir haben für die Querschnittsmodelle logit und mlogit die Gleichung 3.4c über die asymptotische Verteilung des Regressionskoeffizienten abgeleitet.

$$\widehat{(\alpha,\beta)}_{\mathrm{ML}} \xrightarrow{d} \mathcal{N}\Big((\alpha,\beta), \frac{-\mathrm{E}(\mathrm{H}_i(\alpha,\beta))^{-1}}{N}\Big)$$

Für die fixed-effects-Modelle felogit und femlogit wurde die entsprechende Gleichung 3.18c abgeleitet.

$$\widehat{\beta}_{\mathrm{ML}} \xrightarrow{d} \mathcal{N}\Big(\beta, \frac{-\mathrm{E}(\mathrm{H}_i(\beta))^{-1}}{N}\Big)$$

In den weiteren Ausführungen wird der Regressionkoeffizientenvektor für die Querschnittsmodelle (α,β) zusammenfassend als β dargestellt, um die Notation nicht zu verkomplizieren. Die Gleichungen über die asymptotischen Standardfehler setzen für alle vier Modelle voraus, dass die Dichten, wie sie in den Gleichungen 3.2 und 3.15 definiert sind, richtig spezifiziert sind (vgl. Cameron und Trivedi 2009, S. 142 und Wooldridge 2010, S. 469). Greene (2000, S. 823) stellt ausführlich dar, was die Ursache für eine fehlspezifizierte Dichte sein kann: "[...] any form of heteroscedasticity, unmeasured heterogeneity, omitted variables (even if they are orthogonal to the included ones), nonlinearity of the functional form of the index, or an error in the distributional assumption [...]." Wenn die Dichte-Funktionen fehlspezifiziert sind, ist darüber hinaus auch der Schätzer für die Regressionskoeffizienten $\widehat{\beta}_{\mathrm{ML}}$ inkonsistent.

Wenn eine Fehlspezifikation vorliegt, d. h. die angenommene Dichte $f^*_{y|X,\beta}$ nicht der „wahren" Dichte entspricht, kann man einen „quasi-maximum-likelihood"-Schätzer (QML) verwenden (vgl. hierzu Cameron und Trivedi 2009, S. 146f. und Wooldridge 2010, S. 502–504). Der QML-Schätzer maximiert die logarithmierte Likelihood-Funktion $\mathrm{L}^*(\beta)$, die aus der fehlspezifizierten Dichte folgt:

(3.20a) $\widehat{\beta}_{\mathrm{QML}} = \max_{\beta} \mathrm{L}^*(\beta)$.

22 Vgl. hierzu Greene (2000, S. 350–370).

Dieses Maximum ist ein konsistenter Schätzer für den falschen Regressionskoeffizienten $\beta^* \neq \beta$:

(3.20b) $\widehat{\beta}_{\text{QML}} \xrightarrow{p} \beta^* \neq \beta$.

Der QML-Schätzer ist wieder asymptotisch normalverteilt. Durch die fehlspezifizierte Dichte müssen die Standardfehler des Schätzer mit dem Huber-White-Sandwichschätzer geschätzt werden:

(3.20c) $$\widehat{\beta}_{\text{QML}} \xrightarrow{d} \mathcal{N}\left(\beta^*, \frac{-\text{E}\left(\text{H}_i(\beta^*)\right)^{-1}\text{E}\left(\text{s}_i(\beta^*)'\text{s}_i(\beta^*)\right) - \text{E}\left(\text{H}_i(\beta^*)\right)^{-1}}{N}\right).$$

Die Hesse-Matrix ist wieder die Matrix der zweiten partielle Ableitungen der Likelihoodfunktion nach dem Vektor der Regressionskoeffizienten, d. h. formal:

$$-\text{E}(\text{H}_i(\beta^*)) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 \ln \ell_i^*(\beta)}{\partial\beta'\partial\beta}\bigg|_{\widehat{\beta}_{\text{QML}}}.$$

Der Ausdruck $\text{s}_i(\beta^*)$ ist die „score"-Funktion, d. h. die Ableitung der Likelihoodfunktion nach dem Vektor der Regressionskoeffizienten, d. h. formal gilt:

$$\text{E}\left(\text{s}_i(\beta^*)'\text{s}_i(\beta^*)\right) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \ell_i^*(\beta)}{\partial\beta'}\frac{\partial \ell_i^*(\beta)}{\partial\beta}\bigg|_{\widehat{\beta}_{\text{QML}}}.$$

Es stellt sich die Frage, warum man den inkonsistenten QML-Schätzer verwenden sollte. Aus Sicht der robusten Standardfehler erhält man mit dem QML-Schätzer bei den logit-, mlogit-, felogit- und femlogit-Modellen korrekte Standardfehler für einen Schätzer, der in unbestimmter Weise vom wahren Wert β abweicht (vgl. hierzu Greene 2000, S. 823f., Cameron und Trivedi 2009, S. 467, 497, Wooldridge 2010, S. 569). Der QML-Schätzer ist zwar nicht als unverzerrter oder konsistenter Schätzer interpretierbar, aber er ist auch nicht völlig unbestimmt. Der QML-Schätzer $\widehat{\beta}_{\text{QML}}$ minimiert die Distanz zwischen der fehlspezifizierten Dichte $f^*_{y|X,\beta}$ und der wahren *unbedingten* Dichte f_y. Die Distanz ist hier durch das „Kullback-Leibler information criterion" (KLIC) definiert (vgl. Cameron und Trivedi 2009, S. 146f. und Wooldridge 2010, S. 502f.):

$$\text{KLIC} = \text{E}\left(\ln\left(\frac{f_y}{f^*_{y|X,\beta}}\right)\right).$$

Explizit bedeutet das, dass die KLIC-Distanz zwischen der fehlspezifizierten Dichte f^* und der wahren Dichte f für den QML-geschätzten Regressionskoeffizienten $\widehat{\beta}_{\text{QML}}$ minimal ist:

$$\min_{\beta} \text{E}\left(\ln\left(\frac{f_y}{f^*_{y|X,\beta}}\right)\right) = \text{E}\left(\ln\left(\frac{f_y}{f^*_{y|X,\widehat{\beta}_{\text{QML}}}}\right)\right).$$

Für die beiden Querschnittsmodelle logit und mlogit birgt die Schätzung robuster Standardfehler mit dem QML-Schätzer keine Vorteile. Die geschätzten Standardfehler sind nur

dann inkonsistent, wenn die Dichtefunktion fehlspezifiziert ist. Dann ist aber auch immer der Schätzer für die Regressionskoeffizienten inkonsistent.

Bei Modellen mit Paneldaten, die mit dem maximum-likelihood-Verfahren geschätzt werden, sind grundsätzlich Fehlspezifikationen der Dichtefunktion denkbar, die mit robusten Schätzern eingefangen werden können. Beim maximum-likelihood-Verfahren ist der Schätzer von der korrekt spezifizierten Dichte abhängig. Häufig wird bei Modellen über Paneldaten serielle Unabhängigkeit der Fehlerterme über die Messzeitpunkte angenommen. Auch beim felogit- und femlogit-Modell wird diese Annahme mit Annahme 14 getroffen. Diese Annahme ist häufig verletzt und der geschätzte Regressionskoeffizient $\widehat{\beta}_{\text{MKL}}$ ist dann entsprechend inkonsistent. Bei manchen Modellen mit Paneldaten schafft eine „partial-maximum-likelihood"(PML)-Schätzung Abhilfe (vgl. hierzu Cameron und Trivedi 2009, S. 150 und Wooldridge 2010, S. 486–492). Wenn die bedingte Dichtefunktion über die Zeitreihe der abhängigen Variablen bedingt auf die unabhängigen Variablen für eine Beobachtungseinheit $f_{y_i|X_i}$ in das Produkt der entsprechenden bedingten Dichten für jeden einzelnen Messzeitpunkt zerlegt werden kann, d. h.

$$f_{y_i|X_i} = \prod_{t=1}^{T_i} f_{y_{it}|X_{it}},$$

kann man das zugehörige Modell mit dem PML-Verfahren schätzen, und so auf die Annahme der seriellen Unabhängigkeit der Fehlerterme verzichten.

Bei den in dieser Arbeit betrachteten Modelle felogit und femlogit ist dies aber nicht möglich, da die Dichtefunktionen für jeden einzelnen Messzeitpunkt die unbeobachtete Heterogenität enthält, und umgekehrt die Dichtefunktion ohne die unbeobachtete Heterogenität für einen einzelnen Messzeitpunkt keine Entsprechung findet.[23] D. h. die Schätzung panel-robuster Standardfehler ist mit dem Huber-White-Schätzer bei den felogit- und femlogit-Modellen nicht möglich. Die einzige Möglichkeit zur Schätzung panel-robuster Standardfehler ist das „panel-bootstrap"-Verfahren (vgl. Cameron und Trivedi 2009, S. 377f., 788f.).[24]

Das heisst aber nicht, dass man bei den felogit- und femlogit-Modellen keine robusten Standardfehler schätzen kann. Eine Implementation des Huber-White-Schätzers der Fehlervarianz ergibt die Fehlervarianz für den QML-Schätzer, wie er oben beschrieben wurde. Wie im Kapitel 4 dargestellt wird, wird bei der Implementation des felogit- und femlogit-Modells auf die Umsetzung des Huber-White-Schätzers der Fehlervarianz verzichtet, da damit nicht das Problem der Panelkorrelation gelöst werden kann.

---

23 Dies rührt im Wesentlichen daher, dass die Annahme der seriellen Unabhängigkeit der Fehlerterme notwendig ist, damit $\sum_{T_i} y_{it}$ eine suffiziente Statistik für α_i ist. Ohne diese Annahme ist der Ansatz der Dichtefunktion bedingt für die suffiziente Statistik fruchtlos, da hierfür eine Dichtefunktion über die gesamte Zeitreihe y_i benötigt wird.

24 Es ist anzumerken, dass z. B. im Statistikpaket Stata für bestehende Umsetzungen z. B. der binären logistischen Regression mit fixed-effects mit `clogit` und `xtlogit` eine Schätzung robuster Standardfehler angeboten wird. Hier handelt es sich aber um *heteroskedastie-robuste* Standardfehler. Bei diesen wird die Annahme über die Korrelation der Messfehler über die *Beobachtungseinheiten* – nicht über die Messzeitpunkte – abgeschwächt (vgl. hierzu z. B. Cameron und Trivedi 2010, S. 334).

Zusammenfassend ist die Schätzung robuster Standardfehler bei den Modellen, die in dieser Arbeit im Mittelpunkt stehen, nur bei den fixed-effects-Modellen sinnvoll anwendbar. Bei den Querschnittsmodellen sind entweder sowohl der Schätzer für den Regressionskoeffizienten als auch die zugehörigen Standardfehler inkonsistent oder beide sind konsistent. Bei den fixed-effects-Modellen sollten robuste Standardfehler geschätzt werden, wenn die Annahme der seriellen Korrelation über die Messzeitpunkte verletzt sind. In diesem Fall kann man aber nur mit dem panel-bootstrap-Verfahren panel-robuste Standardfehler schätzen.

Alternativen-spezifische Kovariate

Bei den bisherigen Ausführungen zu den Querschnitts- und fixed-effects-Modellen haben wir implizit angenommen, dass sich die unabhängigen Variablen über die Alternativen j hinweg *nicht* unterscheiden, d. h. für alle Alternativen j ist $X_j = X$. Umgekehrt wird für die Regressionskoeffizienten β_j angenommen, dass diese über die Alternativen hinweg variieren können. Zur Identifikation muss eine Alternative B als Referenzkategorie bestimmt werden, für die $\beta_B = 0$ gesetzt wird. In diesem Abschnitt sollen mögliche Alternativen dieser Operationalisierung dargestellt werden (vgl. Cameron und Trivedi 2009, S. 491–495, 500).[25] Die Operationalisierungsalternativen sind grundsätzlich sowohl für die Querschnitts- als auch die fixed-effects-Modelle anwendbar. Der einzige Unterschied besteht darin, dass die unabhängigen Variablen bei den fixed-effects-Modellen immer über die Messzeitpunkte variieren müssen. Um die Notation einfach zu halten, wird der Subskript t für die Messzeitpunkte weggelassen.

Den Ausgangspunkt der Operationalisierungsalternativen bildet das lineare Modell über die latente Variable y_j^*, die für die Querschnittsmodelle mit Annahme 1 und für die fixed-effects-Modelle mit Annahme 9 eingeführt spezifiziert werden:

$$y_j^* = \alpha_j + X\beta_j' + \epsilon_j.$$

Hier sind die unabhängigen Variablen über die Alternativen hinweg konstant und die Regressionskoeffizienten können sich unterscheiden. Solche Regressoren werden häufig als „alternativen-invariante" Variablen bezeichnet. Demgegenüber stehen Variablen, die über die Beobachtungseinheiten *und* über die Alternativen hinweg variieren. Diese werden häufig als „alternativen-spezifische" Variablen bezeichnet.[26]

Alternativen-spezifische Variablen können ohne große Modifikationen in Querschnitts- und fixed-effects-Modelle aufgenommen werden. Zur besseren Anschaulich-

25 Vgl. hierzu auch Long (1997, S. 178–182), Greene (2000, S. 857–865), Train (2009, S. 20–22) und Wooldridge (2010, S. 643–649). Schon bei McFadden (1974) findet man die hier dargestellten Operationalisierungen. Ausführlichere Darstellungen finden sich bei Maier und Weiss (1990) und Ben-Akiva und Lerman (1994). Eine klare Darstellung der technischen Umsetzung findet man bei Kühnel (1993, S. 143–165).

26 Neben dem Begriff der alternativen-invariante Variablen wie bei Cameron und Trivedi (2009) findet man auch die Begriffe sozioökonomische Regressoren (Maier und Weiss 1990; Ben-Akiva und Lerman 1994), individuelle Charakteristika oder Attribute (McFadden 1974; Greene 2000; Wooldridge 2010) oder soziodemographische Variablen (Train 2009). Alternativen-spezifische Variablen (wie bei Train 2009) werden auch als „alternative-varying regressors" (Cameron und Trivedi 2009), „choice"- oder Alternativenattribute (McFadden 1974; Greene 2000; Wooldridge 2010), generische Variablen (Ben-Akiva und Lerman 1994) oder Alternativencharakteristika (Maier und Weiss 1990) bezeichnet.

keit seien alternativen-invariante Variablen mit S und alternativen-spezifische Variablen mit C notiert.[27] Dann gilt für die latente Variable für die Alternativen j:

$$y_j^* = \alpha_j + S\beta_j' + C_j\gamma' + \epsilon_j.$$

Zur Identifikation der Regressionskoeffizienten für die alternativen-invarianten Variablen müssen wir den entsprechenden Koeffizienten β_B für die Referenzkategorie null setzen. Analog muss für die Identifikation der Regressionskoeffizienten für die alternativen-spezifischen Variablen die Ausprägung für Referenzkategorie $C_B = 0$ gesetzt werden. Für die Linearkombination $X\beta'$ für die Auswahlwahrscheinlichkeit ergibt sich dann dieser Ausdruck:

$$\Pr(y = j) = \frac{\exp(\alpha_j + S(\beta_j - \beta_B)' + (C_j - C_B)\gamma')}{1 + \sum_{k \neq B} \exp(\alpha_k + S(\beta_k - \beta_B)' + (C_j - C_B)\gamma')} \quad \text{für } j \neq B,$$

$$\Pr(y = B) = \frac{1}{1 + \sum_{k \neq B} \exp(\alpha_k + S(\beta_k - \beta_B)' + (C_j - C_B)\gamma')}.$$

Dadurch, dass $\beta_B = 0$ und $C_B = 0$ gesetzt wurden, vereinfacht sich in diesem Ausdruck die Linearkombination zu $\alpha_j + S\beta_j' + C_j\gamma'$. Bei den alternativen-spezifischen Variablen ist zu beachten, dass man den Effekt der Differenzen dieser Variablen auf die Auswahlwahrscheinlichkeit schätzt.

In der Literatur wird ein Modell, dass nur alternativen-spezifische Variablen enthält, nach McFadden (1974) als „conditional logit"-Modell bezeichnet. Modelle, die nur alternativen-invariante Variablen enthalten, werden als „multinomiale logit"-Modelle bezeichnet. Modelle, bei denen beiden Variablenformen verwendet werden, werden als „mixed logit"-Modelle bezeichnet.[28]

Bei der technischen Umsetzung, d. h. der konkreten Schätzung in einer Statistiksoftware, steht in der Regel eine Implementation des „conditional logit"-Modells oder des multinomialen Logitmodells oder beide zur Verfügung.[29] Es ist zu beachten, dass sich die Implementationen des „conditional logit"-Modells und des multinomialen logit-Modells grundsätzlich nicht unterscheiden. Der Unterschied liegt in der Regel in der Struktur, in der die unabhängigen und abhängigen Variablen vorliegen.

Da beim multinomialen logit-Modell nur alternativen-invariante Variablen verwendet werden, die sich über die Alternativen nicht unterscheiden, liegen hier die Daten im sogenannten „wide"-Format vor. Wie in Tabelle 3.1 dargestellt, entspricht ein „Fall", bzw. eine Zeile der Datenmatrix, einer Beobachtungseinheit. Bei fixed-effects-Modellen entspricht eine Zeile einem Messzeitpunkt für eine bestimmte Beobachtungseinheit. Beim „conditional logit"-Modell ist die Datenmatrix im „long"-Format. Wie in Tabelle 3.2 dargestellt, ist

27 Die Kürzel S und C folgen den Begriffen von Maier und Weiss (1990).

28 Der Begriff der multinomial-logit-Modelle suggeriert hier, dass die drei Operationalisierungsformen nur bei mehreren Alternativen (J>2) auftreten. Grundsätzlich ist die Unterscheidung auch bei nur zwei Alternativen möglich, wobei sie bei mehreren Alternativen mehr Gehalt bekommt. Der Begriff der mixed-logit-Modelle ist irreführend, da er auch für random-effects-Modelle bei discrete-choice-Modellen verwendet wird. Vgl. hierzu Train (2009, Kap. 6).

29 In Stata liegt mit dem Befehl `asclogit` eine Implementation des mixed-logit-Modells vor (StataCorp LP 2009d, S. 86–96).

Tabelle 3.1: Datenmatrix im wide-Format

Beobachtungs-einheit	Gewählte Alternative	Alternativen-invariante Variable S
1	2	10
2	1	12
3	2	9

eine Zeile der Datenmatrix eine mögliche Alternative für eine bestimmte Beobachtungseinheit.

Tabelle 3.2: Datenmatrix im long-Format

Beobachtungs-einheit	Alternativen	Gewählte Alternative	Alternativen-spezifische Variable C
1	1	0	5
1	2	1	10
1	3	0	7
2	1	1	4
2	2	0	3
2	3	0	10
3	1	0	1
3	2	1	5
3	3	0	7

Wie Kühnel (1993, S. 143–165) ausführlich zeigt, kann man die beiden Formen wechselseitig ineinander überführen. D. h. man kann den Einfluss alternativen-spezifischer Variablen mit einer multinomialen logit-Implementation schätzen, und umgekehrt den Einfluss alternativen-invarianter Variablen mit einer conditional-logit-Implementation. Für die Schätzung des Einflusses der alternativen-spezifischen Variablen C in einem multinomialen logit-Modell führt man für jeden Kontrast zwischen einer Alternativen $j \neq B$ und der Referenzkategorie B eine Variable $C_j - C_B$ ein. Bei der Schätzung werden für die zugehörigen Regressionskoeffizienten γ zusätzliche lineare Beschränkungen eingeführt:

$$\gamma_1 = \gamma_2 = \cdots = \gamma_J,$$
$$\gamma_B = 0.$$

Dadurch erhält man für den Effekt der alternativen-spezifischen Variable, wie bei der Implementation im conditional-logit-Modell, genau einen Effekt. Diese Operationalisierung spiegelt sich in der Gleichung für die Auswahlwahrscheinlichkeit folgendermaßen wieder:

$$\Pr(y = j) = \frac{\exp(\alpha_j + (C_1 - C_B)\gamma' + (C_2 - C_B)\gamma' + \cdots + (C_J - C_B)\gamma')}{1 + \sum_{k \neq B} \exp(\alpha_k + \sum_{l=1}^{J} (C_l - C_B)\gamma')}.$$

Durch diese Operationalisierung wird aus der alternativen-spezifischen Variablen C die $J-1$ alternativen-invarianten Variablen $C_1 - C_B$, ..., $C_J - C_B$ gebildet. In Tabelle 3.3 ist dargestellt, wie sich die Operationalisierung der alternativen-spezifischen Variablen aus der Datenmatrix aus Tabelle 3.2 in der Datenmatrix im wide-Format niederschlägt, wobei die Alternative 1 als Referenzkategorie gewählt ist.

Tabelle 3.3: Alternativen-spezifische Variablen im wide-Format

Beobachtungs-einheit	Gewählte Alternative	Alternativen-spezifische Variablen $C_2 - C_1$	$C_3 - C_1$
1	2	$10 - 5 = 5$	$7 - 5 = 2$
2	1	$3 - 4 = -1$	$10 - 4 = 6$
3	2	$5 - 1 = 4$	$7 - 1 = 6$

Umgekehrt lassen sich auch alternativen-invariante Variablen so operationalisieren, dass man deren Einfluss mit conditional-logit-Implementierungen schätzen kann. Grundsätzlich können bei conditional-logit-Modellen, wie man sie in der Literatur findet, immer Variablen eingefügt werden, die sich nicht über die Alternativen hinweg unterscheiden. Der entscheidende Unterschied zu den multinomialen logit-Modellen besteht darin, dass bei ersteren *ein* Regressionskoeffizient für alle Alternativen geschätzt wird, und bei letzteren für *jede* Alternative ein Koeffizient geschätzt wird. Mit anderen Worten operationalisiert man so eine Interaktion mit alternativen-spezifischen Konstanten K_1, ..., K_J und der konstanten alternativen-invarianten Variablen S. In Tabelle 3.4 ist dargestellt, wie sich diese Operationalisierung der alternativen-invarianten Variable in Tabelle 3.1 in den Daten widerspiegelt.

Tabelle 3.4: Datenmatrix im long-Format

Beobachtungseinheit	Alternativen	Gewählte Alternative	Operationalisierte alternativen-invariante Variable						
			K_1	K_2	K_3	S	K_1S	K_2S	K_3S
1	1	0	1	0	0	10	10	0	0
1	2	1	0	1	0	10	0	10	0
1	3	0	0	0	1	10	0	0	10
2	1	1	1	0	0	12	12	0	0
2	2	0	0	1	0	12	0	12	0
2	3	0	0	0	1	12	0	0	12
3	1	0	1	0	0	9	9	0	0
3	2	1	0	1	0	9	0	9	0
3	3	0	0	0	1	9	0	0	9

Man schätzt dann für die Variablen K_1S, K_2S, K_3S, die nun alternativen-spezifische Variablen sind, jeweils einen eigenen Effekt γ_{K_1S}, γ_{K_2S} und γ_{K_3S}.

Zusammenfassend kann man sowohl Variablen, die über die Alternativen hinweg variieren, als auch solche, die über die Alternativen hinweg konstant sind, in die hier betrachteten Modelle integrieren. Es ist zu beachten, dass bei den fixed-effects-Modellen die Variablen über die Messzeitpunkte hinweg variieren müssen. Das femlogit-Modell, dass in dieser Arbeit im Mittelpunkt steht, wird, wie im nächsten Kapitel dargestellt wird, im wide-Format implementiert. D. h. dass für die Berücksichtigung alternativen-spezifischer Variablen lineare Beschränkungen eingeführt werden müssen.

3.3 Diskussion

In diesem Kapitel wurden die statistischen Grundlagen der Querschnitts- und fixed-effects-Modelle dargestellt. Bei diesen Ausführungen wurde nur am Rande angesprochen, unter welchen Umständen eine Modellgruppe der anderen Gruppe vorzuziehen ist. Im Weiteren sollen verschiedene Vor- und Nachteile der Querschnitts- und fixed-effects-Modelle diskutiert werden.

3.3.1 Varianz über die Zeit auf den unabhängigen Variablen

Erstens wird an den fixed-effects-Modellen im Allgemeinen kritisiert, dass sie gegenüber den Querschnitts-, pooled- und random-effects-Modellen ineffiziente Schätzer liefern. Aus der Annahme 16 folgt, dass für Variablen, die über die Messzeitpunkte hinweg konstant sind, keine Effekte geschätzt werden. Da die Beobachtungseinheiten ihre eigenen Kontrollgruppen darstellen, wird nur die sogenannte „within”-Information genutzt, und die „between”-Information vernachlässigt. Halaby (2004, S. 522) führt aus, dass dieser Aspekt der fixed-effects-Modelle häufig abgelehnt wird: “[F]ixed effects [estimation] can be interpreted substantively as ‘throwing away’ all between-[unit] variation present in the data. [...] [R]andom effects [estimation] is asymptotically efficient relative to fixed effects [estimation].” Durch die Opferung der Varianz zwischen den Beobachtungseinheiten und die Beschränkung auf die Varianz innerhalb der Beobachtungseinheiten erhält man konsistente Schätzer für die Regressionskoeffizienten, wobei für den Zusammenhang zwischen unbeobachteten zeitkonstanten Variablen und den beobachteten zeitlich variierenden Variablen keine Annahmen getroffen werden müssen:

> “[...] the luxury of ‘throwing out’ between variation is the very source of the advantage that panel data provide over cross-sectional data and that within-group estimators, such as fixed effects and first difference, exploit to avoid heterogeneity bias. Throwing out between variation is not wasting data: It buys protection against biased and inconsistent parameter estimates.” (Halaby 2004, S. 523)

Den Verweis auf die relative Ineffizienz der fixed-effects-Schätzer findet man in vielen Quellen.[30] Brüderl (2010, S. 971) stellt die Beschränkung auf die „within”-Information in einen größeren Zusammenhang. Dadurch, dass man Beobachtungseinheiten mit Varianz auf den unabhängigen Variablen über die Messzeit berücksichtigt, schätzt man letztlich einen „average treatment effect of the treated” (ATET). Brüderl führt weiter aus:

30 Vgl. z. B. Allison (1994, S. 191), Maddala (2001, S. 576), Hsiao (2003, S. 41f., 48f.), Allison (2009, S. 3), Cameron und Trivedi (2009, S. 702f.), Wooldridge (2010, S. 301, 304).

> „[Es] trägt oft nur ein kleiner Teil der Panel-Stichprobe zum FESchätzer bei. Das ist in Ordnung so, weil nur der Teil der Daten verwendet wird, der tatsächlich Within-Information liefert. Aber man muss sich bewusst sein, dass dies ein nur mit Vorsicht generalisierbarer ATET-Schätzer ist."

3.3.2 Varianz über die Zeit auf der abhängigen Variable

Das zweite Problem der in dieser Arbeit diskutierten Modelle ist der Beschränkung auf die „within"-Information ähnlich. Grundsätzlich berücksichtigen alle fixed-effects-Modelle nur die Varianz über die Messzeitpunkte innerhalb der Beobachtungseinheiten. Bei den nicht linearen fixed-effects-Modellen kommt darüber hinaus eine weitere Beschränkung hinzu. Wie auf Seite 46 bereits kurz dargestellt wurde, werden nur solche Beobachtungseinheiten berücksichtigt, bei denen neben der Varianz auf den unabhängigen Variablen auch Varianz auf der abhängigen Variablen vorliegt. Dieser Umstand wird an vielen Stellen in der Literatur ohne weitere Anmerkungen benannt.[31] Brüderl (2010, S. 987) führt hierzu aus, dass diese Beschränkung häufig irrigerweise als eine selektive Stichprobenauswahl problematisiert wird. Wooldridge stellt fest, dass diese Beschränkung unproblematisch ist:

> "One should not expend energy worrying about the loss of observations for which $y_{i1} = y_{i2}$. Because $f_{\alpha_i|X_i}$ is unrestricted, α_i is free to vary as much as required to make $\Pr(y_{it} = 1|X_{it}, \alpha_i)$ arbitrarily close to one (if $y_{it} = 1$, $t = 1, 2$) or arbitrarily close to zero (if $y_{it} = 0$, $t = 1, 2$). When y_{it} does not change across t, α_i can adjust to make both observed outcomes occur with certainty. Clearly, such data points contain no information for estimating β, and so they *should* drop out of estimation." (Wooldridge 2010, S. 621f., Hevorhebung im Original, Notation angepasst)

Mit anderen Worten sind für die Beobachtungseinheiten, in denen keine Varianz auf der abhängigen Variablen vorliegt, die entsprechenden unbeobachteten Heterogenitäten α_{ij} so extrem ausgeprägt, dass die unabhängigen Variablen keine Änderung mehr hervorbringen können. Für eine Beobachtungseinheit, die eine bestimmte Alternative j immer wählt, ist die entsprechende Heterogenität unendlich, d. h. $\alpha_{ij} = \infty$. Wenn die Alternative j nie gewählt wird, ist die entsprechende Heterogenität minus unendlich, d. h. $\alpha_{ij} = -\infty$. Um Identifizierbarkeit zu gewährleisten, werden Fälle, bei denen eine Alternative für alle Messzeitpunkte gewählt wird, für die Schätzung nicht berücksichtigt.

3.3.3 Annahmen-Dilemma zwischen den Querschnitts- und den fixed-effects-Modellen

Zur Identifikation müssen bei den fixed-effects-Modellen gegenüber den Querschnittsmodellen die Annahme 12 über strikte Exogenität und die Annahme 14 über serielle Unabhängigkeit getroffen werden. Um diese Einschränkung richtig zu verstehen, muss man den richtigen Vergleich zwischen den fixed-effects-Modellen und den Querschnittsmodellen wählen. Bei den Querschnittsmodellen ist α ein konstanter, zu schätzender Parameter für die gesamte Population, d. h. damit ist der Parameter grundsätz-

31 Vgl. z. B. Hsiao (1986, S. 162), Baltagi (1995, S. 179f.), Baltagi (2001, S. 207), Hsiao (2003, S. 196), Baltagi (2005, S. 210f.), Allison (2009, S. 36), Cameron und Trivedi (2009, S. 796).

lich unkorreliert mit den unabhängigen Variablen X. Bei den fixed-effects-Modellen ist α eine Zufallsvariable in der Population, die beliebig mit den unabhängigen Variablen korrelieren kann. Darüber hinaus wird bei den fixed-effects-Modellen strikte Exogenität und serielle Unabhängigkeit der Fehlerterme angenommen. Bei den Querschnittsmodellen findet die Annahme der seriellen Unabhängigkeit keine Entsprechung, da es für jede unabhängige Beobachtungseinheit nur eine Beobachtung gibt. Daneben wird bei den Querschnittsmodellen nur eine kontemporäre Exogenität für die Fehlerterme angenommen.[32] Die übrigen Annahmen sind für die fixed-effects-Modelle und die Querschnittsmodelle im Wesentlichen gleich, abgesehen davon, dass bei den fixed-effects-Modellen lineare Unabhängigkeit über die Differenzen über die Messzeit angenommen wird. Damit werden für die Querschnittsmodelle die gleichen Annahmen getroffen wie für die sogenannten „pooled"-Modelle,[33] wobei bei den pooled-Modellen gegenüber den Querschnitts-Modellen teilweise Annahmen über die Beziehungen der Fehlerterme über die Zeitpunkte hinweg getroffen werden.[34]

Aus dem Vergleich zwischen den pooled- und fixed-effects-Modellen ergibt sich nun, dass streng genommen keines der beiden Modelle allgemeiner ist als das andere. Bei den fixed-effects-Modellen ist die unbeobachtete Heterogenität α eine Zufallsvariable, für die nur angenommen wird, dass sie unabhängig vom idiosynkratischen Fehlerterm ϵ ist. Bei den pooled-Modellen muss mindestens angenommen werden, dass die Heterogenität α von den unabhängigen Variablen unabhängig ist und dass sie einer festgelegten Verteilung folgt (vgl. Wooldridge 2010, S. 613). D. h. in dieser Hinsicht sind die fixed-effects-Modelle allgemeiner als die pooled- und damit die Querschnittsmodelle. Umgekehrt muss bei den fixed-effects-Modellen aber für die Fehlerterme strikte Exogenität und serielle Unabhängigkeit angenommen werden. Bei den pooled-Modellen muss hingegen mit der kontemporären Exogenität keine Annahme über die Zusammenhänge über die Messzeitpunkte hinweg getroffen werden. D. h. hier sind die pooled-Modelle allgemeiner als die fixed-effects-Modelle, da bei letzteren z. B. der Einfluss von zeitlich verzögerten abhängigen Variablen ausgeschlossen ist (vgl. hierzu Wooldridge 2010, S. 610f.).

Leider gibt es keine eindeutige Entscheidungsregel zur Lösung dieses Dilemmas. Aufgrund der Verschränktheit ist ein Hausman-Ansatz nicht möglich, wie es z. B. bei der Entscheidung über die Gültigkeit über die Annahmen der random-effects- und der fixed-effects-Annahmen möglich ist (vgl. z. B. Cameron und Trivedi 2009, S. 788). Hierfür müsste eine Modellgruppe wirklich mit weniger Annahmen auskommen als die andere, und die Modellgruppe mit den stärkeren Annahmen müsste präzisere, d. h. effizientere Schätzer liefern, sofern die spezielleren Annahmen gültig sind. Zur Entscheidung über

32 Siehe hierzu Wooldridge (2010, S. 163f.). Da für eine Beobachtungseinheit nur ein Messzeitpunkt vorliegt, werden nur Annahmen über die Beziehung zwischen dem Fehlerterm und den unabhängigen Variablen zum selben Messzeitpunkt getroffen. Beziehungen zwischen unabhängigen Variablen zu einem Zeitpunkt und dem Fehlerterm zu einem zumindest theoretisch denkbar anderen Zeitpunkt sind unbeschränkt.

33 Vgl. die Darstellung der linearen pooled-Modelle bei (Cameron und Trivedi 2009, S. 699, 702) und (Wooldridge 2010, S. 191–194) und die Darstellung der nicht linearen pooled-Modelle bei Cameron und Trivedi (2009, S. 787) und Wooldridge (2010, S. 609–610).

34 Wenn auf diese Annahme bei den pooled-Modellen verzichtet wird, müssen diese mit panel-robusten Standardfehlern geschätzt werden.

die Gültigkeit der Annahmen der Querschnitts- und der fixed-effects-Modelle bleiben nur Plausibilitätsüberlegungen. Die Verletzung der Unabhängigkeit der unbeobachteten Heterogenität von den unabhängigen Variablen lässt sich häufig leicht konstruieren: Z. B. können sich Gruppen mit bestimmten Heterogenitäten selbst in das Treatment selektieren. Weiterhin können für bestimmte Gruppen bestimmte Messfehler auf den unabhängigen Variablen vorliegen. Umgekehrt ist die Verletzung der strikten Exogenität gegenüber der kontemporären Exogenität selten offensichtlich. Man findet zwar häufig Verweise auf Antizipations- und Verharrungseffekte, die man im Modell berücksichtigen muss. Es stellt sich aber die Frage, ob diese Effekte nicht durch die Kontrolle von zeitverzögerten und zeitlich nachgelagerten *unabhängigen* Variablen adäquat berücksichtigt werden können.

3.3.4 Unterschiede in der Interpretation

Das vierte Problem bei der Anwendung der fixed-effects-Modelle, die in dieser Arbeit dargestellt werden, ist die erschwerte Interpretierbarkeit. Wie bereits auf Seite 47 dargestellt, können die fixed-effects-Modelle gegenüber den Querschnitts-Modellen nur eingeschränkt interpretiert werden. Das Problem der Interpretierbarkeit zieht sich im Wesentlichen darauf zurück, dass man grundsätzlich eine Interpretation des Effektes auf die latente Variable y^* vermeiden möchte und die anschaulichere Interpretation des Effektes auf die realisierte Alternative y bevorzugt wird. Wie ausführlich in den Abschnitten bei S. 29ff. und S. 47ff. dargestellt wurde, benötigt man für die Bezugnahme auf die realisierte Alternative y die Ausprägungen der *gesamten* Linearkombination $\widehat{\alpha_j} + X\widehat{\beta}'_j$. Da bei den nicht linearen fixed-effects-Modellen durch die Konditionierung auf die suffiziente Statistik für α_{ij} die unbeobachtete Heterogenität nicht geschätzt werden kann, und auch sonst keine Verteilungsannahmen getroffen werden, ist die Linearkombination $\widehat{\alpha_j} + X\widehat{\beta}'_j$ nur unvollständig bekannt. Das hat zur Konsequenz, dass man bei den fixed-effects-Modellen entweder nur Effekte auf die latente Neigung zu einer Alternativen y_j^* interpretieren kann, oder abgeschwächt Effekte auf die Odds-Ratios interpretiert.

3.3.5 „Neglected heterogeneity"

Die vier bisher dargestellten Aspekte der in dieser Arbeit dargestellten Modelle sind mehr oder weniger schwerwiegende Probleme der fixed-effects-Modelle. Der fünfte Aspekt der Modelle ist dagegen grundsätzlich ein Vorteil der fixed-effects-Modelle gegenüber den Querschnittsmodellen. Wie bei den Annahmen 2 und 10 dargestellt wurde, haben beide Modellfamilien das Problem der „neglected heterogeneity". Wooldridge (2010, S. 582ff.) bezeichnet mit diesem Begriff unbeobachtete und anderweitig nicht berücksichtigte Regressoren, die zwar einen Einfluss auf die abhängige Variable haben, aber von den berücksichtigten unabhängigen Variablen *unabhängig* ist. Bei linearen Modellen ist diese Form der unvollständigen Spezifikation der Regressoren unproblematisch. Eine auf diese Weise vernachlässigte Variable z_i wird dem Fehlerterm zugeschlagen, der von den berücksichtigten unabhängigen Variablen ebenso unabhängig ist:

$$\begin{aligned} y_i &= \alpha + X_i\beta' + z_i + \epsilon_i \\ &= \alpha + X_i\beta' + (z_i + \epsilon_i). \end{aligned}$$

Bei nicht linearen Modellen, die mit dem maximum-likelihood-Verfahren geschätzt werden, muss die bedingte Dichte $f_{y|X} = f_\epsilon$ spezifiziert werden. In unserem Beispiel bedeutet das, dass durch die Vernachlässigung der Variablen z der zusammengesetzte Fehler $(z_i + \epsilon_i)$ fehlspezifiziert ist. Dies führt dazu, dass die geschätzten Regressionskoeffizienten β von der Korrelation der vernachlässigten Variable z mit der abhängigen Variable und ihrer Varianz abhängig ist (vgl. hierzu Allison 1999, Mood 2010, Wooldridge 2010, S. 582–585).

Der Vorteil der fixed-effects-Modelle gegenüber den Querschnittsmodellen ergibt sich nun daraus, dass bei den fixed-effects-Modellen für alle Variablen, die über die Messzeitpunkte konstant sind, implizit kontrolliert wird. D. h. bei den fixed-effects-Modellen schrumpft das Problem der „neglected heterogeneity" auf solche vernachlässigte Regressoren zusammen, die über die Messzeitpunkte hinweg variieren. Dieser Vorzug wird jedoch auch teilweise durch den folgenden Nachteil der fixed-effects-Modelle relativiert: Das üblicherweise empfohlene Verfahren zur Abmilderung des Problems der „neglected heterogeneity" ist die Betrachtung von gemittelten Effekten, d. h. „average partial effects" (vgl. die Darstellungen bei S. 29ff. und S. 47ff.). Gerade diese Interpretationsform ist bei den fixed-effects-Modellen nicht direkt anwendbar, da die unbeobachtete Heterogenität nicht geschätzt wird. Wenn man aber, wie im Abschnitt zur Interpretation der fixed-effects-Modelle dargestellt, die suffiziente Statistik als Schätzer für die unbeobachtete Heterogenität α_i verwendet, kann man grundsätzlich auch APE- und AME-Interpretation auch bei fixed-effects-Modellen vornehmen und so den Effekt von etwaig vernachlässigten zeitlich variierenden Variablen auf die geschätzten Regressionsparameter herausrechnen.

3.3.6 Fazit

Die in dieser Arbeit im Mittelpunkt stehenden fixed-effects-Modelle weisen also gegenüber den Querschnittsmodellen verschiedene Probleme auf. Erstens ist man bei den Analysen auf Beobachtungseinheiten beschränkt, bei denen sich sowohl die abhängige als auch die unabhängigen Variablen über die Messzeitpunkte hinweg ändern. Das beinhaltet natürlich, dass man auch nur Effekte für zeitlich variierende unabhängige Variablen schätzen kann. Zweitens muss man gegenüber den Querschnittsmodellen strikte Exogenität und serielle Unabhängigkeit annehmen. Drittens sind die Schätzer im Vergleich zu den Querschnittsmodellen schwerer interpretierbar, da die unbeobachtete Heterogenität weder gemessen noch geschätzt werden kann. Demgegenüber steht, dass bei den fixed-effects-Modellen das Problem der „neglected heterogeneity" grundsätzlich geringer ist.

Auf den ersten Blick scheinen die Nachteile der fixed-effects-Modelle zu überwiegen. In der Regel sind aber fixed-effects-Modelle gegenüber Querschnittsmodellen vorzuziehen, da man bei diesen Modellen keine Annahmen über den Zusammenhang zwischen der unbeobachteten Heterogenität und den unbeobachteten Variablen treffen muss. Dadurch können zwei große Fehlerquellen der Sozialforschung ausgeschlossen werden (vgl. hierzu z. B. Hsiao 1986, S.5–8, Brüderl 2010).

Erstens kann bei praktischen Anwendungen nicht ausgeschlossen werden, dass man alle relevanten unbeobachteten Variablen gemessen und in das Modell aufgenommen hat,

d. h. es liegt ein omitted-variable-bias vor. Eine Variable ist bei nicht linearen Modellen in diesem Sinne relevant, wenn sie mit der abhängigen Variable unter Kontrolle der anderen unabhängigen Variablen korreliert ist. Bei Hsiao und Brüderl wird diese Fehlerquelle auch als Problem der unbeobachteten Heterogenität bezeichnet.

Zweitens besteht bei vielen Anwendungen das Problem, dass sich die Beobachtungseinheiten in den Ausprägungen der interessierenden unabhängigen Variablen unterscheiden. In diesem Zusammenhang werden die unabhängigen Variablen als Treatmentvariablen bezeichnet. Im Gegensatz zu echten Experimenten ist die Zuordnung der Beobachtungseinheiten zur Treatmentgruppe und damit zu den Ausprägungen der unabhängigen Variablen bei Surveydaten nicht zufällig. Bei Surveydaten hängen die unabhängigen Variablen mit anderen Variablen zusammen, d. h. die Beobachtungseinheiten selektieren sich in Anlehnung an die Experimental-Terminologie *selbst* in die Experimental- und Kontrollgruppe.

Die beiden Fehlerquellen sind sehr ähnlich und statistisch häufig äquivalent. Beim omitted-variablen-bias liegt die Betonung auf dem Zusammenhang zwischen den unbeobachteten Variablen mit der abhängigen Variablen; beim Problem der Selbstselektion liegt das Augenmerk auf dem Zusammenhang zwischen der unberücksichtigten Variablen und der interessierenden unabhängigen Variablen.

Die Nachteile der fixed-effects-Modelle beschränken die generelle Verwendung dieser Modelle, aber sofern entsprechende Paneldaten vorliegen, bzw. erhebbar sind, sollten die fixed-effects-Verfahren vorgezogen werden. Selbst wie bei einer bestimmten Fragestellung die Erhebung geeigneter Längsschnittsdaten problematisch sind, sollten die Ergebnisse aus Querschnittsmodellen mindestens mit Ergebnissen aus Längsschnittsmodellen untermauert werden.

4 Implementation

Wir haben im vorherigen Kapitel ausgehend von den Annahmen in Abgrenzung zum Querschnittsmodell für das femlogit-Modell die Likelihoodfunktion abgeleitet. Der Regressionskoeffizientenvektor, der diese Likelihoodfunktion maximiert, ist der konsistente ML-Schätzer. In diesem Kapitel soll nun dargestellt werden, wie das femlogit-Modell in Stata implementiert wird. D. h. wir zeigen die in Stata bereitgestellten Routinen, mit denen man das Maximum der Likelihoodfunktion berechnet.

Das Kapitel besteht aus zwei Teilen. Der erste Teil befasst sich mit der technischen Implementation des femlogit-Modells in Stata. Im zweiten Teil wird ein Performanztest der Implementation dargestellt.

Im Einzelnen werden wir im ersten Teil zuerst auf die allgemeine numerische Optimierung eingehen. Danach wird schematisch die allgemeine Struktur der in Stata bereitgestellten Routinen vorgestellt, mit denen ML-Schätzer geschätzt werden. Dann wird diese Struktur konkret auf die Implementierung des femlogit-Schätzers angewendet. Zuerst werden die notwendigen Ableitungen der logarithmierten Likelihoodfunktion erläutert. Danach wird die konkrete Implementation geschildert. Die Darstellung beginnt bei einem Minimalaufruf der bereitgestellten Stataroutine. Daraufhin wird der eigentliche Kern des Schätzers ausführlich erläutert, und danach werden ausgehend vom Minimalbeispiel die Erweiterungen dargestellt, die für einen eigenständigen Statabefehl notwendig sind.

Im zweiten Teil des Kapitels wird ein einfacher Performanztest der Implementation – wiederum in zwei Schritten – vorgestellt. Im ersten Schritt wird die Implementation des femlogit für den Spezialfall mit nur zwei Alternativen mit der in Stata vorhandenen Umsetzung verglichen. Der zweite Schritt des Performanztests wird die Implementation an einem simulierten Datensatz mit bekannten Regressionskoeffizienten getestet.

4.1 Numerische Maximierung der Likelihoodfunktion

Da die Likelihoodfunktionen im Allgemeinen so wie auch beim femlogit-Modell nicht lineare Funktionen sind, lassen sich die Maxima im Gegensatz zu linearen Modellen nicht analytisch bestimmen. Alternativ dazu wird das Maximum, also der maximum-likelihood-Schätzer numerisch durch ein iteratives Verfahren bestimmt (vgl. Cameron und Trivedi 2009, S. 336–353, Gould, Pitblado, und Poi 2010 und Wooldridge 2010, S. 431–436).

Das gängigste Verfahren zur Berechnung ist das Newton-Raphson-Verfahren (NR), dass im Weiteren ausführlicher dargestellt wird, und für die Schätzung des femlogit-Modells verwendet wird. Andere Möglichkeiten sind das BHHH-, das DFP- und das BFGS-Verfahren. Das BHHH-Verfahren ist sowohl leichter zu berechnen als auch annahmenärmer als das NR-Verfahren, wird aber immer noch seltener verwendet. Der Vorteil besteht darin, dass im Gegensatz zum NR-Verfahren die zweite Ableitung der Likelihoodfunktion nicht berechnet werden muss. Das DFP- und BFGS-Verfahren kommen ebenso ohne die Auswertung der zweiten Ableitung der Likelihoodfunktion aus. Diese Verfahren sind aber im Allgemeinen unstabiler als das NR- und das BHHH-Verfahren (Cameron und Trivedi 2009, S. 344). Die in dieser Arbeit vorgestellte Implementation des femlogit-Modells in Stata erlaubt grundsätzlich eine Schätzung mit allen vier genannten Verfahren. Im gegenwär-

tigen Zustand werden die Schätzer aber mit dem geläufigsten NR-Verfahren geschätzt. Daher soll im Weiteren auch nur dieses Verfahren ausführlicher besprochen werden.

Allgemein ist das NR-Verfahren ein iteratives Optimierungsverfahren. Das Verfahren findet das Maximum einer Funktion, indem es die Nullstelle der Ableitung der Funktion nach dem gesuchten Parameter sucht. Zur Veranschaulichung betrachten wir eine Beispielfunktion über eine Variable x: $f(x) = x/\exp(x)$. Der Graph der Funktion und ihre Ableitung nach x sind in Abbildung 4.1 dargestellt.

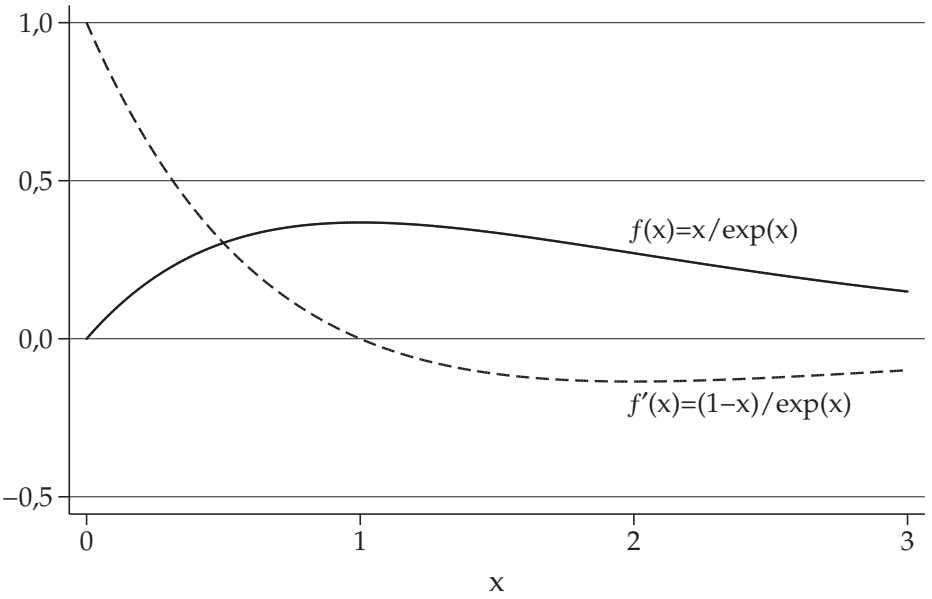


Abbildung 4.1: Beispielfunktion

Die Beispielfunktion hat ein Maximum an der Stelle $x = 1$, d. h. die Ableitungsfunktion $f'(x) = (1 - x)/\exp(x)$ hat an dieser Stelle eine Nullstelle. Diese Nullstelle kann bei nicht linearen Funktionen im Allgemeinen nicht analytisch bestimmt werden. In diesem konkreten Beispiel ist diese Nullstelle leicht ersichtlich, aber prinzipiell nicht analytisch lösbar, da es sich um eine nicht lineare Funktion handelt.

Das iterative NR-Verfahren verläuft folgendermaßen. Man wählt eine beliebige Startstelle x_0 – in unserem Beispiel sei diese Stelle $x_0 = 0$. Nun bestimmt man die Tangente an die Ableitungsfunktion $f'(x)$ an der Stelle β_0:

$$\begin{aligned} T(x_0) &= f'(x_0) + f''(x_0)(x - x_0) \\ &= \frac{1 - x_0}{\exp(x_0)} + \frac{x_0 - 2}{\exp(x_0)}(x - x_0) \\ &= -2x + 1. \end{aligned}$$

Die Nullstelle dieser Tangente $T(x_0)$ ergibt die Stelle für den nächsten Schritt x_1. In unserem Beispiel ist $x_1 = 1/2$. Im zweiten Schritt bildet man die Tangente $T(x_1)$ an die Ableitungsfunktion $f'(x)$ an der Stelle x_1:

$$\begin{aligned} T(x_1) &= f'(x_1) + f''(x_1)(x - x_1) \\ &= \frac{1 - x_1}{\exp(x_1)} + \frac{x_1 - 2}{\exp(x_1)}(x - x_1) \\ &= \frac{1}{2\exp(1/2)} - \frac{3}{2\exp(1/2)}(x - 1/2) \\ &= \left(-\frac{3}{2}x + \frac{5}{4} \right) \exp\left(\frac{1}{2}\right). \end{aligned}$$

In Abbildung 4.2 ist der erste und zweite Iterationsschritt dargestellt.

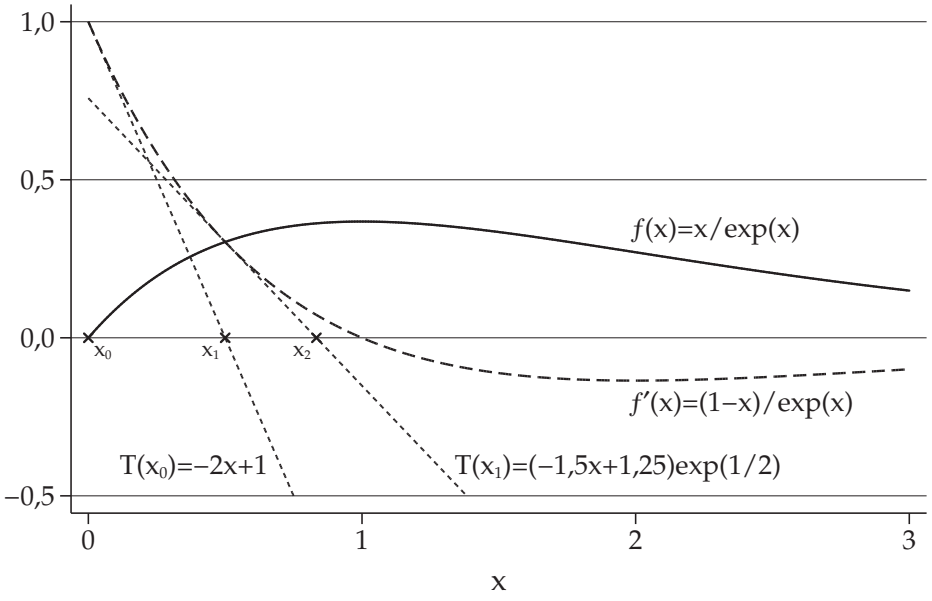


Abbildung 4.2: Beispielfunktion mit Tangente

Allgemein ist für einen beliebigen Iterationsschritt r der Folgewert x_{r+1} folgendermaßen definiert:

$$x_{r+1} = x_r - \frac{f'(x_r)}{f''(x_r)}.$$

Das Verfahren lässt sich einfach von einer Funktion über eine Variable auf eine Funktion über mehrere Variablen verallgemeinern, wie sie bei der logarithmierten Likelihoodfunktion $\mathrm{L}(\beta) = \mathrm{E}(\ln \ell_i(\beta))$ vorliegt. Wir betrachten also z. B. $\mathrm{L}(\beta)$ als die Funktion, für die

das Maximum über den Vektor der Regressionskoeffizienten β gefunden werden muss. Im Maximum ist die Funktion $\mathrm{L}(\beta)$ flach, d. h. im Maximum sind alle Ableitungen nach allen Parametern β_{jm} null. D. h. analog zu oben suchen wir die Nullstelle für die vektorwertige score-Funktion $\mathrm{s}(\beta)$:

$$\begin{aligned}\mathrm{s}(\beta) &= \frac{\partial \mathrm{L}(\beta)}{\partial \beta} = \left(\frac{\partial \mathrm{L}(\beta)}{\partial \beta_1}, \ldots, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_{B-1}}, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_{B+1}}, \ldots, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_J}\right) \\ &= \left(\frac{\partial \mathrm{L}(\beta)}{\partial \beta_{11}}, \ldots, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_{B-1,M}}, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_{B+1,1}}, \ldots, \frac{\partial \mathrm{L}(\beta)}{\partial \beta_{JM}}\right).\end{aligned} \tag{4.1}$$

Da $\mathrm{s}(\beta)$ vektorwertig ist, ist die Nullstelle die Stelle, für die alle Ableitungen null sind. Für das Iterationsverfahren wählt man wie oben einen Startvektor β_0, z. B. den Nullvektor:

$$\beta_0 = (\beta_{11_0}, \ldots, \beta_{B-1,M_0}, \beta_{B+1,1_0}, \ldots, \beta_{JM_0}) = (0, \ldots, 0).$$

Für jeden Iterationsschritt r ist der Folgewert β_{r+1} gegeben durch:

$$\beta_{r+1} = \beta_r - \mathrm{H}(\beta_r)^{-1}\mathrm{s}(\beta_r)'.$$

In diesem Ausdruck ist $\mathrm{H}(\beta_r)$ die Matrix über die zweiten Ableitungen der Likelihoodfunktion über die Regressionskoeffizienten, d. h.

$$\mathrm{H}(\beta) = \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta' \partial \beta} \tag{4.2}$$

$$= \begin{pmatrix} \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_1' \partial \beta_1} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_1' \partial \beta_{B-1}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_1' \partial \beta_{B+1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_1' \partial \beta_J} \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1}' \partial \beta_1} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1}' \partial \beta_{B-1}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1}' \partial \beta_{B+1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1}' \partial \beta_J} \\ \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1}' \partial \beta_1} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1}' \partial \beta_{B-1}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1}' \partial \beta_{B+1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1}' \partial \beta_J} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_J' \partial \beta_1} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_J' \partial \beta_{B-1}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_J' \partial \beta_{B+1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_J' \partial \beta_J} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{11} \partial \beta_{11}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{11} \partial \beta_{B-1,M}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{11} \partial \beta_{B+1,1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{11} \partial \beta_{JM}} \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1,M} \partial \beta_{11}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1,M} \partial \beta_{B-1,M}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1,M} \partial \beta_{B+1,1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B-1,M} \partial \beta_{JM}} \\ \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1,1} \partial \beta_{11}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1,1} \partial \beta_{B-1,M}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1,1} \partial \beta_{B+1,1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{B+1,1} \partial \beta_{JM}} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{JM} \partial \beta_{11}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{JM} \partial \beta_{B-1,M}} & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{JM} \partial \beta_{B+1,1}} & \cdots & \frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{JM} \partial \beta_J M} \end{pmatrix}.$$

Das NR-Verfahren konvergiert im Allgemeinen sehr schnell gegen die Nullstelle, vorausgesetzt, dass die Startstelle ausreichend nahe an der Nullstelle gewählt wird. Wenn die betrachtete Funktion lokale Extrema oder Sattelpunkte hat, kann das NR-Verfahren auch dort konvergieren. Wenn die Funktion aber global konkav ist, was bei der Likelihoodfunktion der logistischen Regression nach McFadden (1974, S. 115) immer gegeben ist,

tritt dieses Problem nicht auf, d. h. bei den in dieser Arbeit betrachteten Modellen sollte das NR-Verfahren immer gegen das globale Maximum der Likelihoodfunktion konvergieren. Technisch betrachtet man das Verfahren als konvergiert, wenn die Veränderung der Funktionswerte eine vorher gewählte, sehr kleine Schranke unterschreitet.

Aus diesen Ausführungen folgt für die Implementation, dass man neben der analytischen Ableitung der Likelihoodfunktion – wie in Gleichung 3.3 und 3.17 – die erste und zweite Ableitung der Likelihoodfunktion nach dem Vektor der Regressionskoeffizienten benötigt. Wenn die Ableitungen nicht als analytisch abgeleitete Ausdrücke vorliegen, können sie auch numerisch bestimmt werden (vgl. Stoer und Bulirsch 1983, S. 122, Späth 1994, Kap. 3). Da dies aber höheren Rechenaufwand darstellt, sollte man soweit möglich immer die abgeleiteten Ausdrücke verwenden.

Das hier dargestellte Verfahren ist das einfache Newton-Verfahren. In Stata wird das sehr ähnliche Newton-Raphson-Verfahren (NR) verwendet. Der Vorteil dieses Verfahrens besteht in der schnelleren Berechnung. Beim einfachen Newton-Verfahren wird für jeden Iterationsschritt die analytische oder numerische Berechnung der ersten beiden Ableitungen benötigt. Beides benötigt im Allgemeinen einen großen Rechenaufwand. Beim NR-Verfahren wird der Rechenaufwand dadurch reduziert, dass bei jedem Iterationsschritt die „Richtung" der Änderung $\mathrm{H}(\beta_r)^{-1}\mathrm{s}(\beta_r)'$ einmal berechnet wird, und vom ausgehenden Regressionskoeffizientenvektor β_r verschiedene Schrittweiten in Richtung der Änderung „ausprobiert" werden, bis sich die entsprechende Likelihood nicht mehr vergrößert. Erst dann wird die erste und zweite Ableitung für den so gefundenen Vektor β_{r+1} neu berechnet, und damit eine neue „Änderungsrichtung". Dadurch wird im Allgemeinen Rechenzeit gespart, da die mehrfache Auswertung der Likelihoodfunktion für die einzelnen „ausprobierten" Schrittweiten schneller berechnet werden können, als die Ableitungen, selbst wenn so nicht immer in der optimalen „Richtung" gesucht wird. Für die konkrete Implementation bedeutet das, dass man bei der Iteration grundsätzlich unterscheidet, ob man beim jeweiligen Berechnungsschritt eine neue Schrittweite ausprobiert, oder ob man eine neue Richtung berechnet.

Unterschiede zwischen dem „conditional" und „unconditional maximum likelihood"-Ansatz

An dieser Stelle ist darauf hinzuweisen, dass die in Stata bereitgestellten Routinen für das ML-Verfahren *nicht* die Likelihoodfunktion und deren erste und zweite Ableitung in der Form erwarten, wie wir sie in den obigen Ausführungen definiert haben. Wir haben die logarithmierte Likelihoodfunktion $\mathrm{L}(\beta)$ als den Erwartungswert der logarithmierten Likelihood-Funktion $\ln \ell_i(\beta)$ für die Beobachtungseinheit i über die gesamte Stichprobe definiert, d. h.: $\mathrm{L}(\beta) = \mathrm{E}(\ln \ell_i(\beta)) = 1/N \sum_{i=1}^{N} \ln \ell_i(\beta)$. Entsprechend zieht sich der Erwartungswert durch die Ausdrücke für die erste Ableitung $\mathrm{s}(\beta) = \mathrm{E}(\mathrm{s}_i(\beta)) = 1/N \sum_{i=1}^{N} \mathrm{E}(\mathrm{s}_i(\beta))$ und die zweite Ableitung $\mathrm{H}(\beta) = \mathrm{E}(\mathrm{H}_i(\beta)) = 1/N \sum_{i=1}^{N} \mathrm{E}(\mathrm{H}_i(\beta))$.

Die Routinen in Stata erwarten dagegen für die logarithmierte Likelihoodfunktion nur die Summe der logarithmierten Likelihoodfunktionen für jede Beobachtung über die Stichprobe: $\widetilde{\mathrm{L}}(\beta) = \sum_{i=1}^{N} \ln \ell_i(\beta)$. Entsprechend werden auch bei den Ableitungen die entsprechenden Ausdrücke *ohne* den Faktor $1/N$ erwartet. Der Unterschied ergibt sich daraus, dass unsere Definition aus dem „conditional maximum likelihood"-Ansatz folgt,

wohingegen Stata vom „unconditional maximum likelihood"-Ansatz ausgeht (vgl. Wooldridge 2010, S. 470f.).[1]

Wie bei der konkreten Umsetzung noch genauer erläutert wird, werden zur besseren Vergleichbarkeit mit der bestehenden Implementierung des felogit-Modells die logarithmierten Likelihoodfunktionen und die Ableitungen „unconditional maximum likelihood"-Ansatz ohne den Faktor $1/N$ verwendet. Für die Schätzung der Regressionskoeffizienten ist der Unterschied völlig irrelevant, da der Faktor natürlich von β unabhängig ist. Der einzige Unterschied besteht darin, dass grundsätzlich unterschiedliche Likelihoodwerte berichtet werden müssen. Die typischen Fitmaße wie das Pseudo-R^2 von McFadden (1974) und die LR-Statistik sind nicht von diesem Unterschied betroffen (vgl. Wooldridge 2010, S. 481f.).

4.2 Umsetzung des femlogit-Modells in Stata

In diesem Abschnitt soll dargestellt werden, wie das femlogit-Modell in Stata konkret umgesetzt wird. Die Ausführungen der in Stata bereitgestellten Befehle folgen Gould et al. (2010), StataCorp LP (2009d) und StataCorp LP (2009a).

Dieser Teil bildet den eigentlichen wissenschaftlichen eigenständigen Kern der vorliegenden Arbeit. Wie in der Einleitung in Abschnitt 1.2 bereits dargestellt wurde, wurde das femlogit-Modell zwar schon in der Literatur verwendet. Jedoch verwenden die bestehenden Umsetzungen – soweit bekannt – einen Datentransformationstrick, so dass das Modell mit einer mlogit-Modell-Implementation geschätzt werden kann. Hierbei muss aber die Anzahl der Alternativen und die Anzahl der Messzeitpunkte stark eingeschränkt werden. Die im Weiteren dargestellte Implementation überwindet diese Beschränkung und liefert eine allgemein anwendbare Lösung in der gängigen Statistiksoftware Stata.

4.2.1 Allgemeine Struktur des ML-Verfahrens in Stata

Für das maximum-likelihood-Verfahren wird in Stata auf der Skriptsprachenebene, d. h. der bekannten Statacodeebene, das Kommandopaket `ml` bereitgestellt. Auf der darunter liegenden Matrixsprachenebene Mata steht das Funktionenpaket `moptimize()` zur Verfügung. Wenn man mit Stata ein Modell mit dem ML-Verfahren schätzen möchte, für das nicht schon eine vorgefertigte Implementation vorliegt, wird die Verwendung des Statakommandos `ml` empfohlen. Bei komplexeren und rechenzeitintensiven Problemen ist man auf die Matafunktion `moptimize()` angewiesen. Bevor auf die konkrete Umsetzung eingegangen wird, wird die abstrakte Struktur dieser Kommando dargestellt.

Sowohl beim Statakommando `ml` als auch bei der Matafunktion `moptimize()` besteht das Kommando aus zwei Teilen. Wie in Abbildung 4.3 schematisch dargestellt ist, wird mit dem eigentlichen Kommando `ml` oder `moptimize()` das Optimierungsproblem definiert und damit die abhängige und die unabhängigen Variablen mit der Likelihoodfunktion und dem Iterationsverfahren in Beziehung gesetzt.

1 Der Vorteil beim „conditional"-Ansatz besteht darin, dass man nur eine Verteilung für die abhängige Variable bedingt auf die unabhängigen Variablen benötigt. Beim „unconditional"-Ansatz muss man eine gemeinsame Verteilung über die abhängige Variable und die unabhängigen Variablen annehmen, oder die unabhängigen Variablen als konstant annehmen.

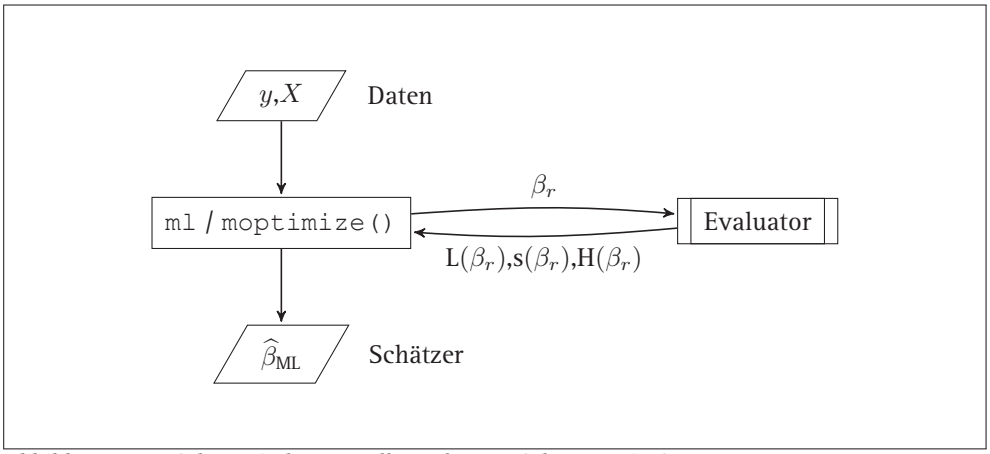


Abbildung 4.3: Schematische Darstellung der ML-Schätzung in Stata

Wie im Weiteren und bei der Umsetzung des femlogit-Modells noch deutlicher wird, wird hier die Analysestichprobe festgelegt. Weiterhin wird der Vektor der bzw. potentiell mehrerer abhängiger Variablen y_d definiert. Ferner legt man mit dem eigentlichen Kommando die Matrix der unabhängigen Variablen X fest. Neben dem Bezug zu den konkreten Daten wird hier auch der Bezug zur Likelihoodfunktion definiert. Hier werden die der Likelihoodfunktion zugrunde liegenden Modellgleichungen definiert, d. h. es wird bestimmt, wie viele Gleichungen es gibt und wie viele Regressionskoeffizienten und andere zu schätzende Parameter es gibt.

Der zweite Teil des Kommandos ist ein separates „Evaluator”-Unterprogramm, dass aus einem gegebenen Koeffizientenvektor β_r den zugehörigen logarithmierten Likelihoodwert und die Auswertung der ersten und zweiten Ableitungen für diesen Koeffizientenvektor besteht. Nach dem Aufruf durch `ml` oder `moptimize()` schickt Stata mit jedem Iterationsschritt r den Koeffizientenvektor β_r an den Evaluator. Dieser gibt L(β_r), s(β_r) und H(β_r) zurück. Daraus wird der Koeffizientenvektor des nächsten Iterationsschritts β_{r+1} berechnet. Wenn die Änderung des Koeffizientenvektors unter eine bestimmte Schranke fällt, gilt das Verfahren als konvergiert, und der letzte Koeffizientenvektor $\widehat{\beta}_{\text{ML}}$ wird als Ergebnis ausgegeben.

Bei der Implementation des femlogit-Modells verwenden wir für beide Komponenten die Mata-Funktionen, d. h. die Daten werden an `moptimize()` übergeben, und der Evaluator ist ebenso eine Mata-Funktion. Wie im nächsten Abschnitt genauer erläutert wird, haben die Mata-Funktionen gegenüber den Stata-Befehlen für unsere Implementation mehrere Vorteile: Erstens können wir durch die Verwendung von `moptimize()` statt `ml` den Umstand, dass die Likelihoodfunktion und deren Ableitungen für einzele Beobachtungseinheiten definiert, direkt umsetzen. Zweitens können wir durch die Umsetzung des Evaluators in Mata im Gegensatz zu einer Umsetzung mit Stata die Likelihoodfunktion direkt umsetzen. Drittens wird eine Umsetzung mit Mata – insbesondere beim separaten Evaluator – in der Regel in Stata schneller ausgeführt.

4.2.2 Score-Funktion und Hesse-Matrix

Wie oben ausgeführt wurde, benötigt man für das iterative Verfahren die ersten beiden Ableitungen. Entweder werden diese aus den Likelihoodfunktion numerisch bestimmt oder man gibt die Ableitungen als analytisch abgeleitete Ausdrücke an. Im zweiten Fall erhöht sich sowohl die Rechengenauigkeit als auch die Rechengeschwindigkeit. Für die vorliegende Implementation benötigen wir also die logarithmierte Likelihoodfunktion für die einzelne Beobachtungseinheit $\ln \ell_i(\beta)$. Weiterhin benötigen wir die erste Ableitung der logarithmierten Likelihoodfunktion für die einzelne Beobachtungseinheit $\mathrm{s}_i(\beta)$ und die zweite Ableitung der logarithmierten Likelihoodfunktion für die gesamte Stichprobe $\mathrm{H}(\beta) = \mathrm{E}(\mathrm{H}_i(\beta))$. Die Likelihoodfunktion und deren erste Ableitung für die Stichprobe $\mathrm{L}(\beta)$ und $\mathrm{s}(\beta)$ werden durch die Mittlung über die Beobachtungseinheiten von Stata automatisch gebildet.[2]

Wie haben die Likelihoodfunktion für die Beobachtungseinheit i in Abschnitt 3.2.2 in Gleichung 3.16 bereits definiert.

$$\ell_i(\beta) = \frac{\exp(\sum_{t=1}^{T_i} \sum_{j=1,j\neq B}^{J} \delta_{y_{it},j} X_{it}\beta_j')}{\sum_{d_i \in \Delta_i} \exp(\sum_{t=1}^{T_i} \sum_{j=1,j\neq B}^{J} \delta_{d_{it},j} X_{it}\beta_j')}.$$

Wenn man den Ausdruck logarithmiert, erhält man:

$$\ln \ell_i(\beta) = \sum_{t=1}^{T_i} \sum_{j=1,j\neq B}^{J} \delta_{y_{it},j} X_{it}\beta_j' - \ln \sum_{d_i \in \Delta_i} \exp(\sum_{t=1}^{T_i} \sum_{j=1,j\neq B}^{J} \delta_{d_{it},j} X_{it}\beta_j'). \tag{4.3}$$

Die erste Ableitung der logarithmierten Likelihood-Funktion $\ln \ell_i(\beta)$ für die Beobachtungseinheit i nach dem Vektor der Regressionskoeffizienten haben wir oben als score-Funktion eingeführt. Der score-Vektor für die Beobachtungseinheit i ist gegeben mit:

$$\begin{aligned} \mathrm{s}_i(\beta) &= \frac{\partial \ln \ell_i(\beta)}{\partial \beta} \\ &= \left(\frac{\partial \ln \ell_i(\beta)}{\partial \beta_1}, \ldots, \frac{\partial \ln \ell_i(\beta)}{\partial \beta_{B-1}}, \frac{\partial \ln \ell_i(\beta)}{\partial \beta_{B+1}}, \ldots, \frac{\partial \ln \ell_i(\beta)}{\partial \beta_J} \right) \\ &= \left(\frac{\partial \ln \ell_i(\beta)}{\partial \beta_{11}}, \ldots, \frac{\partial \ln \ell_i(\beta)}{\partial \beta_{B-1,M}}, \frac{\partial \ln \ell_i(\beta)}{\partial \beta_{B+1,1}}, \ldots, \frac{\partial \ln \ell_i(\beta)}{\partial \beta_{JM}} \right). \end{aligned} \tag{4.4}$$

2 Die Form, in der die Likelihoodfunktion und die Ableitungen an Stata übergeben werden, hängt vom Modell und der konkreten Ausgestaltung des Evaluators ab (vgl. Gould et al. 2010, S.48–51). Die hier gewählte Form trägt erstens dem Umstand Rechnung, dass die Likelihoodfunktion nur für Beobachtungseinheiten und nicht für Beobachtungen definiert ist. Zweitens erlaubt die Spezifikation grundsätzlich eine einfache Berücksichtigung von komplexen Stichproben und heteroskedastie-robuster Standardfehler.

Jeder einzelne Zelleintrag des Vektors ist also die partielle Ableitung der logarithmierten Likelihood-Funktion $\ln \ell_i(\beta)$ für die Beobachtungseinheit i nach dem entsprechenden Regressionskoeffizienten β_{jm}:

$$\frac{\partial \ln \ell_i(\beta)}{\partial \beta_{jm}} = \sum_{t=1}^{T_i} \delta_{y_{it},j} x_{itm} - \frac{\sum_{d_i \in \Delta_i} \left(\left(\sum_{t=1}^{T_i} \delta_{d_{it},j} x_{itm} \right) \exp\left(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} \delta_{d_{it},j} X_{it} \beta_j' \right) \right)}{\sum_{d_i \in \Delta_i} \exp\left(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} \delta_{d_{it},j} X_{it} \beta_j' \right)}. \tag{4.5}$$

Um den Ausdruck der zweiten Ableitung der logarithmierten Likelihoodfunktion für die gesamte Stichprobe $\mathrm{E}(\mathrm{H}_i(\beta))$ zu entwickeln, gehen wir schrittweise vor. Wir haben bereits oben in Gleichung 4.2 ausgeführt, dass die zweite Ableitung der logarithmierten Likelihoodfunktion für die gesamte Stichprobe $\mathrm{H}(\beta)$ eine $(J-1)M \times (J-1)M$-Matrix ist:

$$\mathrm{H}(\beta) = \left(\frac{\partial \mathrm{L}(\beta)^2}{\partial \beta_{jm} \partial \beta_{lk}} \right) = \left(\frac{\partial \mathrm{E}\left(\ln \ell_i(\beta) \right)^2}{\partial \beta_{jm} \partial \beta_{lk}} \right) = \left(\mathrm{E}\left(\frac{\partial \ln \ell_i(\beta)^2}{\partial \beta_{jm} \partial \beta_{lk}} \right) \right). \tag{4.6}$$

Der letzte Schritt ergibt sich daraus, dass der Erwartungswert und die Differentiation lineare Funktionen sind, d. h. die Reihenfolge vertauscht werden können.

Der Ausdruck erhält seinen Sinn erst durch die Zelleinträge. In jeder Zelle ist der Erwartungswert der zweifachen Ableitung der logarithmierten Likelihoodfunktion $\ln \ell_i(\beta)$ nach dem Regressionskoeffizienten β_{mj} und nach dem Regressionskoeffizienten β_{lk}. Wie schon gesagt, ergibt sich daraus eine Matrix über alle Kombinationen. Jeder Zelleintrag ist also gegeben mit:

$$\begin{aligned}
&\mathrm{E}\left(\frac{\partial \ln \ell_i(\beta)^2}{\partial \beta_{jm} \partial \beta_{lk}} \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \Bigg(\Bigg(\sum_{d_i \in \Delta_i} \Big(\Big(\sum_{t=1}^{T_i} d_{itj} x_{itm} \Big) \exp\Big(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} d_{itj} X_{it} \beta_j' \Big) \Big) \\
&\quad \cdot \sum_{d_i \in \Delta_i} \Big(\Big(\sum_{t=1}^{T_i} d_{itl} x_{itk} \Big) \exp\Big(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} d_{itj} X_{it} \beta_j' \Big) \Big) \Bigg/ \\
&\quad \Big(\exp\Big(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} d_{itj} X_{it} \beta_j' \Big) \Big)^2 \Bigg) - \Bigg(\sum_{d_i \in \Delta_i} \Big(\Big(\sum_{t=1}^{T_i} d_{itj} x_{itm} \Big) \\
&\quad \cdot \Big(\sum_{t=1}^{T_i} d_{itl} x_{itk} \Big) \exp\Big(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} d_{itj} X_{it} \beta_j' \Big) \Big) \Bigg/ \\
&\quad \Big(\exp\Big(\sum_{t=1}^{T_i} \sum_{j=1, j \neq B}^{J} d_{itj} X_{it} \beta_j' \Big) \Big) \Big) \Bigg).
\end{aligned} \tag{4.7}$$

Es ist zu beachten, dass Stata – wie oben bereits erläutert wurde – die unbedingte Likelihood mit den entsprechenden Ableitungen erwartet, d. h. im übergebenen Ausdruck wird der Quotient $\frac{1}{N}$ weggelassen.

Hilfsvariablen

Für eine effiziente Implementation werden die komplexen Ausdrücke der logarithmierten Likelihoodfunktion und der ersten beiden Ableitungen als einfache Ausdrücke über Hilfsterme gebildet. Der Vorteil besteht erstens in einer besser organisierten und leichter nachvollziehbareren Darstellung der Programmierung und zweitens in der effizienten Vermeidung von unnötigen doppelten Berechnungen von Teilausdrücken.

Zur einfacheren Darstellung und einfacheren Berechnung werden also folgende Hilfsterme definiert. Auf der tiefsten Ebene definieren wir zwei elementare Bausteine Z1 und Z2. Für jede Beobachtungseinheit i und zusätzlich für jede Permutation d_i aus der Menge der Permutationen Δ_i, wie wir sie bereits in Abschnitt 3.2.2 definiert haben, ist Z1 dabei eine skalare Größe und Z2 ein Vektor mit einem Zelleintrag für alle möglichen Regressionskoeffizienten β_{mj}. D. h. wir definieren also:

$$
\begin{aligned}
\mathrm{Z1} &= \exp\Big(\sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{d_{it},j}X_{it}\beta_j'\Big),\\
\mathrm{Z2} &= \big(\mathrm{Z2}_{11},\ldots,\mathrm{Z2}_{B-1,M},\mathrm{Z2}_{B+1,1},\ldots,\mathrm{Z2}_{JM}\big)\\
&= \Bigg(\sum_{t=1}^{T_i}\delta_{d_{it},1}x_{it1},\ldots,\sum_{t=1}^{T_i}\delta_{d_{it},B-1}x_{itM},\sum_{t=1}^{T_i}\delta_{d_{it},B+1}x_{it1},\ldots,\sum_{t=1}^{T_i}\delta_{d_{it},J}x_{itM}\Bigg).
\end{aligned}
$$

Weiter definieren wir unter Zuhilfenahme dieser elementaren Bausteine die darüber liegenden Hilfsterme A, B, C, D und E. A und B sind dabei für jede Beobachtungseinheit i skalare Größen. C und D sind für jede Beobachtungseinheit i Vektoren über alle Regressionskoeffizienten β_{mj}. Schließlich ist E eine Matrix über alle Kombinationen der Regressionskoeffizienten β_{mj} und β_{lk}. D. h. die Hilfsterme sind folgendermaßen definiert:

$$
\begin{aligned}
\mathrm{A} &= \sum_{t=1}^{T_i}\sum_{j=1,j\neq B}^{J}\delta_{y_{it},j}X_{it}\beta_j',\\
\mathrm{B} &= \sum_{d_i\in\Delta_i}\mathrm{Z1},\\
\mathrm{C} &= \big(\mathrm{C}_{11},\ldots,\mathrm{C}_{B-1,M},\mathrm{C}_{B+1,1},\ldots,\mathrm{C}_{JM}\big)\\
&= \Bigg(\sum_{t=1}^{T_i}\delta_{y_{it},1}x_{it1},\ldots,\sum_{t=1}^{T_i}\delta_{y_{it},B-1}x_{itM},\sum_{t=1}^{T_i}\delta_{y_{it},B+1}x_{it1},\ldots,\sum_{t=1}^{T_i}\delta_{y_{it},J}x_{itM}\Bigg),\\
\mathrm{D} &= \big(\mathrm{D}_{11},\ldots,\mathrm{D}_{B-1,M},\mathrm{D}_{B+1,1},\ldots,\mathrm{D}_{JM}\big)\\
&= \sum_{d_i\in\Delta_i}\mathrm{Z2}\cdot\mathrm{Z1},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E} &= \begin{pmatrix}
\mathrm{E}_{1111} & \cdots & \mathrm{E}_{1,1,B-1,M} & \mathrm{E}_{1,1,B+1,1} & \cdots & \mathrm{E}_{11JM} \\
\vdots & \ddots & \vdots & \vdots & & \vdots \\
\mathrm{E}_{B-1,M,1,1} & \cdots & \mathrm{E}_{B-1,M,B-1,M} & \mathrm{E}_{B-1,M,B+1,1} & \cdots & \mathrm{E}_{B-1,M,J,M} \\
\mathrm{E}_{B+1,1,1,1} & \cdots & \mathrm{E}_{B+1,1,B-1,M} & \mathrm{E}_{B+1,1,B+1,1} & \cdots & \mathrm{E}_{B+1,1,J,M} \\
\vdots & & \vdots & \vdots & \ddots & \vdots \\
\mathrm{E}_{JM11} & \cdots & \mathrm{E}_{J,M,B-1,M} & \mathrm{E}_{J,M,B+1,1} & \cdots & \mathrm{E}_{JMJM}
\end{pmatrix} \\
&= \sum_{d_i \in \Delta_i} \mathrm{Z2}'\mathrm{Z2} \cdot \mathrm{Z1}.
\end{aligned}
$$

Mit diesen Hilfstermen lassen sich die logarithmierte Likelihoodfunktion $\ln \ell_i(\beta)$, die score-Funktion $\mathrm{s}_i(\beta)$ und die Hesse-Matrix $\mathrm{H}(\beta)$ folgendermaßen vereinfacht darstellen:

$$
\begin{aligned}
\ln \ell_i(\beta) &= \mathrm{A} - \ln B, \\
\mathrm{s}(\beta) &= C - \frac{D}{B}, \\
\mathrm{H}(\beta) &= \left(\sum_{i=1}^{N} \frac{D'D}{B^2} - \frac{E}{B} \right).
\end{aligned}
\tag{4.8}
$$

4.2.3 Programmcode

Nachdem wir nun analytische Ausdrücke für die Ableitungen entwickelt haben, können wir nun die Umsetzung des femlogit-Modells in Stata angehen. Die Darstellung der Implementation erfolgt von „innen nach aussen”, d. h. wir beginnen mit einer sehr konkreten, einfachen, minimalen Umsetzung und entwickeln diese danach schrittweise zum eigenständigen Statakommando.

Minimalbeispiel des Aufrufs der `moptimize()`-Funktion

Wie schon erwähnt, verwenden wir bei der Implementation des femlogit-Modells für die Übergabe der Daten und der Definition des Maximierungsproblems die Mata-Funktion `moptimize()`. Ebenso wird das Evaluator-Unterprogramm als Mata-Funktion geschrieben. Wir betrachten für die Implementation die abhängige Variable y, die in Stata als `y` abgebildet ist, und die unabhängigen Variablen X, die als `X1`,...,`XM` abgebildet sind. Ferner liegt eine Indikatorvariable `id` vor, die bezeichnet, welche Beobachtungen zur selben Beobachtungseinheit gehören. Wir gehen vorerst davon aus, dass die Anzahl der Ausprägungen J der abhängigen Variable von vornherein bekannt ist. Außerdem legen wir die Referenzkategorie mit $B = 1$ fest. Die einfachste Umsetzung eines femlogit-Modells ist dann folgendermaßen aufgebaut (vgl. StataCorp LP 2009a; Gould et al. 2010).

Aufruf von moptimize()

```
1  // Übergang von Stata zu Mata
2  mata:
3
4  // Initializiere Problem
5  M=moptimize_init()
6
7  // Abhängige Variable
8  moptimize_init_depvar(M,1,"y")
9
10 // Unabhängige Variablen
11 for(j=2;j<=J;j++) {
12    moptimize_init_eq_indepvars(M,j,"X1 … XM")
13    moptimize_init_eq_cons(M,j,"off")
14 }
15
16 // Startwerte
17 moptimize_init_search(M,"off")
18 moptimize_init_search_rescale(M,"off")
19
20 // Evaluator
21 moptimize init evaluator(M,&femlogit_eval_gf2())
22 moptimize init evaluatortype("gf2")
23
24 // Aufruf des Optimierungsalgorithmus
25 moptimize(M)
26
27 // Ausgabe der Ergebnistabelle
28 moptimize result display(M)
```

Was bewirken diese Kommandozeilen? Gehen wir die Codezeiten schrittweise durch: Mit dem Befehl in Zeile 2 geht man von der Stataskript-Ebene zur Mata-Ebene über. Ab dieser Stelle werden Statabefehle nicht mehr akzeptiert, sondern nur noch Matafunktionsaufrufe. Die Übergabe der Daten und die Definition erfolgt mit einer Reihe von Befehlen. In Zeile 5 wird das Optimierungsproblem initialisiert. D. h. es wird die abstrakte Variable `M` definiert, in der alle Bestandteile des Optimierungsproblems, dass mit dem maximum-likelihood-Verfahren gelöst werden soll, gesammelt werden.[3] In Zeile 8 wird die abhängige Variable `y` definiert. In Zeile 12 wird für jede Gleichung, d. h. also jedem Kontrast zwischen den Ausprägungen $2, \ldots, J$ zur Referenzkategorie 1, der gleiche Vektor der unabhängigen Variablen $X = (X_1, \ldots, X_M)$ bestimmt. In Zeile 13 wird klargestellt, dass der zu den unabhängigen Variablen zugehörige Vektor der Regressionskoeffizienten keine Konstante enthalten soll, sondern nur Koeffizienten für die Variablen. In den Zeilen ab 16 wird der Startwert für die Regressionskoeffizienten festgelegt. Standardmäßig wird ein Nullvektor als Startwert verwendet. Dieser soll nicht durch eine Zufallssuche verbessert werden, und der Vektor soll auch nicht durch Reskalierungsverfahren verbessert werden. In Zeile 21 wird nun die Evaluatorfunktion `femlogit_eval_gf2()` zugeordnet. Mit dieser Funktion, die wir im Folgenden ausführlich diskutieren, werden die Likelihoodfunktion und die ersten beiden Ableitungen berechnet. In Zeile 22 wird der Typ der Evaluatorfunktion definiert. Hier wird festgelegt, ob der Evaluator die Likelihoodfunktion für jede Beobachtung, jede unabhängige Beobachtungseinheit oder für die

3 Es ist anzumerken, dass man mit `moptimize()`-Funktion neben dem maximum-likelihood-Verfahren auch eine quadratische Optimierung zur Bearbeitung von GLS-Verfahren verwenden kann.

gesamte Stichprobe zurückgibt. Weiterhin wird mit dem Typ festgelegt, ob der Evaluator die ersten beiden Ableitungen durch einen analytischen Ausdruck selbst berechnet, oder ob diese aus der Likelihoodfunktion numerisch bestimmt werden. Schließlich wird in Zeile 25 das eigentliche maximum-likelihood-Verfahren gestartet, und mit Zeile 25 nach Erreichen der Konvergenz die Ergebnistabelle ausgegeben.

Evaluator `femlogit_eval_gf2()`

Es bleibt nun eine große Blackbox übrig, nämlich der Evaluator selbst. Hier wird im Wesentlichen in jedem Iterationsschritt für einen gegebenen Regressionskoeffizientenvektor die Likelihoodfunktion und deren Ableitungen bestimmt, d. h. dieses Unterprogramm ist der eigentliche Kern eines in Stata implementierten maximum-likelihood-Schätzers. Im Weiteren soll die für das femlogit-Modell verwendete Evaluatorfunktion `femlogit_eval_gf2()` ausführlich dargestellt werden. Es ist zu beachten, dass hier die vollständige und endgültige Fassung vorgestellt wird, so dass bei einzelnen technischen Details vorgegriffen werden muss:[4]

femlogit_eval_gf2()

```
1  mata set matastrict on
2  void femlogit_eval_gf2(transmorphic scalar ML, real scalar todo, /*
3     */ real rowvector b, real colvector lnfj, real matrix S, /*
4     */ real matrix H) {
5
6     // declare variables
7     real colvector lnli, touse, id, yi, d
8     real matrix gi, Hi, panelinfo, Xi, out2eq, X, E
9     real scalar N, M, J, i, A, B, j, m, Z1
10    real rowvector C, D, permuteinfo, Z2
11
12    // get things from Stata
13    st_view(touse=.,.,st_local("touse"))
14    st_view(id=.,.,st_local("group"),st_local("touse"))
15    st_view(X=.,.,st_local("rhs"),st_local("touse"))
16
17    // moptimize_util_depvar(M,1)
18    out2eq=st_matrix(st_local("out2eq"))
```

In diesem ersten Teil finden im Wesentlichen zwei Dinge statt. Erstens werden im Teil ab Zeile 6 alle im Evaluator verwendeten Variablen typen-deklariert.[5] Zweitens werden alle notwendigen Informationen aus Stata bzw. aus der aufrufenden Mata-Funktion in den Evaluator geholt. Die Übergabe erfolgt an zwei Stellen. In Zeile 2 werden aus der `moptimize()`-Funktion die Problemdefinition `ML`, ein Zähler `todo`, der aktuelle Regressionskoeffizientenvektor `b` und die Variablen `lnfj`, `S` und `H` übergeben. Die Variable `ML` ist die Variable `M` von oben, die hier, um Verwechslung mit der Anzahl der unabhängigen Variablen M zu vermeiden, umbenannt ist. Mit der Variable `ML` können verschiedene, mit der Problemdefinition festgelegte Informationen an den Eva-

4 Aus Platzgründen wurde bei der Darstellung auf einzelne Kommentare verzichtet.

5 Bei Mata ist Typen-Deklaration nicht zwingend notwendig, erleichtert aber erstens die Fehlersuche, und führt zweitens zu einer geringfügig höheren Rechengeschwindigkeit.

luator übergeben werden. Hier ist dies nur die abhängige Variable y, die im Evaluator mit `moptimize_util_depvar(ML,1)` abgerufen wird.

Mit dem `todo`-Zähler wird der Ablauf des Evaluators effizienter organisiert. Wie in Abschnitt 4.1 auf Seite 69 kurz dargestellt wurde, muss man beim NR-Verfahren den Evaluator so organisieren, dass man die Ableitungen nur unter bestimmten Umständen bestimmt. Erst wenn beim NR-Verfahren keine Schrittweite gefunden werden kann, mit der sich bei der gegenwärtigen Richtung die Likelihood noch vergrößert, muss eine neue Änderungsrichtung berechnet werden. D. h. der Evaluator muss so organisiert sein, dass beim Ausprobieren neuer Schrittweiten nur die Likelihoodwerte zurückgegeben werden, und keine Berechnungszeit für die Ableitungen verschwendet wird. Erst wenn eine neue Richtung berechnet werden muss, sollen die Ableitungen bestimmt werden. Um diese Teile zu trennen, gibt `moptimize()` an den Evaluator die `todo`-Variable weiter. Wenn der Skalar `todo`$= 0$ ist, wird nur eine neue Likelihoodauswertung erwartet. Wenn der `todo`$=1$, wird zusätzlich die erste Ableitung erwartet, und wenn `todo`$= 2$, wird zusätzlich die zweite Ableitung erwartet.[6]

Der Regressionskoeffizientenvektor `b`, also β_r, ist der zentrale Baustein, aus dem die aktuelle logarithmierte Likelihood $\ell_i(\beta)$ und die zugehörigen ersten beiden Ableitungen berechnet werden. Nach der Berechnung werden diese in die Variablen `lnfj`, `S` und `H` geschrieben und an `moptimize()` zurückgegeben.

Im zweiten Teil der Übergabe der Informationen aus Stata in den Evaluator werden im Abschnitt nach Zeile 12 eine Analysesampleindikator `touse`, die Panelindikatorvariable `id` und die Matrix der unabhängigen Variablen X an den Evaluator gegeben. Das oben dargestellte Minimalbeispiel des Aufrufs der `moptimize()`-Funktion verzichtet auf den Analysesampleindikator, da hier implizit davon ausgegangen wird, dass alle Beobachtungen das Analysesample ergeben. Die Matrix der unabhängigen Variablen enthält, wie im Minimalbeispiel die Variablen `X1`,...,`XM`.[7] In Zeile 18 wird schließlich die Matrix `out2eq` aus Stata an den Evaluator übergeben. Zum Verständnis dieser Matrix betrachten wir folgendes Beispiel: Angenommen, die abhängige Variable y habe die Ausprägungen $-1, 0, 1, 3, 5$ und die Referenzkategorie sei $B = 1$. Für dieses Beispiel ist die Matrix `out2eq` folgendermaßen aufgebaut:

$$\texttt{out2eq} = \begin{pmatrix} -1 & 1 \\ 0 & 2 \\ 1 & 0 \\ 3 & 3 \\ 5 & 4 \end{pmatrix}$$

Damit bildet die erste Spalte der Matrix die möglichen Ausprägungen der abhängigen Variable. Die zweite Spalte ist einerseits der Vektor über die Laufindezes der Alternativen, die nicht die Referenzkategorie bilden, und andererseits bezeichnet die Null in dieser

6 Für das NR-Verfahren ist `todo` entweder null oder zwei. Für die alternativen Iterations-Verfahren, die kurz in Abschnitt 4.1 erwähnt wurden, ist `todo` bei bestimmten Schritten auch eins.

7 Die unabhängigen Variablen werden in der gegenwärtigen Fassung doppelt übergeben. Die Dopplung ist teilweise unnötig und ineffizient, erleichtert aber die Ausgabe der Ergebnistabelle.

Spalte die Referenzkategorie. Diese Matrix wird benötigt, wenn man beliebige Alternativenanzahlen und beliebige, dynamisch wechselnde Referenzkategorien verwenden will. Da wir im Minimalbeispiel von einer bekannten Alternativenanzahl J mit vorher bekannter, statischer Referenzkategorie $B = 1$ ausgegangen sind, wurde diese Matrix ignoriert.

Fortsetzung

```
19
20      // derived information
21      J=rows(out2eq)
22      M=cols(X)
23      panelinfo=panelsetup(id,1)
24      N=panelstats(panelinfo)[1]
25
26      // init lnli, gi, Hi
27      lnli=J(N,1,0)
28      if (todo>0) {
29         gi=J(N,(J-1)*M,0)
30         if (todo==2) {
31            Hi=J(N,((J-1)*M)^2,0) // panelwise Hessian component in wide-format
32         }
33      }
34
```

Im zweiten Schritt werden im Abschnitt ab Zeile 20 aus den übergebenen Informationen die Konstanten N, J und M bestimmt, die die Lesbarkeit der folgenden Berechnungsteile und den Zusammenhang zu den oben dargestellten Ausdrücken für die Ableitungen erleichtern soll. Weiterhin werden im Abschnitt ab Zeile 26 die Variablen `lnli`, `gi` und `Hi` initialisiert, d. h. auf null gesetzt. In den Spaltenvektor `lnli` über alle Beobachtungseinheiten soll die logarithmierte Likelihood $\ell_i(\beta_r)$ geschrieben werden. In die Matrix `gi` über alle Beobachtungseinheiten und alle Regressionskoeffizienten sollen die ersten Ableitungen $\mathrm{s}_i(\beta_r)$ geschrieben werden. In die Matrix `Hi` über alle Beobachtungseinheiten und jedes Paar von Regressionskoeffizienten wird die zweite Ableitung der Likelihoodfunktion für jede Beobachtungseinheit $\mathrm{H}_i(\beta_r)$ geschrieben. Man beachte, dass die Matrix für jede Beobachtungseinheit eine Zeile mit $(J-1)\cdot M$ Spalteneinträgen enthält, d. h. die quadratische Matrix $\mathrm{H}_i(\beta_r)$ für jede Beobachtung im wide-Format vorliegt.

Im folgenden Teil wird nun die logarithmierte Likelihood und die Ableitungen unter Bezugnahme auf die oben definierten Hilfsvariablen berechnet.

Fortsetzung

```
35    // calculate lnli, gi, Hi
36    for(i=1;i<=N;i++) { // loop over panels
37       // create panel-wise variables (only one call per panel)
38       yi=moptimize_util_depvar(ML,1)[|panelinfo[i,1]\panelinfo[i,2]|]
39       Xi=X[|panelinfo[i,1],.\panelinfo[i,2],.|]
40
41       // init major auxiliary variables (A,B,C,D,E)
42       A=0
43       B=0
44       if (todo>0) {
45          C=J(1,(J-1)*M,0)
46          D=J(1,(J-1)*M,0)
47          if (todo==2) {
48             E=J((J-1)*M,(J-1)*M,0)
49          }
50       }
51
52       // calculate A,C
53       for(j=1;j<=J;j++) { // loop over outcomes
54          if (out2eq[j,2]!=0) { // exclude base outcome
55             A=A+quadcolsum((yi:==out2eq[j,1]):* /*
56             */ ((Xi*(colshape(b,M)'))[.,out2eq[j,2]]))
57             if (todo>0) {
58                for(m=1;m<=M;m++) { // loop over indep. vars
59                   C[1,(out2eq[j,2]-1)*M+m]=quadcolsum((yi:==out2eq[j,1]):* /*
60                   */ (Xi[.,m]))
61                }
62             }
63          }
64       }
65
```

Die Berechnung der logarithmierten Likelihoodfunktion $\ln \ell_i(\beta)$ und der ersten Ableitung $s_i(\beta)$ für jede Beobachtungseinheit und der zweiten Ableitung für die Stichprobe $H(\beta)$ verläuft als Schleife wiederholt für jede einzelne Beobachtungseinheit. Der Teil, der als Schleife wiederholt wird, beginnt in Zeile 36 oben und endet in Zeile 111 unten. Zu Beginn werden in Zeile 37 dann der Vektor der abhängigen Variable y_i und die Matrix der unabhängigen Variablen X_i für die betrachtete Beobachtungseinheit in die Variablen `yi` und `Xi` geschrieben.

Zur Berechnung von $\ln \ell_i(\beta)$, $s_i(\beta)$ und $H(\beta)$ werden die oben definierten Hilfsvariablen A, B, C, D und E berechnet. Dafür werden die entsprechenden Variablen im Teil ab Zeile 41 zunächst auf null gesetzt.

Im Teil von Zeile 52 bis Zeile 64 werden die Hilfsvariablen A und C für die Beobachtungseinheit i berechnet.

Fortsetzung

```
66       // calculate B,D,E
67       // generate Delta_i=Set of permutations of y_i
68       permuteinfo=cvpermutesetup(yi)
69       // loop over permutations of y_i (d_i in Delta_i)
70       while((d=cvpermute(permuteinfo))!=J(0,1,.)) {
71          // init minor auxiliary variables
72          Z1=0
73          if (todo>0) {
74             Z2=J(1,(J-1)*M,0)
75          }
76
```

```
77          // calculate Z1,Z2
78          for(j=1;j<=J;j++) { // loop over outcomes
79             if (out2eq[j,2]!=0) { // exclude base outcome
80                Z1=Z1+quadcolsum((d:==out2eq[j,1]):*((Xi* /*
81                */ (colshape(b,M)'))[.,out2eq[j,2]]))
82                if (todo>0) {
83                   for(m=1;m<=M;m++) {
84                      Z2[1,(out2eq[j,2]-1)*M+m]= /*
85                      */ quadcolsum((d:==out2eq[j,1]):*(Xi[.,m]))
86                   }
87                }
88             }
89          }
90          Z1=exp(Z1)
91
92          // fill up B,D,E with minor aux. var's
93          B=B+Z1
94          if (todo>0) {
95             D=D+Z2:*Z1
96             if (todo==2) {
97                E=E+(quadcross(Z2,Z2)):*Z1
98             }
99          }
100      }
```

Danach werden die Hilfsvariablen B, D und E berechnet. Diese setzen sich technisch betrachtet aus den Hilfsvariablen Z1 und Z2 zusammen. Inhaltlich beziehen sich die Hilfsvariablen B, D und E auf die Menge der Permutation Δ_i für jede Beobachtungseinheit i. Die Hilfsvariablen A und B oben beziehen sich dagegen auf die Zeitreihe der tatsächlich gewählten Alternativen y_i. Für die Berechnung von B, D und E bedeutet das nun zweierlei. Erstens werden – von der inhaltlichen Seite betrachtet – zunächst die Bestandteile Z1 und Z2 berechnet, und aus diesen die Hilfsvariablen B, D und E zusammengesetzt. Zweitens läuft – von der technischen Seite betrachtet – *innerhalb* eines Schleifengliedes für eine bestimmte Beobachtungseinheit i der oben angesprochenen Schleife über alle Beobachtungseinheiten (Zeile 36–111) eine *zweite* Schleife über die Menge aller Permutation Δ_i der von Beobachtungseinheit i realisierten Zeitreihe der gewählten Alternativen y_i ab.

Die Menge der Permutationen von y_i wird in Zeile 68 mit der Mata-Funktion `cvpermutesetup` gebildet.[8] Die Schleife über die Δ_i wird in Zeile 70 und wird in Zeile 100 geschlossen.

Innerhalb dieser Schleife über die Permutationen werden die Hilfsvariablen B, D und E berechnet. Zuerst werden hier im Abschnitt ab Zeile 71 die Bestandteile Z1 und Z2 null gesetzt. Im Abschnitt nach Zeile 77 werden dann Z1 und Z2 berechnet. Diese werden im Prinzip wie die Hilfsvariablen A und C berechnet, jedoch statt für die realisierte Zeitreihe y_i für die jeweilige Permutation d_i, über die die innere Schleife gerade läuft. Im Teil ab Zeile 92 werden dann die Hilfsvariablen B, D und E aus den Z1 und Z2 zusammengesetzt.

8 Die Mata-Funktion `cvpermutesetup`, die die direkte Berechnung der Likelihoodfunktion erlaubt, ist der wesentliche Grund, warum der Evaluator in Mata implementiert ist.

Fortsetzung

```
101
102         // fill up lnli,gi,Hi with major aux. var's A,B,C,D,E
103         lnli[i]=A-ln(B)
104         if (todo>0) {
105            gi[i,.]=C-D:/B
106            if (todo==2) {
107               // fill up Hi in wide format
108               Hi[i,.]=rowshape(((quadcross(D,D)):/(B^2))-(E:/B),1)
109            }
110         }
111      }
112
113      // fill up lnf,h,H with panel-wise lnli,gi,Hi
114      lnfj=lnli
115      if (todo>0) {
116         S=gi
117         if (todo==2) {
118            H=colshape(quadcolsum(Hi),((J-1)*M))
119         }
120      }
121   }
```

Ende

Im letzten Teil des Evaluators werden nach der Berechnung der Hilfsvariablen A, B, C, D und E die logarithmierte Likelihoodfunktion $\ln \ell_i(\beta)$ für die Beobachtungseinheit i und die beiden Ableitungen $\mathrm{s}_i(\beta)$ und $\mathrm{H}_i(\beta)$ berechnet. Dies geschieht im Abschnitt ab Zeile 102. Im letzten Abschnitt ab Zeile 113 werden dann die Likelihoodfunktion und die score-Funktion in die entsprechenden Variablen `lnfj` und `S` geschrieben, die an die `moptimize()`-Funktion zurückgegeben werden. Außerdem wird hier die Summe der zweiten Ableitung der logarithmieren Likelihoodfunktion $\mathrm{H}_i(\beta)$ über die Stichprobe gebildet, und in die Variable `H` geschrieben, die an die `moptimize()`-Funktion zurückgegeben wird.

Eigenständiger Stata-Befehl `femlogit`

Wir haben nun oben minimalen Aufruf der `moptimize()`-Funktion dargestellt und das vollständige Evaluator-Unterprogramm `femlogit_eval_gf2()`, dass von `moptimize()` aufgerufen wird. Hier sollen nun die Erweiterungen des minimalen Aufrufs ausgeführt werden, die es ermöglichen, das femlogit-Modell mit Stata wie mit einem gewöhnlichen Statabefehl zu schätzen. Aus Platzgründen wird hier nicht vollständig auf jedes Detail des Stataskripts eingegangen. Sowohl der Stata-Code als auch der Mata-Code sind in ungekürzter Fassung im Anhang A zu finden.

Um aus dem minimalen Aufruf der `moptimize()`-Funktion ein eigenständiges Kommando zu machen, müssen verschiedene Dinge gewährleistet sein.[9] Zunächst muss das Programm, dass in Stata „ado" heißt, eine allgemeine Eingabe verarbeiten. D. h. eine beliebige bzw. gerade noch gültige Angabe der abhängigen und der unabhängigen Variablen, des Panelgruppen-Indikators und der linearen Beschränkungen muss so verarbeitet werden, dass das richtige Modell geschätzt wird und die richtigen Ergebnisse ausgegeben werden. Die Eingabe wird mit dem Kommando `syntax` verarbeitet.

9 Siehe hierfür Baum (2009), StataCorp LP (2009c) und Cameron und Trivedi (2010).

Weiterhin muss das Programm in eindeutiger und korrekter Weise das Analysesample bilden. Hierzu werden mit dem Kommando `marksample` die fehlenden Angaben „listwise” gelöscht. Weiterhin werden mit dem Kommando `_rmcoll` kollineare unabhängige Variablen aus der Analyse entfernt, was dazu führen kann, dass Beobachtungen verloren gehen. Ferner werden zur Bestimmung des Analysesamples, im Code an separater Stelle mit Routinen, die aus der Implementierung des `clogit`-Kommandos übernommen wurde, solche Panelgruppen entfernt, bei denen die abhängige Variable über die Messzeitpunkte nicht variiert. Schließlich werden bei dieser Stelle auch unabhängige Variablen aus der Analyse genommen, bei denen bei keiner Beobachtungseinheit keine Varianz über die Messzeitpunkte vorliegt.

In Anlehnung an die vorhandene Umsetzung des `mlogit`-Kommandos wird geprüft, ob genug Arbeitsspeicher freigegeben wurde. Weiterhin wird die vom Evaluator erwartete Indikatormatrix `out2eq` definiert, die eine schnelle und einfache Zuordnung von allen Alternativen zu den Alternativen ohne die Referenzkategorie erlaubt. Ferner wird für den LR-Test, der bei Modellen ohne lineare Restriktion ausgegeben wird, die logarithmierte Likelihood für die Stichprobe *ohne* Kovariate berechnet. Die Berechnung folgt hier der Umsetzung des gegebenen `clogit`-Kommandos. Weiterhin werden die Startwerte wie beim `clogit`-Kommando gebildet, d. h. als Startwerte werden die Schätzer der gepoolten multinomialen logistischen Regression verwendet.[10]

Ferner wird geprüft, ob die linearen Restriktionen so spezifiziert wurden, dass sie verarbeitet werden können. Der größte Teil der Verarbeitung der linearen Restriktion verläuft automatisch in der `moptimize()`-Funktion. Neben der Eingabenprüfung muss aber auch zusätzlich die Anzahl der Freiheitsgrade korrigiert werden.

Nach dieser Vorverarbeitung wird dann eine erweiterte Fassung des minimalen `moptimize()`-Aufrufs ausgeführt. Zuerst werden, um eine schnelle und stabile Berechnung im Evaluator zu gewährleisten, die Daten umsortiert. Die Originalsortierung der Daten wird am Ende wieder hergestellt. Dann werden alle angegebenen Informationen über die abhängige, die unabhängigen Variablen und den Panelgruppenindikator in das Optimierungsproblem übergeben. Hier wird auch die Analysestichprobe und die linearen Restriktionen übergeben. Mit diesen Informationen und dem oben erläuterten Evaluator wird das iterative ML-Verfahren ausgeführt. Danach werden die Ergebnisse ausgegeben. Hier wird im Wesentlichen die Regressionskoeffiziententabelle ausgegeben. Die Tabelle stellt zusätzlich den LR-Test[11] inklusive des p-Wertes und das Modellfitmaß Pseudo-R^2 von McFadden (1974) dar. Zuletzt werden die notwendigen Informationen in Systemvariablen geschrieben, so dass der Befehl an andere einfache Auswertungskommandos anschlussfähig ist.

Vorausgesetzte Datenstruktur und Syntax

Abschließend soll noch kurz erläutert werden, wie der hier dargestellte Befehl `femlogit` konkret in Stata angewendet wird. Der Befehl erwartet, dass die Daten im

10 Man beachte, dass die gepoolte multinomiale logistische Regression, die für die Startwerte gerechnet wird, im Gegensatz zum femlogit-Modell eine Konstante für jeden Kontrast enthält.

11 Bei linearen Restriktionen wird der LR-Test durch den Wald-Test ersetzt.

„long"-„wide"-Format vorliegen. Gemeint damit ist, dass jede Beobachtung bzw. jeder Fall ein Messzeitpunkt für eine Beobachtungseinheit ist. Zu jedem Messzeitpunkt entscheidet die Beobachtungseinheit über mehrere Alternativen, wobei die Entscheidung über die Alternative als eine Variable pro Messzeitpunkt abgebildet ist. D. h. in der Zeitdimension liegen die Daten im „long"-Format vor und in der Alternativendimension im „wide"-Format. Das Datenformat, wie es der Befehl erwartet, ist in Tabelle 4.1 dargestellt.

Tabelle 4.1: Vorausgesetzte Datenstruktur

id	t	y	x
1	1	3	2
1	2	1	1
1	3	2	3
2	1	2	1
2	2	3	1
2	3	1	2

In der Tabelle sind zwei Beobachtungseinheiten mit `id` $= 1$ und `id` $= 2$. Für beide Einheiten liegen jeweils drei Messzeitpunkte vor. Die Einheiten entscheiden über drei Alternativen. Die Beobachtungseinheit mit `id` $= 1$ wählt zum Zeitpunkt $t = 1$ die Alternative mit der Ausprägung 3, zum Zeitpunkt $t = 2$ die Alternative 1 und zum Zeitpunkt $t = 3$ die Alternative 2. Die Ausprägungen der einzigen unabhängigen Variablen x sind für die Beobachtungseinheit zum Zeitpunkt $t = 1$ 2, zum Zeitpunkt $t = 2$ 1 und zum Zeitpunkt $t = 3$ 3. Für die zweite Beobachtungseinheit gilt das Gleiche analog.

Die Syntax des Befehls `femlogit` ist folgendermaßen definiert:

```
femlogit depvar [indepvars] [if] [in], group(varlist) /*
*/ [baseoutcome(#) constraints(clist) difficult]
```

D. h. es muss eine abhängige Variable angegeben werden. Es können, müssen aber keine unabhängigen Variablen spezifiziert werden. Die Analysestichprobe lässt sich wie gewohnt mit den `if`- und `in`-Befehlen eingrenzen. Die Panelindikatorvariable ist ein Pflichtelement wie die abhängige Variable. Die Referenzkategorie kann, muss aber nicht, als die entsprechende Ausprägung angegeben werden. Die linearen Restriktionen werden, der Stata-Konvention folgend, vorher definiert, und können hier optional zugeordnet werden (vgl. hierzu StataCorp LP 2009d, S. 316). Schließlich kann mit dem optionalen Befehl `difficult` angegeben werden, dass statt dem in Abschnitt 4.1 dargestellten konventionellen NR-Verfahren ein „Hybrid"-Verfahren verwendet wird.[12] Die Unterstreichungen zeigen an, dass manche Optionen abgekürzt angegeben werden können.

12 Die „difficult"-Option ersetzt das Standardverfahren, dass technisch auch als „modified marquart"-Verfahren bezeichnet wird, durch das „hybrid"-Verfahren. Vgl. hierzu Gould et al. (2010, S. 15–17).

4.3 Performanztest

In diesem Abschnitt soll die oben dargestellte Implementation des femlogit-Modells getestet werden. Der Test besteht aus zwei Teilen. Im ersten Teil wird die Implementation des femlogit-Modells mit dem bereits vorliegenden Statabefehl `clogit` verglichen, das für zwei Alternativen die gleichen Ergebnisse produzieren sollte. Im zweiten Teil wird ein simulierter Datensatz mit mehr als zwei Alternativen mit bekannten Regressionskoeffizienten erzeugt, an dem die Implementation getestet wird.

4.3.1 Vergleich mit `clogit`

Für den Spezialfall der logistischen Regression mit fixed effects mit nur zwei Alternativen ist bereits das Kommando `clogit` implementiert.[13] Daher kann dieser Befehl als Vergleichsstandard für die Implementation des femlogit-Modells herangezogen werden. Zumindest für zwei Alternativen sollte die oben dargestellte Implementation die gleichen Ergebnisse liefern.

Zur allgemeinen Vergleichbarkeit und Nachvollziehbarkeit verwenden wir für den Test einen allgemein verfügbaren Datensatz. Stata verwendet als Beispieldatensatz für den `clogit`-Befehl eine Teilstichprobe aus dem „NLS of Young Women” (vgl. StataCorp LP 2009d, S. 274, StataCorp LP 2009b, S. 6), den Stata öffentlich zur Verfügung stellt.[14] Die Teilstichprobe „union.dta” ist ein Datensatz über 4 434 Frauen, die im Jahr 1968 zwischen 14 und 24 Jahre alt waren. Der Messzeitraum ist von 1970 über 1988.

In unserem einfachen Testbeispiel verwenden wir aus diesem Datensatz die Gewerkschaftszugehörigkeit als dichotome abhängige Variable (`union`) und das Alter in Jahren (`age`) und die Schulbildung in Bildungsjahren (`grade`) als unabhängige Variablen. Wir betrachten also das folgende Analysemodell:

$$\Pr(\texttt{union}_{it} = 1) = \frac{\exp(\alpha_i + \beta_1 \texttt{age}_{it} + \beta_2 \texttt{grade}_{it})}{1 + \exp(\alpha_i + \beta_1 \texttt{age}_{it} + \beta_2 \texttt{grade}_{it})}. \tag{4.9}$$

Das betrachtete Modell hat keine inhaltliche Motivation. Da wir nur die Performanz der Implementation des femlogit-Modells schätzen ist das aber auch kein Problem. Von den 4 434 Frauen weisen 2 744 keine Varianz auf der abhängigen Variablen auf, d. h. die Analysestichprobe beschränkt sich auf 1 690 Frauen mit insgesamt 12 035 Messzeitpunkten.

Betrachten wir zuerst den „Nennwert” der Implementation, indem wir die expliziten Ergebnisausgaben des `clogit`- und des `femlogit`-Befehls miteinander vergleichen. Der `clogit`-Befehl erzeugt diese Ausgabe:

13 Das felogit-Modell ist in Stata mit dem Kommando `clogit` und mit dem nahezu äquivalenten Kommando `xtlogit, fe` implementiert (vgl. StataCorp LP (2009d)).

14 Der Datensatz „`union.dta`” kann auf der Verlagsseite heruntergeladen werden: `http://www.stata-press.com/data/r11/union.dta`.

clogit

```
. clogit union age grade, group(idcode)
note: multiple positive outcomes within groups encountered.
note: 2744 groups (14165 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -4536.0531
Iteration 1:   log likelihood = -4534.5985
Iteration 2:   log likelihood = -4534.5984

Conditional (fixed-effects) logistic regression   Number of obs   =      12035
                                                  LR chi2(2)      =      31.17
                                                  Prob > chi2     =     0.0000
Log likelihood = -4534.5984                       Pseudo R2       =     0.0034

------------------------------------------------------------------------------
       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0164108   .0041344     3.97   0.000     .0083075     .024514
       grade |   .0846214   .0416914     2.03   0.042     .0029078    .1663351
------------------------------------------------------------------------------
```

Die Ausgabe gibt wieder, was wir oben zu der Beschränkung der Analysestichprobe erläutert haben. Die Ausgabe des `femlogit`-Befehls ist diese:

femlogit

```
. femlogit union age grade, group(idcode)
note: 2744 groups (14165 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -4536.0531
Iteration 1:   log likelihood = -4534.5985
Iteration 2:   log likelihood = -4534.5984

Fixed-effects multinomial logistic regression     Number of obs   =      12035
                                                  LR chi2(2)      =      31.17
                                                  Prob > chi2     =     0.0000
Log likelihood = -4534.5984                       Pseudo R2       =     0.0034

------------------------------------------------------------------------------
       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_0           |  (base outcome)
-------------+----------------------------------------------------------------
_1           |
         age |   .0164108   .0041344     3.97   0.000     .0083075     .024514
       grade |   .0846214   .0416914     2.03   0.042     .0029078    .1663351
------------------------------------------------------------------------------
```

Es zeigt sich, dass die Ergebnisse der Implementation des femlogit-Modells sich selbst in den letzten Nachkommastellen nicht von den Ausgaben des `clogit`-Befehls unterscheiden. Die Ausgabe unterscheidet sich an zwei Stellen. Erstens wird beim `femlogit`-Befehl nicht darauf hingewiesen, dass bestimmte Ausprägungen mehrfach innerhalb einer Beobachtungseinheit vorkommen. Zweitens folgt die Darstellung der Regressionskoeffizienten der Ausgabe des implementierten `mlogit`-Befehls. Die Referenzkategorie erhält eine eigene Zeile und die Ausprägungen der abhängigen Variablen werden explizit dargestellt.

Über den ersten Eindruck hinaus gibt es aber Unterschiede zwischen den beiden Implementationen. Erstens ist die Berechnungszeit beim `clogit`-Befehl wesentlich kürzer als beim `femlogit`-Befehl. Auf einem gewöhnlichen Arbeitsrechner[15] dauert es ca. 2 Sekunden, das dargestellte Modell mit dem `clogit`-Befehl zu schätzen. Mit dem `clogit`-Befehl dauert die Schätzung desselben Modells ca. 69 Sekunden. Ferner sind die geschätzten Regressionskoeffizienten nicht exakt gleich. Der Unterschied im Vektor der Regressionskoeffizienten ist bei diesem Beispiel $(2,082 \cdot 10^{-17}, -1,540 \cdot 10^{-15})$. Der Unterschied in der Varianzmatrix ist

$$\begin{pmatrix} 3.219 \cdot 10^{-19} & -1.702 \cdot 10^{-19} \\ -1.702 \cdot 10^{-19} & 6.536 \cdot 10^{-16} \end{pmatrix}.$$

Insgesamt zeigt der erste Test, dass die Implementation des femlogit-Modells die Ergebnisse des `clogit`-Befehls einerseits nicht exakt reproduzieren. Andererseits sind die Unterschiede bei dem hier dargestellten Beispiel für zwei Alternativen so gering, dass sie für praktische Anwendungen vernachlässigt werden können. Jedoch ist die Implementation des femlogit-Modells wesentlich langsamer als der bereitstehende `clogit`-Befehl. Im nächsten Abschnitt soll die Testung auf mehr als zwei Alternativen ausgedehnt werden.

4.3.2 Anwendung auf simulierte Daten

Für mehr als zwei Alternativen steht kein Vergleichsstandard zur Verfügung. Die einzige Möglichkeit eines Leistungstests besteht darin, einen simulierten Datensatz zu analysieren, bei dem die Koeffizienten vorher festgelegt werden. Zielsetzung der Simulation ist es, zu zeigen, dass die Implementation auch bei mehreren Alternativen konsistente Schätzer liefert.

Der simulierte Datensatz wird hierfür so konstruiert, dass er einer hinreichend realistischen Anwendung entspricht. Im Einzelnen heißt das, dass wir eine Stichprobe von 4 000 Beobachtungseinheiten ($N = 4000$) mit jeweils fünf Messzeitpunkten ($T = 5$) betrachten, die über fünf Alternativen ($J = 5$) entscheiden. Eine Einschränkung in Hinsicht auf den Realismus der Simulation wird bei der Anzahl der unabhängigen Variablen vorgenommen. Der Einfachheit halber beschränken wir uns auf nur eine unabhängige Variable ($M = 1$).

Diese Einschränkung ist vertretbar, da das Hauptinteresse bei fixed-effects-Modellen beim Zusammenhang von konstanten Variablen und einer interessierenden, zeitlich variierenden Variable liegt. Wie schon ausführlich diskutiert wurde, kann man mit fixed-effects-Modellen auf die Annahme verzichten, dass man alle relevanten antezedierenden zeitlich konstanten Variablen in das Modell aufgenommen hat, da man für diese implizit kontrollieren kann.

Im simulierten Datensatz konstruieren wir parallel zu unabhängigen Variablen x für jede Alternative zeitkonstante unbeobachteten Heterogenitäten α_j, wobei die Heterogenität für die Referenzkategorie $B = 1$ null gesetzt wird: $\alpha_1 = 0$. Die Heterogenitäten und die unabhängige Variable sind gemeinsam normalverteilt mit den Erwartungswer-

15 Der verwendete Rechner hat einen Intel Pentium D Prozessor 805 auf einem Asus 4CoreDual-SATA2 Motherboard mit 1 GB DDR3 RAM.

ten $\mathrm{E}(x) = \mathrm{E}(\alpha_j) = 0$ und den Varianzen $\mathrm{Var}(x) = \mathrm{Var}(\alpha_i) = 1$. Der entscheidende Punkt ist die Korrelation zwischen der unabhängigen Variablen und der Heterogenitäten: $\mathrm{r}_{x,\alpha_j} = \frac{1}{5}$.[16] Die Heterogenitäten α_j selbst sind untereinander unabhängig. Außerdem ist die unabhängige Variable über die Messzeitpunkte hinweg unabhängig. Aus diesen Restriktionen ergibt sich folgende Korrelationsmatrix über die unbeobachteten Heterogenitäten α_2,...,α_J und den Messzeitpunkten der unabhängigen Variablen x_1,...,x_T:

	α_2	α_3	α_4	α_5	x_1	x_2	x_3	x_4	x_5
α_2	1	0	0	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
α_3	0	1	0	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
α_4	0	0	1	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
α_5	0	0	0	1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
x_1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	1	0	0	0	0
x_2	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0	1	0	0	0
x_3	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0	0	1	0	0
x_4	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0	0	0	1	0
x_5	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0	0	0	0	1

Weiterhin wird ein idiosynkratischer Fehlerterm ϵ als Zufallsvariable erzeugt. Für jede Beobachtungseinheit wird für jede Kombination der zehn Messzeitpunkte und der fünf Alternativen 50 Fehlerterme ϵ_{itj} erzeugt. Die Fehlerterme werden als Zufallsvariablen so gebildet, dass sie identisch und unabhängig Gumbel-verteilt sind mit dem Erwartungswert $\mathrm{E}(\epsilon_{itj}) = \gamma$ und der Varianz $\mathrm{Var}(\epsilon_{itj}) = \frac{\pi^2}{6}$.[17]

Weiterhin werden für den simulierten Datensatz die Regressionskoeffizienten a priori angenommen. Für die fünf Alternativen und der einen unabhängigen Variable x ergeben sich fünf Regressionskoeffizienten β_1, β_2, ..., β_5. Der erste Regressionskoeffizient, der zur Referenzkategorie gehört, wird wie die Heterogenität null gesetzt: $\beta_1 = 0$. Die übrigen Regressionskoeffizienten werden auf die folgenden Werte gesetzt: $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (0, 1, 2, 3, 4)$.

Aus den Heterogenitäten, der unabhängigen Variablen, dem Fehlerterm und den Regressionskoeffizienten lässt sich nun die latente Variable y^*_{itj} für jede einzelne Alternative bilden. Für jede Alternative j gilt:

$$\begin{aligned} &\text{Für } j = 1: \quad y^*_{it1} = \epsilon_{it1}, \\ &\text{für } j \neq 1: \quad y^*_{itj} = \alpha_{ij} + \beta_j x_{it} + \epsilon_{itj}. \end{aligned}$$

16 Die Korrelation zwischen den Heterogenitäten und der unabhängigen Variablen wird auf diesen scheinbar niedrigen Wert gesetzt, da die Korrelation durch die anderen Restriktionen beschränkt ist. Da die Heterogenitäten untereinander und die unabhängige Variable über die Messzeitpunkte hinweg unabhängig sind, ist die Korrelation beschränkt, da die zugehörige Korrelationsmatrix positiv semidefinit sein muss, um eine sinnvolle gemeinsame Verteilung zu definieren.

17 Eine Gumbel-verteilte Zufallsvariable wird technisch erzeugt, indem man in die Inverse der Verteilungsfunktion gleichverteilte Zufallsvariablen einsetzt (vgl. z. B. Cameron und Trivedi 2010, S. 132f.). Für die Standard-Gumbel-Verteilung ist die Verteilungsfunktion $F_\epsilon(x) = \exp(-\exp(-x))$ (vgl. Johnson et al. 1994, Kap. 22). Damit ist die Inverse der Verteilungsfunktion $F_\epsilon^{-1}(x) = -\ln(-\ln(x))$.

Daraus wird dann im simulierten Datensatz die abhängige Variable y_{it} so gebildet, dass die abhängige Variable für jede Beobachtungseinheit zu jedem Messzeitpunkt diejenige Alternative ist, bei der in diesem Messzeitpunkt die latente Variable maximal ist:

$$y_{it} = j \Leftrightarrow \max_{k \in \{1,\ldots,J\}} y^*_{itk} = y^*_{itj}.$$

Nachdem wir nun einen Datensatz nach diesen Vorgaben erzeugt haben, können wir die Implementation des femlogit-Modells darauf anwenden. Um den Effekt der Verzerrung durch die nicht berücksichtigte unbeobachtete Heterogenität zu veranschaulichen, stellen wir dem fixed-effects-Modell die Schätzer des gepoolten Querschnittsmodell mit panelrobusten Standardfehlern gegenüber. In Abbildung 4.4 sind die Schätzer der Regressionskoeffizienten und die zugehörigen Konfidenzintervalle für die beiden Modelle dargestellt.

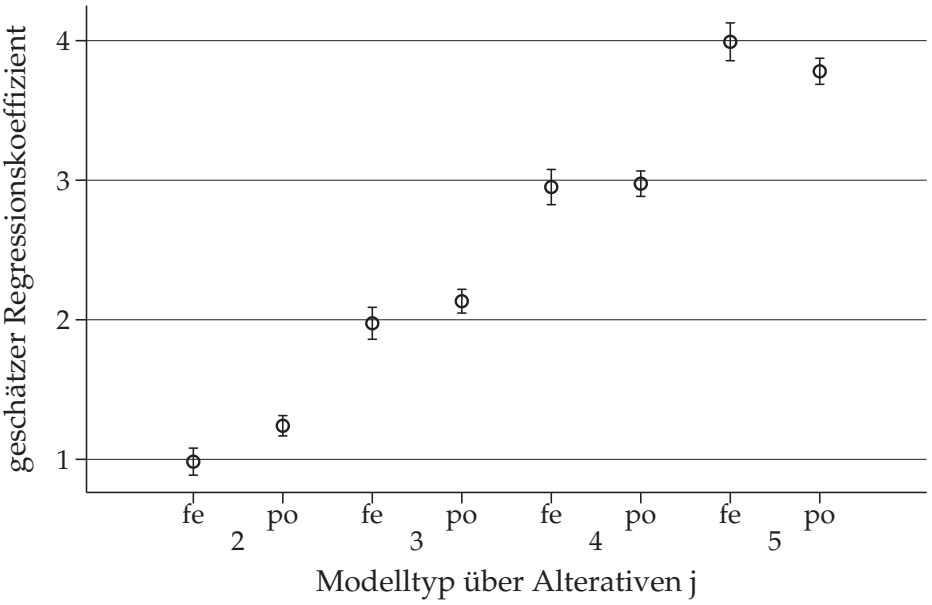


Abbildung 4.4: Geschätzte Regressionsmodelle nach Modelltyp

Die Abbildung zeigt, dass das fixed-effects-Modell die a priori gesetzten Regressionskoeffizienten konsistent schätzt. Dagegen sind die Schätzer des gepoolten Modells inkonsistent. Auch in diesem Beispiel zeigt sich wieder die deutlich längere Laufzeit des fixed-effects-Modells. Das gepoolte Modell wird in 1,3 Sekunden geschätzt, während das fixed-effects-Modell 100,2 Sekunden dauert.[18]

18 Stichprobenartige Untersuchungen zeigen, dass die Rechenzeit linear mit der Anzahl der Beobachtungseinheiten und der Anzahl der Permutationen der Zeitreihe der abhängigen Variablen wächst. D. h. dass die Rechenzeit in der Regel *faktoriell* mit der Anzahl der Messzeitpunkte wächst.

Es lässt sich zusammenfassen, dass die Implementation für eine beliebige Anzahl an Alternativen und Messzeitpunkten die gewünschten konsistenten Schätzer berechnet. Die Rechenzeit ist bei nur zwei Alternativen länger als beim äquivalenten `clogit`-Befehl. Außerdem steigt die Rechenzeit linear mit der Anzahl der Beobachtungseinheit und in Allgemeinen faktoriell mit der Anzahl der Alternativen und Anzahl der Messzeitpunkte. Grundsätzlich kann die Verwendung bei realistischen Anwendungen mit sehr langen Zeitreihen und vielen Alternativen eine technische Grenze der Berechenbarkeit erreichen. Bei den durchschnittlichen Anwendungen sollte man aber höchstens mit einer sehr langen Rechenzeit rechnen, nicht aber damit, dass der Schätzer nicht berechnet werden kann.

5 Zusammenfassung

In diesem Teil der Arbeit wurde in zwei Abschnitten zunächst der statistische Hintergrund des femlogit-Modells und danach die Implementation des Modells in Stata dargestellt. Im ersten Abschnitt wurde zunächst eine allgemeine Notation eingeführt. Danach wurde in drei Unterabschnitten zunächst der statistische Hintergrund für die Querschnittsmodelle, dann der Hintergrund für die fixed-effects-Modelle erläutert und im dritten Schritt verschiedene Erweiterungen dieser Modelle ergänzend diskutiert. Am Ende des ersten Abschnitts dieses Teils wurden die Probleme der fixed-effects-Modelle diskutiert. Der zweite Abschnitt befasste sich mit der eigentlichen Umsetzung des femlogit-Modells in Stata. Hierfür wurde zuerst die Theorie der Maximum-Likelihood-Schätzung veranschaulicht. Dann wurde die Umsetzung in Stata schrittweise vom Abstrakten zum Konkreten vorgestellt und erläutert. Die Darstellung der Implementation endete mit einem Performanztest, der zeigte, dass die Umsetzung erwartungsgemäß konsistente Schätzer lieferte.

Zur Implementation sind folgende Dinge zu beachten. Wie im Fazit des Abschnitts 3.3 erläutert wurde, ist das fixed-effects-Modell nicht uneingeschränkt anwendbar. Neben den üblichen Problemen der fixed-effects-Modelle besteht das besondere Problem des femlogit-Modell in der sehr eingeschränkten Interpretierbarkeit. Man erkauft die Möglichkeit, auf *jegliche* Annahmen über die gemeinsame Verteilung der unabhängigen Variablen und der unbeobachteten Heterogenität verzichten zu können, damit, dass man keine Effekte auf die Realisationswahrscheinlichkeiten interpretieren und statt dessen nur Odds-Ratio-Effekte interpretieren kann.

Weiterhin wurde beim Abschnitt 4.3 über den Performanztest kurz erwähnt, dass die Rechenzeit überexponentiell mit der Anzahl der Messzeitpunkte ansteigt. Das hat zur Folge, dass das Versprechen, dass die Implementation allgemein für alle denkbaren Fragestellungen angewendet werden, nicht voll eingehalten werden kann. Der zweite Teil der Arbeit widmet sich der Anwendung des femlogit-Modells auf realistische Fragestellungen. Hier wird deutlich werden, dass bei realen Anwendungen die Anzahl der Messzeitpunkte gekürzt werden muss, um die Rechenzeit in einem realisierbaren Rahmen zu halten.

Teil II

Anwendungen

6 Einleitung

Im zweiten Teil der Arbeit wird die Implementation des femlogit-Modells für zwei Replikationen angewendet. Die erste der beiden Anwendungen ist eine Replikation einer Analyse aus der Dissertation von Jette Schröder (2010), die zweite Anwendung repliziert ein Kapitel aus der Dissertation von Ulrich Kohler (2002). Die Anwendung, die sich auf die Arbeit von Schröder bezieht, widmet sich dem Effekt der Fertilität auf die Erwerbstätigkeit bei Frauen. Genauer soll der negative Effekt der Kinderzahl und des Alters des jüngsten Kindes auf die Erwerbstätigkeit in Vollzeit bzw. in Teilzeit möglichst unverzerrt geschätzt werden. Die zentrale Ursache der Verzerrung ist die antezedierende, zeitlich konstante, unbeobachtete „Neigung” zur Mutterschaft und zur beruflichen Karriere. Die zweite Anwendung, die sich auf die Arbeit von Kohler bezieht, untersucht den Einfluss der sozialen Klassenzugehörigkeit und der Zugehörigkeit zur anderen soziostrukturell definierten sozialen Gruppen auf die Parteipräferenz in Westdeutschland. Auch hier ist das Ziel, diese Effekte möglichst unverzerrt zu schätzen, wobei die Hauptursache für eine Verzerrung die fehlenden antezedierenden Variablen sind. Bei beiden Anwendungen können fixed-effects-Modelle, angewendet auf Längsschnittdaten, das Problem der unberücksichtigten, zeitlich konstanten, antezedierenden Variablen beheben.

Die Replikationen der Arbeiten von Kohler (2002) und Schröder (2010) beschränken sich auf die Teile der beiden Arbeiten, in denen die Autoren fixed-effects-Modelle verwenden. Die Autoren gehen in ihren Arbeiten weit über die in der vorliegenden Arbeit erläuterten Analysen hinaus. Auf die Darstellung bzw. Replikation dieser Analysen wird in dieser Arbeit verzichtet, da das femlogit-Modell keinen Mehrwert für diese Untersuchungen darstellt.

Beide replizierte Arbeiten verwenden bereits fixed-effects-Modelle, beschränken sich aber auf die bereitstehenden Implementation der binären logistischen Regression mit fixed-effects. Im Einzelnen untersucht Schröder jeweils getrennt den Effekt der Fertilität einerseits auf den Kontrast zwischen der Erwerbstätigkeit gegenüber der Nichterwerbstätigkeit und andererseits auf den Kontrast zwischen Erwerbstätigkeit in Vollzeit gegenüber Teilzeit. Kohler untersucht getrennt den Effekt der sozialen Gruppenzugehörigkeit auf die Präferenz für die SPD, die CDU und FDP und die Grünen, jeweils im Kontrast zu einer Präferenz für eine andere Partei oder für keine andere Partei. Das in dieser Arbeit im Mittelpunkt stehende femlogit-Modell schafft hier also insoweit Abhilfe, als dass die getrennten Analysen in einem Modell geschätzt werden können. Dadurch werden zwei Probleme der getrennten Analyse gelöst. Erstens beschränkt man sich bei den getrennten Analysen jeweils immer nur die Teilstichprobe der Beobachtungseinheiten, die beide Alternative des jeweils betrachteten dichotomen Kontrastes gewählt haben. Die Beschränkung führt dazu, dass die Schätzer im Allgemeinen weniger effizient sind als bei einem gemeinsamen Schätzung (Vgl. Long 1997, S. 149–151). Diese Ineffizienz kann man mit binären Modellen damit reduzieren, indem man Vergleichsgruppen zusammenwirft. Statt die Kontraste A vs. B, A vs. C und B vs. C getrennt zu betrachten, untersucht man z. B. den Kontrast A vs. (B oder C). Hieraus ergibt sich aber das zweite Problem. Bei dieser Analyse muss man annehmen, dass der Effekt einer unabhängigen Variablen X auf den Kontrast

A vs. B gleich dem Effekt von X auf den Kontrast A vs. C ist. Mit dem femlogit-Modell kann man sowohl immer die Gesamtstichprobe analysieren und getrennte, potentiell unterschiedliche Effekte für alle Alternativenpaare identifizieren.

Die Auswahl der beiden Arbeiten für die hier dargestellten Replikation erfolgte willkürlich, da kein Universum von möglichen replizierbaren Arbeiten definiert, und entsprechend auch keine kontrollierte Stichprobe aus einer definierten Menge von Arbeiten gezogen wurde. Die beiden Arbeiten erfüllen aber mehrere günstige Eigenschaften. Erstens können die Analysen in beiden Arbeiten zweifelsfrei von einer Verwendung des femlogit-Modells profitieren. Zweitens bearbeiten die Arbeiten aktuelle Forschungsfragen und decken mit der Familiensoziologie und Wahlsoziologie ein breites Themenfeld in der Sozialforschung ab. Drittens sind bei beiden Arbeiten die notwendigen Analysedaten und die Aufbereitungscodes hinreichend verfügbar. Insgesamt erscheint die Auswahl der beiden Arbeiten insoweit vertretbar, als dass mit den Replikationen hinreichend die Anwendbarkeit und die Relevanz des femlogit-Modells gezeigt wird.

An dieser Stelle ist anzumerken, dass in diesem Teil keine der in Abschnitt 3.2.3 diskutierten Erweiterungen umgesetzt wird. Eine besondere Berücksichtigung von selektivem Attrition wurde nicht implementiert. Eine Verletzung der Annahme der seriellen Unabhängigkeit der Fehlerterme über die Messzeitpunkte hinweg kann nach Cameron und Trivedi (2009) mit dem Panelbootstrap-Verfahren behandelt werden. Da bei den hier dargestellten Anwendung die Rechenzeit bei den fixed-effects-Modellen ohnehin schon sehr lang ist, wurde hierauf verzichtet. Weiterhin wurde bei keiner der beiden Anwendungen alternativen-spezifische unabhängige Variablen verwendet, so dass die Implementation der Berücksichtigung der linearen Beschränkungen hier nicht getestet werden kann.

Im nächsten Kapitel wird die Replikation der Arbeit von Schröder (2010) dargestellt. Es wird einer kurzer Überblick über den Stand der Forschung gegeben. Danach werden die Befunde von Schröder kurz zusammengefasst, und darauf die Replikation mit der um das femlogit-Modell erweiterte Analyse dargestellt. Die Replikation von Schröder schließt mit der vergleichenden Diskussion der Ergebnisse. Im dritten Kapitel wird die Replikation von Kohler (2002) dargestellt. Auch wird zuerst der Stand der Forschung überblicksartig kurz dargestellt. Danach werden die Ergebnisse von Kohler zusammenfassend erläutert. Ferner wird die eigene Replikation, wiederum erweitert um die Analyse mit dem femlogit-Modell, dargestellt, und die Unterschiede zur Arbeit von Kohler diskutiert.

7 Einfluss der Fertilität auf die Erwerbstätigkeit der Frau

In diesem Kapitel wird die Replikation der Arbeit von Schröder (2010, S. 97–190) dargestellt. Zunächst wird hierfür die kurz auf die zugrundeliegende Theorie erläutert und die bestehenden empirischen Befunde ausgeführt. Danach werden die konkreten Ergebnisse von Schröder dargestellt. Im Anschluss daran wird die Replikation inklusive der Erweiterung um das femlogit-Modell dargestellt und die entsprechenden Ergebnisse den Ergebnissen vergleichend gegenüber gestellt. Abschließend wird der zusätzliche neue Beitrag des femlogit-Modells bei diesem Untersuchungsgegenstand diskutiert.

7.1 Theorie

Der betrachtete Teil der Arbeit von Schröder (2010) beschäftigt sich mit dem Einfluss von Kindern auf die Erwerbstätigkeit von Frauen. Schröder stützt sich bei ihrem theoretischen Modell auf das Modell der „Zeitallokationsentscheidung nach dem klassischen familienökonomischen Modell” (ebd., S. 100) von Becker (1994).

Bei diesem Ansatz wird der Haushalt als einheitlicher Akteur angenommen. Dem Haushalt steht die Zeit der beteiligten Haushaltsmitglieder als Ressource zur Verfügung, die er zur Produktion von Einkommen durch Erwerbstätigkeit und zur Produktion von Haushaltsgütern durch Arbeit im Haushalt verwenden kann. Daneben kann der Haushalt mit dem Einkommen aus der Erwerbsarbeit ebenso Haushaltsgüter kaufen. Der Haushalt konsumiert sowohl die produzierten Haushaltsgüter als auch die gekauften Marktgüter. Bei den Haushaltsgütern handelt es sich um abstrakte Dinge wie Kinder bzw. deren Erziehung, Prestige, Gesundheit, daneben aber auch konkrete Dinge wie Obdach und Lebensmitteln, sowie deren Instandhaltung bzw. Zubereitung. Die Produktionsarbeit der Güter kann durch den Erwerb der Güter auf dem Markt substituiert werden.

Für den Haushalt als Wirtschaftseinheit besteht nun ein Allokationsproblem unter Restriktion der verfügbaren Ressource Zeit. Hierfür wird erstens eine gemeinsame Nutzenfunktion des Haushaltes über die konsumierten Haushaltsgüter und zweitens für die Haushaltsmitglieder unterschiedliche Produktivitäten beim Erzeugen von Erwerbseinkommen und Haushaltsgütern unter Einsatz von Zeit angenommen.[1] Der Haushalt maximiert also seinen Nutzen über mögliche Allokationen der verfügbaren Zeit über die Haushaltsmitglieder auf Erwerbs- und Hausarbeit.

Die zentrale Annahme von Becker, die Symmetrie der beiden Geschlechter bricht, ist, dass Frauen eine biologische Prädisposition für die Produktion von Haushaltsgütern haben, d. h. konkret für die Geburt und das Stillen von Kindern im Säuglingsalter. Aus diesem komparativen Produktivitätsvorteil der Frau hinsichtlich der Hausarbeit folgt nach Schröder für einen Haushalt, der aus einer Frau und einem Mann besteht, dass Frauen eher in hausarbeitsspezifisches Humankapital, und Männer eher erwerbsarbeitspezifisches Humankapital investieren. Frauen und Männer würden damit die biologisch angelegte Spe-

1 Schröder (2010, S. 103) weist auf die besondere Annahme hin, dass die Zeit von Mann und Frau Substitute sind, d. h. gemeinsam verbrachte Zeit genauso bewertet ist wie getrennt verbrachte Zeit. Beim manchen Haushaltsgütern wie Sexualität – aber auch andere Formen des Auslebens einer Partnerschaft – ist die Annahme sicher ungültig.

zialisierung der Frauen und der daraus implizierte relative Produktivitätsvorteil des Mannes bei der Erwerbsarbeit vollständiger ausschöpfen. Ein zusätzliches, bzw. alternatives Argument für diese heterogene Spezialisierung ist eine mögliche geschlechtsspezifische Diskriminierung auf dem Arbeitsmarkt.

Die zentralen Hypothesen bei Schröder beziehen sich auf den kausalen Effekt der Kinderzahl und des Kinderalters auf die Erwerbstätigkeit. Nach Schröder ist bei Becker der Effekt von Kindern auf die Arbeitsteilung nur am Rande erwähnt: „Becker geht davon aus, dass die Haushaltszeit von Kindern während der Jahre der Kindererziehung, wenn sie [die Frauen] sehr beschäftigt und produktiv sind, mehr wert ist." (ebd., S. 104). Darüber hinaus seien nach Schröder Skaleneffekte zu erwarten, d. h. die Produktivität bzgl. Haushaltsgütern bei Frauen bei zusätzlichen Kindern steigt. Z. B. ließen sich bei gleicher Zeit zwei Kinder beaufsichtigen. Auch steige die Zubereitungszeit von Mahlzeiten unterproportional mit der Anzahl der Esser. D. h. insgesamt, dass die haushaltgüterspezifische Produktivität unabhängig vom Geschlecht mit der Anzahl der Kinder steigen sollte. Schröder argumentiert für den Effekt der Kinderzahl weiter, dass der Wert der Haushaltsarbeit für Frauen mit der Anzahl der Kinder steige. D. h. Frauen hätten damit mit steigender Kinderzahl eine geringere Neigung zur Erwerbsarbeit.

Für den Alterseffekt argumentiert Schröder losgelöst von Becker, dass die Betreuungskosten mit steigendem Kinderalter zurückgehen würden. Mit wachsenden Alter seien Kinder selbständiger und würden zusätzlich in wachsendem Ausmaß durch außerhäusliche Institutionen wie Kindergarten oder Schule betreut. Dadurch steige mit zunehmendem Alter der Kinder die Neigung zur Erwerbsarbeit. Schröder argumentiert weiter, dass aber auch ein gegenteiliger Effekt des Kinderalters denkbar sei. Mit steigenden Kinderalter würde die Frau durch die lange Haushaltsarbeit Erfahrung ansammeln, so dass ein Anstieg an haushaltsspezifischer Produktivität zu erwarten sei. Ferner führe die lange Erwerbsunterbrechung zu einem Absinken der erwerbsspezifischen Produktivität.[2]

Schröder betont, dass diese Ausführungen nur gelten, wenn für Frauen eine höhere Produktivität für Haushaltsarbeiten angenommen wird. Sie erläutert, dass der umgekehrte Fall aber empirisch sehr selten sein sollte. Ferner führt sie aus, dass die bisherigen Ausführungen nur für Paare gelten, und nicht direkt für alleinstehende Frauen. Das Modell für alleinstehende Frauen komme aber im wesentlichen hinsichtlich der Effekte von Kindern zu den gleichen Aussagen.

Schröder (2010, S. 108) leitet also folgende vier Hypothesen aus Becker (1994) ab:

- Die Zahl der Kinder hat einen negativen Effekt auf die Wahrscheinlichkeit der Erwerbsbeteiligung.
- Der negative Effekt von Kindern auf die Erwerbsbeteiligung nimmt mit dem Alter der Kinder ab.
- Die Zahl der Kinder hat einen negativen Effekt auf die Wahrscheinlichkeit der Vollzeiterwerbsbeteiligung.

2 Streng genommen führt eine Erwerbsunterbrechung nach Becker (1994) und Mincer (1962) nur zu einer Stagnation der Berufserfahrung, es sei denn man nimmt zusätzlich an, dass Kenntnisse vergessen oder verlernt werden oder vorhandene Kenntnisse veralten.

- Der negative Effekt von Kindern auf die Vollzeiterwerbsbeteiligung nimmt mit dem Alter der Kinder ab.

7.2 Forschungsstand

In diesem Abschnitt soll der Forschungsstand zum Effekt von Kindern auf die Erwerbstätigkeit von Frauen dargestellt werden.[3]

Zunächst findet man bis heute Arbeiten, die den Zusammenhang zwischen der Fertilität und der Erwerbstätigkeit rein deskriptiv bearbeiten (vgl. Hofbauer 1979; Huinink 1989; Kirner und Schulz 1991; Apps und Rees 2005; Thevenon 2009). Darüber hinaus bearbeiten viele Studien den Zusammenhang zwischen der Fertilität und der Erwerbstätigkeit mit multivariaten Regressionsanalysen mit Querschnittsstudien (vgl. Heckman 1974; Franz 1985; Untiedt 1990; Kravdal 1992; Funk 1993; Calhoun 1994; Gerfin 1996; Waldfogel, Higuchi, und Abe 1999; Van der Lippe 2001; Henkens, Grift, und Siegers 2002; Vlasblom und Schippers 2004; Matysiak und Steinmetz 2008; Berninger 2009; Fouarge, Manzoni, Muffels, und Luijkx 2010; Matysiak und Vignoli 2010). Schröder kritisiert an diesen Arbeit zu Recht, dass deskriptive Analysen und Querschnittsanalysen zur Bestimmung kausaler Effekte ungeeignet sind. Sie kritisiert weiterhin an den Arbeiten, die multivariate Regressionsanalysen mit Querschnittsdaten durchführen, dass hier die Kinderzahl und das Kinderalter mangelhaft operationalisiert sind.

In neueren Arbeiten findet man zunehmend Analysen von Längsschnittdaten. Mit diesen Daten ist es grundsätzlich möglich, Effekte des Lebenszyklus einer Person auf die Erwerbstätigkeit adäquat zu untersuchen. Der bei Weitem überwiegende Teil dieser Arbeiten verwendet Methoden der Ereignisdatenanalyse (vgl. Even 1987; Tölke 1989; Desai und Waite 1991; Wenk und Garrett 1992; Felmlee 1993; Joshi und Hinde 1993; Lauterbach 1994; Giannelli 1996; Rønsen und Sundström 1996; Joesch 1997; Drobnič, Blossfeld, und Rohwer 1999; Saurel-Cubizolles, Romito, Escribà-Aguir, Lelong, Pons, und Ancel 1999; Drobnič 2000; Ondrich, Spiess, Yang, und Wagner 2000; Rønsen und Sundström 2002; Budig 2003; Berger und Waldfogel 2004; Weber 2004; Buchholz und Grunow 2006; Grunow, Hofmeister, und Buchholz 2006; Aisenbrey, Evertsson, und Grunow 2009; Matysiak 2009; Troske und Voicu 2010; Shafer 2011; Steele 2011). Die Interpretation liegt hier in Regel auf Erwerbslosigkeitsdauern nach der Geburt von Kindern oder auf bestimmten Übergangsrisiken mit und ohne Kinder. Der Großteil dieser Arbeiten verwendet die konventionellen Ereignisdatenanalyse-Methoden für kontinuierliche Zeit (vgl. Lancaster 1990; Kalbfleisch und Prentice 2002). Bei den Arbeiten, die Modelle mit diskreter Prozesszeit verwenden, findet man einzelne Arbeiten, die Erwerbstätigkeit in Voll- und Teilzeit und auch weitere Fertilität als wechselseitig bedingende Zustände modellieren. Bei diesen Arbeiten ist die unbeobachtete Heterogenität immer als von den anderen Regressoren unabhängige Variable modelliert.

Schröder kritisiert an diesem Ansatz berechtigt, dass sich erstens die Risikomenge über den Lebensverlauf selektiv verändert. Die Gruppe der Frauen mit vielen Kindern und äl-

3 Die Darstellung des Forschungsstandes deckt sich stark mit dem immer noch sehr aktuellen Überblick in Schröder und Pforr (2009) und Schröder (2010, S. 59–93, 109–115).

teren, d. h. früher geborenen, Kindern unterscheidet sich selektiv von Frauen mit wenigen bzw. ohne Kinder und später geborenen Kindern. Wenn hier nicht für alle antezedierenden d. h. selektierenden Variablen kontrolliert wird, erhält man verzerrte Schätzer. Zweitens kritisiert Schröder – wiederum berechtigt, dass man die Effekte aus Ereignisdatenmodelle nicht direkt mit dem Effekt auf Zustandswahrscheinlichkeiten vergleichen kann. Im ersten Fall betrachtet man Effekte auf Übergangswahrscheinlichkeit von einem bestimmten Ausgangszustand in einem bestimmten Zielzustand. Im zweiten Fall betrachtet man Effekte auf die absoluten Wahrscheinlichkeiten, in einem bestimmten Zustand zu sein.

Den Unterschied zwischen Effekten auf Übergänge und Effekten auf unbedingten Zustandswahrscheinlichkeiten kann man in dreierlei Weise betrachten. (1) Pragmatisch gesehen kann man die Effekte nicht direkt vergleichen. Z. B. lässt sich aus einem positiven Effekt der Kinderzahl auf die Übergangswahrscheinlichkeit von der Erwerbstätigkeit in die Arbeitslosigkeit nicht direkt sagen, ob die Wahrscheinlichkeit der Erwerbstätigkeit steigt, da umgekehrt auch die Übergangswahrscheinlichkeit aus der Arbeitslosigkeit in die Erwerbstätigkeit steigen kann. Dadurch wird nur eine mit der Kinderzahl steigende Unstetigkeit der Erwerbstätigkeit impliziert. Grundsätzlich lassen sich aber bei ausreichend ausführlicher Modellierung der Ereignisdatenmodelle auch die unbedingten Wahrscheinlichkeiten für Zustände betrachten. D. h. aus der forschungspraktischen Perspektive ist besondere Sorgfalt bei der Interpretation geboten. (2) Aus theoretischer Perspektive spricht einiges dafür, Effekte auf unbedingten Zustandswahrscheinlichkeiten zu bevorzugen. Modelle, die der neoklassischen Ökonomie folgen, wie eben die neue Haushaltsökonomie, die den Arbeiten von Becker (1994) folgt, betrachten Gleichgewichtszustände innerhalb einer Produktions- bzw. Konsumeinheit über die Zeit hinweg. D. h. diese Modelle machen in erster Linie Aussagen über einen erwarteten Zustand und nicht über die Änderung. Das impliziert, dass hier statistische Modelle über Zustandswahrscheinlichkeiten adäquater seien sollten als Ereignisdatenmodelle. (3) Grundsätzlich erlauben die Ereignisdaten-Modelle mit der Betrachtung der Übergänge eine differenzierte Modellierung als die ausschließliche Betrachtung von absoluten Zustandswahrscheinlichkeiten. Wenn man Theorien über Übergänge zwischen Zuständen hat, kann man diese adäquat nur mit Ereignisdatenmodellen untersuchen.

Bei den oben erwähnten Arbeiten wird sowohl die vorherige als auch die antizipierte Erwerbstätigkeit, aber auch eine etwaige allgemeine Neigung zur Erwerbstätigkeit und zur Mutterschaft als exogen betrachtet. Diese Annahme ist in der Regel verletzt. In der Literatur finden sich drei Ansätze, diesem Problem entgegenzutreten: Wenn man nicht ausschließen will, dass sich Erwerbstätigkeit und Mutterschaft wechselseitig beeinflussen, selbst wenn dieser Zusammenhang über latente Variablen vermittelt wird, ist man auf den Instrumentvariablenansatz und auf dynamische Längsschnittmodelle angewiesen.

Beim Instrumentvariablenansatz nutzt man im vorliegenden Zusammenhang Variablen über exogene „Fertilitätsschocks", d. h. man hat Variablen, die mit der zentralen unabhängigen Variable Kinderanzahl und Kinderalter zusammenhängen, aber nicht mit der Erwerbstätigkeit unter Kontrolle der Kinderanzahl.[4] Schröder führt hier als Beispiel die

4 Vgl. hierzu die Einführung bei Morgan und Winship (2007).

Arbeiten von Schultz (1987), Bronars und Grogger (1994), Rosenzweig und Wolpin (1994), Angrist und Evans (1998) und Jacobsen, Pearce, und Rosenbloom (1999) an, aber auch in den neueren Arbeiten von Frenette (2011) und Cáceres-Delpiano (2012) findet man diesen Ansatz. Schultz verwendet soziodemographische und Makrovariablen als Instrument, die Schröder zu Recht als völlig ungeeignet kritisiert. Die anderen Arbeiten verwenden Mehrlingsgeburten und das Geschlechterverhältnis als natürliche Experimente der zufällig exogene verringerten bzw. vergrößerten Neigung zu einem weiteren Kind. Schröders Kritik an diesen Arbeiten, dass auch hier die Annahme der Exogenität problematisch ist, wird nicht geteilt, da sowohl eine Mehrlingsgeburt als auch das Geschlecht eines Kindes als zufällig betrachtet werden kann.[5] Dennoch sind diese Arbeiten nicht unproblematisch. Das Mehrlingsgeburt und das Geschlechterverhältnis können zwar als exogen angenommen werden, haben aber einen geringen Effekt auf die Fertilitätsneigung. Außerdem ist beim Geschlechterverhältnis eine Aussage nur für Frauen möglich, die bereits zwei Kinder geboren haben, also für eine selektive Teilpopulation aller Frauen.

Als Beispiele für die Verwendung von dynamischen Längschnittdatenmodellen führt Schröder die Arbeiten von Heckman (1981) und Hyslop (1999). Das Problem bei diesen Arbeiten ist, dass hier angenommen wird, dass die Endogenität vollständig beobachtbar ist. Diese Annahme kann in der Regel als verletzt betrachtet werden.

Die dritte Klasse von Verfahren, die das Problem von endogenen Variablen zumindest teilweise beheben können, sind fixed-effects-Modelle. Hier muss man keine Annahmen über zeitlich konstante antezedierende Variablen treffen, die die Fertilität und die Erwerbstätigkeit beeinflussen. Es ist anzumerken, dass mögliche Effekte sowohl der vorherigen als auch der antizipierten Erwerbstätigkeit ausgeschlossen werden. Als Beispiel hierfür führt Schröder die Arbeiten von Schnabel (1994), Shaw (1994), Charles, Buchmann, Halebsky, Powers, und Smith (2001) und Lange (2007) an, aber auch die Arbeit von Michaud und Tatsiramos (2011) verwendet fixed-effects-Modelle. Diesen Arbeiten ist gemein, dass sie entweder die abhängige Variable nicht als kategoriale Variable modellieren oder bestehende fixed-effects-Modelle so verwenden, dass die keine konsistenten Schätzer liefern, oder bei denen die Operationalisierung der zentralen unabhängigen Variablen Kinderzahl und Kinderalter mangelhaft ist.

Schröder (2010) schließt die Lücke dahingehend, dass sie mit der Verwendung von binären logistischen Regressionsmodellen mit fixed-effects erstens einen Teil der Endogenität einfängt und zweitens durch eine extensive Operationalisierung der Kinderzahl und des Kinderalters eine differenzierte Interpretation der entsprechenden Effekte ermöglicht.

Ein Defizit bleibt aber bestehen, nämlich die notgedrungene statistische Modellierung der abhängige Variable Erwerbstätigkeit als dichotome Variable.[6] Sie diskutiert und modelliert die Erwerbstätigkeit zwar als dreistufige kategoriale Variable mit den Stufen „erwerbslos", „teilzeit erwerbstätig" und „vollzeit erwerbstätig", kann aber die Kontraste nur paarweise aber nicht gemeinsam analysieren. Im Weiteren sollen nun erstens die Er-

5 Beim Geschlechterverhältnis als Instrument besteht das Problem, dass das Geschlechterverhältnis nicht in allen Kulturkreisen eine Rolle für die Familienplanung spielt (vgl. Ebenstein 2009).

6 Siehe Schröder (2010, S. 117, Fussnote 16).

gebnisse von Schröder (2010) repliziert werden und zweitens daran anschließend unter Verwendung des femlogit-Modells die Lücke geschlossen werden.

7.3 Analysen

Im Folgenden soll eine Auswahl der Analysen von Schröder (2010) zum Effekt der Fertilität auf die Erwerbstätigkeit repliziert und im Vergleich mit dem Original dargestellt werden. Schröder wendet bei ihrer Untersuchung des Effekt der Fertilität auf die Erwerbstätigkeit gegenüberstellend dichotome fixed-effects-Modelle, pooled-Modelle und Hybrid-Modelle an.[7] In der vorliegenden Arbeit werden nur die fixed-effects-Schätzmodelle repliziert. Die Zielrichtung liegt hier auf der gemeinsamen Schätzung der drei Alternativen der abhängigen Variablen.

7.3.1 Daten

Schröder (2010) verwendet für die Analyse die Daten des dritten Familiensurveys aus dem Jahr 2000, der vom Deutschen Jugendinstitut (DJI) durchgeführt wurde (vgl. Deutsches Jugendinstitut (DJI), München 2003; Bien und Rathgeber 2000; Infratest Burke Sozialforschung 2000; Bien und Marbach 2003). Im Familiensurvey 2000 wurden 10 318 Personen persönlich befragt, wobei 8 091 Befragte aus einer neuen Stichprobe stammen und die übrigen aus einer Panelstichprobe und einer speziellen Jugendlichen-Stichprobe stammen. Bei den Analysen werden nur die Befragten aus der neuen Stichprobe verwendet.

Analysestichprobe

Weiterhin wird die Stichprobe für die Analyse folgendermaßen eingeschränkt. Es werden nur Befragte, die alle folgenden Eigenschaften erfüllen:

- Frauen und
- Befragte, die zwischen 1944 und 1982 geboren wurden, d. h. zum Befragungszeitpunkt zwischen 18 und 55 Jahre alt waren, und
- mit deutsche Staatsbürgerschaft und
- wohnhaft in Westdeutschland und
  - in Westdeutschland geboren oder
  - Flüchtling aus ehem. deutschen Ostgebieten oder
  - spätestens im Alter von 15 Jahren aus der DDR in die BRD übergesiedelt.

7 Das Hybrid-Modell ist eine gepoolte logistische Regression, bei der für alle zeitlich variierenden Kovariate within-Effekte geschätzt werden. D. h. für zeitlich variierende unabhängige Variable X_{it} berechnet man für jede Beobachtungseinheit den Mittelwert der Variablen über die Messzeitpunkte $\bar{X}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} X_{it}$. In der gepoolten logistischen Regression wird dann der Mittelwert und die Abweichung vom Mittelwert für jede unabhängige Variable und zusätzlich die zeitlich konstanten unabhängigen Variable als Regressoren spezifiziert. Der Vorteil besteht darin, dass man Quasi-fixed-effects-Schätzer erhält und dennoch zusätzlich Effekte für zeitlich konstante Variablen schätzen kann. (vgl. Neuhaus und Kalbfleisch 1998).

Da Schröder (2010) die Analysedaten für eine Replikation der Analyse zur Verfügung gestellt hat, können hier die aufbereiteten Daten exakt wiedergegeben werden. Die Analysestichprobe enthält nach der oben dargestellten Definition 3 022 Beobachtungseinheiten. Nach der Datenbereinigung bleiben 2 237 Beobachtungseinheiten. Da die Daten des Familiensurvey bei Schröder in der maximalen verfügbaren Genauigkeit verwendet werden, liegen für die ca. 2 200 Beobachtungseinheiten insgesamt ca. 480 000 Messzeitpunkte vor. Schon für das binäre fixed-effects-Modell bei Schröder führen diese langen Zeitreihen zu extrem langen Rechenzeiten. Daher reduziert Schröder ihre Stichprobe, indem sie nur die Monate Mai und November jeden Jahres verwendet. Dadurch reduziert sich die Anzahl der Messzeitpunkte auf insgesamt 80 639. Die Reduktion führt dazu, dass die Schätzer zwar konsistent bleiben, aber weniger effizient sind als mit der Gesamtstichprobe.

Diese Fälle gehen nur in die pooled- und Hybrid-Modelle ein. Da bei den fixed-effects-Modellen nur Beobachtungseinheiten berücksichtigt werden, bei denen die abhängige Variable über die Messzeitpunkte hinweg variiert, wird die Stichprobe weiter eingeschränkt. So bleiben für die fixed-effects-Modelle über die Erwerbstätigkeit 1 431 Beobachtungseinheiten mit insgesamt 56 395 Messzeitpunkten übrig. Für die FE-Modelle über die Vollzeiterwerbstätigkeit bleiben 1 479 Beobachtungseinheiten mit insgesamt 59 215 Messzeitpunkten übrig.

Für die Replikation der fixed-effects-Modelle von Schröder mit dem femlogit-Modell ist auch diese Anzahl der Messzeitpunkte noch zu groß.[8] Daher wird die Anzahl der Messzeitpunkte weiter reduziert. Im Gegensatz zu Schröder (2010) wird in der im Weiteren ausgeführten Replikation nicht zwei Messzeitpunkte pro Jahr, sondern ein Messzeitpunkt alle sechs Jahre betrachtet. Dies entspricht einer Reduktion auf ein Sechstel der Daten von Schröder. In der vorliegenden Replikation liegen also insgesamt 2 123 Beobachtungseinheiten mit insgesamt 12 581 Messzeitpunkten vor. Durch das Entfernen uninformativer Beobachtungseinheiten ohne Varianz auf der abhängigen Variable über die Messzeit reduziert sich die Analysestichprobe weiter. Beim Kontrast zwischen Erwerbstätigkeit und Nichterwerbstätigkeit liegen für die fixed-effects-Modelle 996 Einheiten mit 6 845 Messungen vor, für den Kontrast zwischen Vollzeit- und Teilzeiterwerbstätigkeit liegen 1 052 Frauen mit 7 319 Zeitpunkte vor. Für das Vollmodell mit drei Alternativen liegen 1 199 Beobachtungseinheiten mit 8 246 Messzeitpunkten vor.[9]

Operationalisierung

Die verwendeten Variablen wurden folgendermaßen operationalisiert. Schröder (2010) verwendet als abhängige Variable den Beschäftigungsstatus in zwei Formen. Sie betrachtet einerseits den Kontrast zwischen Erwerbstätigkeit und Nichterwerbstätigkeit und andererseits den Kontrast zwischen Vollzeiterwerbstätigkeit und Teilzeit- oder Nichterwerbs-

8 Die Rechenzeit wächst mit der mittlere Anzahl der Messzeitpunkte pro Beobachtungseinheit sowohl beim felogit- als auch beim femlogit-Modell ungefähr fakultativ. In beiden Fällen wächst die Rechenzeit ungefähr linear mit Summe der Permutationen der abhängigen Variable über alle Personen. Eine Schätzung mit den ungekürzten Daten hätte in der gegenwärtigen Fassung der Implementation ungefähr 15 000 Jahre gedauert.

9 Hier zeigt sich der Vorteil der gemeinsamen Analyse mit dem femlogit-Modell gegenüber getrennte dichotomen Modellen, wie er bei Long (1997, S. 149–151) beschrieben wird.

tätigkeit. Die Erwerbsbiographie liegt im Familiensurvey monatsgenau vor. Erwerbsunterbrechungen sind erfasst, wenn sie länger als vier Monate gedauert haben. Eine Frau ist in einem bestimmten Monat vollzeiterwerbstätig, wenn sie mindestens 35 Stunden in der Woche gearbeitet hat, und teilzeiterwerbstätig, wenn sie weniger als 35 Stunden in der Woche gearbeitet hat und nicht im gleichen Monat einer Schul- oder Berufsausbildung nachgegangen ist.

Die Fertilität, d. h. die Kinderzahl und das Kinderalter wird aus der Information über alle eigenen und im Haushalt lebenden Kinder berechnet. Beim Familiensurvey ist auch Information über zum Befragungszeitpunkt bereits weggezogene oder verstorbene Kinder gegeben. Weiterhin ist für die noch lebenden Kinder bekannt, ob diese leiblich, Kinder des aktuellen Partners, adoptiert oder Pflegekinder sind. Für die Analyse werden nur leibliche Kinder betrachtet.[10] Aus den Geburtsdaten dieser Kinder wird für jeden Monat drei Dummies über die Anzahl der Kinder gebildet, die anzeigen, ob die Beobachtungseinheit ein Kind, zwei Kinder oder mehr als zwei Kinder hat. Entsprechend sind Monate, in denen die Frau, die keine Kinder hat, die Referenzkategorie. Weiterhin wird für jeden Monat eine Variable über das Alter des gegenwärtig jüngsten Kindes gebildet. Diese Variable wird zusätzlich quadriert und kubisch aufgenommen, um den nichtlinearen Effekt des Alters des jüngsten Kindes hinreichend glatt abzubilden. Schließlich werden diese drei Variable mit den Kinderdummies selbst interagiert. D. h. für die drei Altersvariablen werden unterschiedliche Effekt geschätzt, je nachdem ob die Frau nur ein Kind, zwei Kinder oder mehr als zwei Kinder hat. Zuletzt wird neben den bereits geborenen Kindern auch die Schwangerschaft berücksichtigt, was einem Effekt des negativen Alters der Kinder entspricht. Hierfür werden drei Dummies für drei Phasen der Schwangerschaft berücksichtigt, wobei Monate ohne Schwangerschaft die Referenzkategorie bilden.

Neben diesen inhaltlichen hauptsächlich relevanten Variablen werden im wesentlichen drei Gruppen von Kontrollvariablen aufgenommen. Wie auch Schröder anführt, ist nach Becker (1994) das Arbeitsangebot vom Lohnsatz abhängig. Schröder operationalisiert den Lohnsatz als zeitlich konstante Variable mit dem höchsten erreichten Bildungsabschluss nach Ende der ersten Ausbildungsphase. Da die Bildung damit eine zeitlich invariante Variable ist, wird sie nur in den pooled- und Hybrid-Modellen aufgenommen. Das zweite Kontrollkonstrukt ist der Familienstand bzw. Partnerschaftsstatus. Erst wenn ein Partner vorhanden ist, kann Arbeitsteilung stattfinden. Außerdem ist die institutionalisierte Ehe eine Versicherung für einseitige Spezialisierung bei etwaiger Trennung. Diese Information wird als zeitlich variierende kategoriale Variable erfasst. Im einzelnen liegen fünf Dummies für jeden Monat vor für den Zustand der in Beziehung, aber getrennt lebenden (LAT), der Zusammenlebenden, der Verheirateten, der Geschiedenen und der Verwitweten. Damit ist der Singlestatus die Referenzkategorie. Schließlich wird für die Prozesszeit, die Kohorte und die Kalenderzeit korrigiert. Alle drei Informationen sind linear abhängig, d. h. sie können nicht perfekt voneinander getrennt werden. Zunächst werden die Geburtskohorten grob kategorisiert als zeitlich konstante Variablen nur in die pooled- und Hybrid-Modelle aufgenommen. Im fixed-effects-Modell sind aber auch die Prozess- und

10 Da beim Familiensurvey das Verwandtschaftsverhältnis zwischen dem Befragten und einem verstorbenen Kind nicht zweifelsfrei bekannt ist, werden Mütter mit verstorbenen Kindern aus der Stichprobe genommen.

die Kalenderzeit linear abhängig. Dafür wird die Kalenderzeit nur grob mit einem Proxy abgebildet, indem die Arbeitslosenquote in Westdeutschland linear und quadriert aufgenommen wird. Die Prozesszeit wird als Folge von Altersdummies in der vollständigen Auflösung aufgenommen.

7.3.2 Ergebnisse

In der Arbeit von Schröder (2010) werden zur Untersuchung des Effekts der Fertilität auf der Erwerbstätigkeit verschiedene quantitative Verfahren verwendet. Schröder gibt einen Überblick über die Ergebnisse und Verfahren in der Literatur und analysiert den Zusammenhang deskriptiv bivariat mit Hilfe von Längsschnittdaten. Die Untersuchung gipfelt in einer vergleichenden Analyse des Effekts der Kinderzahl und des Kinderalters auf die Erwerbstätigkeit an sich und die Erwerbstätigkeit in Vollzeit mit binären pooled-logit und fixed-effects-logit-Verfahren. Schröder findet einen starken negativen Effekt der Kinderzahl und einen starken positiven Effekt des Kinderalters auf die Erwerbstätigkeit und die Vollzeiterwerbstätigkeit.

Im Folgenden sollen nun die Analysen von Schröder teilweise repliziert werden und durch die Anwendung des femlogit-Modells ergänzt werden. Betrachten wir zunächst den Ausgangspunkt, den die Analysen von Schröder (2010) bilden. In Tabelle 7.1 sind die exakt replizierten binären pooled-logit- und fixed-effects-Modelle dargestellt. Bei den Modellen ist die abhängige Variable in zwei Formen gebildet: (1) Erwerbstätigkeit an sich gegenüber der Erwerbslosigkeit und (2) die Erwerbstätigkeit in Vollzeit gegenüber Erwerbstätigkeit in Teilzeit und Erwerbslosigkeit.

Tabelle 7.1: Binäre pooled- und fixed-effects-logit-Modelle nach Schröder (2010)

	Erwerbstätigkeit		Vollzeiterwerbstätig	
	pooled[a]	fixed-effects	pooled[a]	fixed-effects
	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)
ein Kind	0,051***	0,003***	0,056***	0,005***
	(0,005)	(0,000)	(0,006)	(0,000)
zwei Kinder	0,037***	0,001***	0,033***	0,001***
	(0,005)	(0,000)	(0,005)	(0,000)
>2 Kinder	0,030***	0,001***	0,040***	0,001***
	(0,006)	(0,000)	(0,010)	(0,000)
Alter des jüngsten Kindes (1 Kind)	1,418***	1,743***	1,176***	1,132***
	(0,045)	(0,042)	(0,035)	(0,030)
Alter^2 –"–	0,983***	0,977***	0,995*	1,004
	(0,003)	(0,002)	(0,003)	(0,002)
Alter^3 –"–	1,000***	1,000***	1,000	1,000
	(0,000)	(0,000)	(0,000)	(0,000)
Alter des jüngsten Kindes (2 Kinder)	1,353***	1,818***	1,130**	1,221***
	(0,040)	(0,049)	(0,042)	(0,041)

Tabelle 7.1: (Fortsetzung)

	Erwerbstätigkeit		Vollzeiterwerbstätig	
	pooled[a]	fixed-effects	pooled[a]	fixed-effects
	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)
Alter^2 –"–	0,987*** (0,003)	0,976*** (0,002)	1,001 (0,003)	1,008** (0,003)
Alter^3 –"–	1,000*** (0,000)	1,000*** (0,000)	1,000 (0,000)	1,000*** (0,000)
Alter des jüngsten Kindes (>2 Kinder)	1,354*** (0,070)	1,962*** (0,109)	1,137* (0,074)	1,206*** (0,067)
Alter^2 –"–	0,985** (0,005)	0,965*** (0,005)	1,000 (0,006)	1,007 (0,005)
Alter^3 –"–	1,000* (0,000)	1,001*** (0,000)	1,000 (0,000)	1,000 (0,000)
Schwangerschafts-beginn	0,632*** (0,055)	0,524*** (0,067)	0,767** (0,071)	0,536*** (0,079)
Schwangerschafts-mitte	0,556*** (0,041)	0,317*** (0,031)	0,694*** (0,055)	0,395*** (0,042)
Schwangerschafts-ende	0,189*** (0,020)	0,036*** (0,005)	0,243*** (0,029)	0,065*** (0,009)
living apart together	1,121 (0,160)	1,056 (0,095)	0,998 (0,122)	1,224* (0,114)
kohabitierend	0,913 (0,139)	0,753** (0,079)	0,838 (0,121)	0,863 (0,090)
verheiratet	0,568*** (0,084)	0,278*** (0,025)	0,463*** (0,067)	0,298*** (0,029)
geschieden	1,271 (0,264)	0,740* (0,092)	1,573* (0,314)	1,730*** (0,221)
verwitwet	0,921 (0,350)	0,715 (0,134)	0,469* (0,160)	0,265*** (0,049)
Hauptschule mit Ausbildung	1,286 (0,188)		1,593** (0,260)	
mittlere Reife	1,344 (0,247)		1,314 (0,288)	
mittlere Reife mit Ausbildung	1,964*** (0,285)		2,212*** (0,360)	
Abitur	0,747 (0,185)		0,923 (0,248)	
Abitur mit Ausbildung	1,822** (0,336)		1,915** (0,393)	
FH-/Uni-Abschluss	3,320*** (0,657)		2,595*** (0,567)	
Geburtskohorten 1954–1963	1,082 (0,120)		0,696** (0,083)	

Tabelle 7.1: (Fortsetzung)

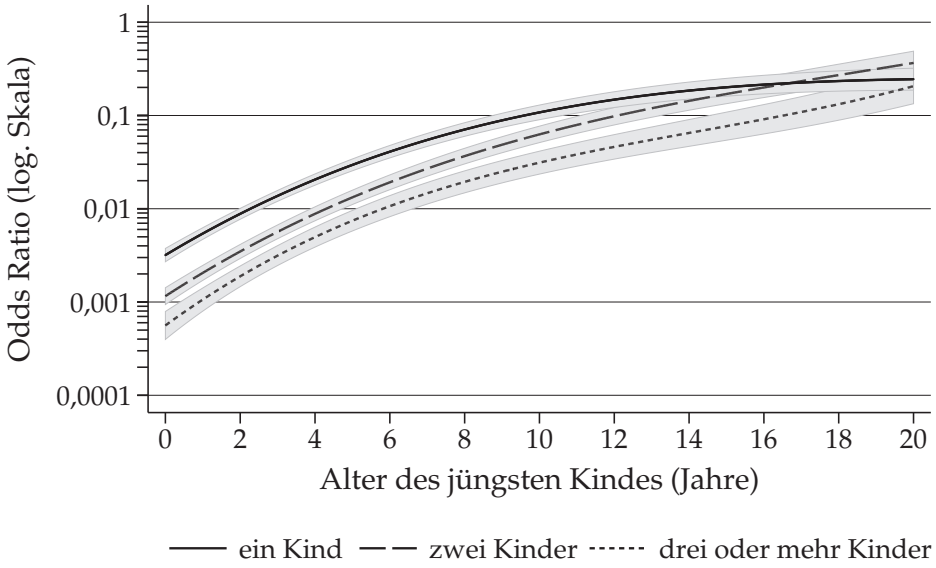
	Erwerbstätigkeit		Vollzeiterwerbstätig	
	pooled[a]	fixed-effects	pooled[a]	fixed-effects
	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)
Geburtskohorten 1964–1973	0,949 (0,119)		0,626*** (0,085)	
Geburtskohorten 1974–1983	0,627* (0,128)		0,333*** (0,067)	
Arbeitslosenquote	0,942 (0,045)	0,706*** (0,021)	1,013 (0,015)	1,008 (0,011)
Arbeitslosenquote^2	1,004 (0,003)	1,020*** (0,002)		
Frauen	2 237	1 431	2 237	1 479
Messzeitpunkte	80 639	56 395	80 639	59 215
χ^2/(df)[b]	1 711,1 (71)	20 646,9 (62)	1 729,8 (70)	29 994,6 (61)
R^2_{MF}	0,216	0,392	0,301	0,545

*: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. [a]: Bei pooled-Modellen heteroskedastie- und panelrobuste Standardfehler. [b]: Bei pooled-Modellen Wald-Test, bei fixed-effects-Modellen LR-Test. Referenzkategorie: Kinderlos, nicht schwanger, Single, kein Schulabschluss oder Hauptschulabschluss ohne Ausbildung, Geburtskohorte 1944–1953. Nicht dargestellte Kontrollvariablen: jahresweise Altersdummies mit Alter 20 als Referenzkategorie.

Die Ergebnisse in Tabelle 7.1 zeigen, dass Ergebnisse von Schröder (2010) vollständig und exakt repliziert werden können (vgl. ebd., S. 154, 172). Grundsätzlich entsprechen die Effekte der Kinderzahl und des Kinderalters der Erwartungen der Theorie. Mit steigender Kinderzahl sinkt das Odds, erwerbstätig gegenüber nicht erwerbstätig zu sein, bzw. in Vollzeit zu arbeiten gegenüber nicht in Vollzeit zu arbeiten. Für Mütter erhöht sich mit steigendem Alter des jüngsten Kindes das Odds, erwerbstätig zu sein, bzw. in Vollzeit zu arbeiten. Diese Effekte sind im Wesentlichen für beide Kontraste gleich. Da die Vollzeiterwerbstätigkeit die höhere Hürde darstellt, sind hier die Effekt betragsmäßig kleiner.

Auch hier findet man den kontraintuitiven Befund von Schröder: Die „negativen” OR-Effekte der Kinderzahl-Variablen bei den fixed-effects-Modellen sind entgegen der Erwartung betragsmäßig größer, d. h. sie liegen näher an null. Die positiven Effekte des Alters der Kinder bei den fixed-effects-Modellen sind ebenso betragsmäßig größer als bei den pooled-Modellen.

Die zentralen Effekte der Kinderzahl und des Kinderalters aus den binären fixed-effects-logit-Modellen sind in den Abbildungen 7.1 und 7.2 dargestellt. In den Abbildungen sind jeweils die Odds-Ratio-Effekte eines Kindes auf die Erwerbstätigkeit bzw. die Vollzeiterwerbstätigkeit dargestellt.

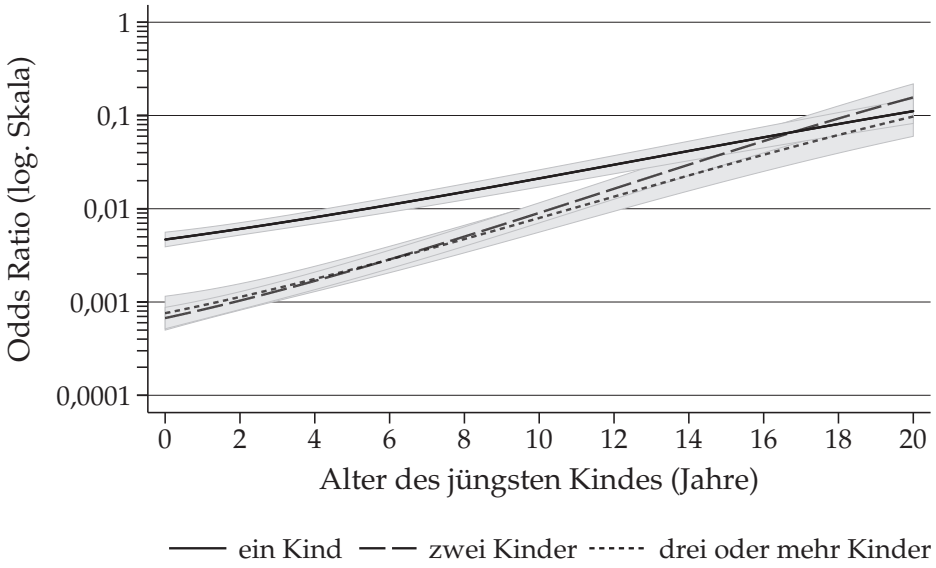


Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).

Abbildung 7.1: Odds Ratio von Erwerbstätigkeit mit Kindern vs. ohne Kinder nach Kinderalter aus dem fixed-effects-Modell über alle Zeitpunkte

In Abbildung 7.1 sieht man, dass das erste neugeborene Kind (mit Alter $= 0$) das Odds, erwerbstätig zu sein, um mehr als das 300-fache senkt. Das zweite neugeborene Kinder senkt dann das Odds sogar um einen Faktor von über 1 000 und weitere neugeborene Kinder senken das Odds gegenüber Kinderlosen um ungefähr das 4 000-fache. Mit steigendem Alter geht der Unterschied zwischen kinderlosen Frauen und Müttern im Odds, erwerbstätig zu sein, zurück. Das Odds, erwerbstätig zu sein, ist für eine Mutter eines 14-jährigen Einzelkindes oder 14-jährigen Kindes mit einem älteren Geschwister knapp ein fünftel des entsprechenden Odds einer kinderlosen Frau. Das Odds für eine Mutter eines 14-jährigen Kindes mit mehr als einem älteren Geschwister ist ungefähr 70 % des Odds einer Frau ohne Kinder. Die Konfidenzintervalle in dieser Abbildung sind wie bei den folgenden Abbildungen über Odds-Ratio-Effekte in Anlehnung an Brambor, Clark, und Golder (2006) berechnet.

Abbildung 7.2 zeigt, dass der Effekt von neugeborenen Kindern auf das Odds, vollzeiterwerbstätig zu sein, insgesamt ähnlich ist wie auf das Odds, überhaupt erwerbstätig zu sein. Das erste Neugeborene (mit Alter $= 0$) senkt das Odds der Vollzeiterwerbstätigkeit im Vergleich zu Kinderlosen um den Faktor von ca. $0,004$. Neugeborene nach dem ersten Kind senken das Odds wie oben noch einmal um mehr als das Vierfache. Auch hier gleichen sich die Odds für Mütter von einem oder mehreren Kindern mit dem Alter des jüngsten Kindes im Vergleich zu den kinderlosen Frauen an. Wenn das jüngste Kind ungefähr 16 Jahre alt, sind die Odds ungefähr gleich.



Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).

Abbildung 7.2: Odds Ratio von Vollzeiterwerbstätigkeit mit Kindern vs. ohne Kinder nach Kinderalter aus dem fixed-effects-Modell über alle Zeitpunkte

Man erkennt aber auch Unterschiede zu den Odds der Erwerbstätigkeit. Erstens steigt das Odds der Vollzeiterwerbstätigkeit mit dem Alter des jüngsten Kindes nach der Geburt weniger schlagartig wieder an wie das Odds der Erwerbstätigkeit. Hier ist der Anstieg unabhängig von der Kinderzahl ungefähr linear. Bei der Erwerbstätigkeit im Allgemein flacht der Anstieg sichtbar mit steigendem Alter des jüngsten Kindes ab. Zweitens ist zu erkennen, dass das Odds für Vollzeiterwerbstätigkeit für Mütter im Allgemeinen gegenüber kinderlosen Frauen auf einem niedrigen Niveau bleibt wie das Odds der Erwerbstätigkeit. Selbst wenn das jüngeste Kind erwachsen ist, ist das Odds vollzeiterwerbstätig ca. 20 mal kleiner als das der Kinderlosen. Das entsprechende Odds, überhaupt erwerbstätig zu sein, ist nur ca. viermal kleiner als das der kinderlosen Frauen.

Insgesamt können die Befunde der binären fixed-effects-logit-Modelle von Schröder (2010) vollständig und exakt repliziert werden. Im Weiteren soll nun dargestellt werden, welchen Beitrag die Erweiterung dieser Analysen um eine gemeinsame Schätzung in einem multinomialen Modell mit fixed-effects bringt.

Wie bereits oben erläutert wurde, müssen für die Verwendung des femlogit-Modell auf die Analysen von Schröder die Daten gekürzt werden, um die Rechenzeit in einem handhabbaren Bereich zu halten. Schröder verwendet zwei Messzeitpunkte pro Jahr. Diese Frequenz wird für die Verwendung des femlogit-Modell auf einen Messzeitpunkt alle drei Jahre verringert.

Tabelle 7.2: Binäre fixed-effects-logit-Modelle nach Schröder (2010) und multinomiales logit-Modell über die gekürzten Daten

	Binäres Logit		Multinomiales Logit[a]	
	Erwerbstätigkeit	Vollzeiterwerbstätig	Teilzeit	Vollzeit
	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)
ein Kind	0,005***	0,013***	0,085***	0,005***
	(0,001)	(0,003)	(0,030)	(0,002)
zwei Kinder	0,002***	0,003***	0,054***	0,001***
	(0,001)	(0,001)	(0,022)	(0,000)
>2 Kinder	0,002***	0,008***	0,023***	0,002***
	(0,001)	(0,004)	(0,013)	(0,001)
Alter des jüngsten Kindes (1 Kind)	1,687***	1,131	1,847***	1,392***
	(0,104)	(0,074)	(0,151)	(0,107)
Alter^2 –"–	0,978***	1,004	0,969***	0,995
	(0,005)	(0,005)	(0,006)	(0,006)
Alter^3 –"–	1,000**	1,000	1,000**	1,000
	(0,000)	(0,000)	(0,000)	(0,000)
Alter des jüngsten Kindes (2 Kinder)	1,753***	1,126	1,997***	1,473***
	(0,119)	(0,096)	(0,170)	(0,147)
Alter^2 –"–	0,980***	1,016*	0,963***	1,003
	(0,006)	(0,007)	(0,007)	(0,009)
Alter^3 –"–	1,000*	1,000*	1,001***	1,000
	(0,000)	(0,000)	(0,000)	(0,000)
Alter des jüngsten Kindes (>2 Kinder)	1,775***	0,971	1,840***	1,293
	(0,242)	(0,139)	(0,290)	(0,229)
Alter^2 –"–	0,971*	1,023	0,964*	1,002
	(0,014)	(0,013)	(0,016)	(0,018)
Alter^3 –"–	1,001	0,999	1,001	1,000
	(0,000)	(0,000)	(0,000)	(0,000)
Schwangerschaftsbeginn	0,443*	0,608	0,420	0,444
	(0,172)	(0,262)	(0,269)	(0,207)
Schwangerschaftsmitte	0,633	0,508*	1,302	0,612
	(0,192)	(0,165)	(0,527)	(0,221)
Schwangerschaftsende	0,029***	0,103***	0,057***	0,064***
	(0,012)	(0,047)	(0,043)	(0,031)
living apart together	1,005	1,054	1,031	1,066
	(0,290)	(0,330)	(0,526)	(0,361)
kohabitierend	0,561	0,654	0,783	0,651
	(0,177)	(0,201)	(0,399)	(0,229)
verheiratet	0,239***	0,315***	0,372*	0,233***
	(0,062)	(0,089)	(0,167)	(0,072)
geschieden	0,686	1,937	0,327*	1,559
	(0,237)	(0,686)	(0,175)	(0,630)

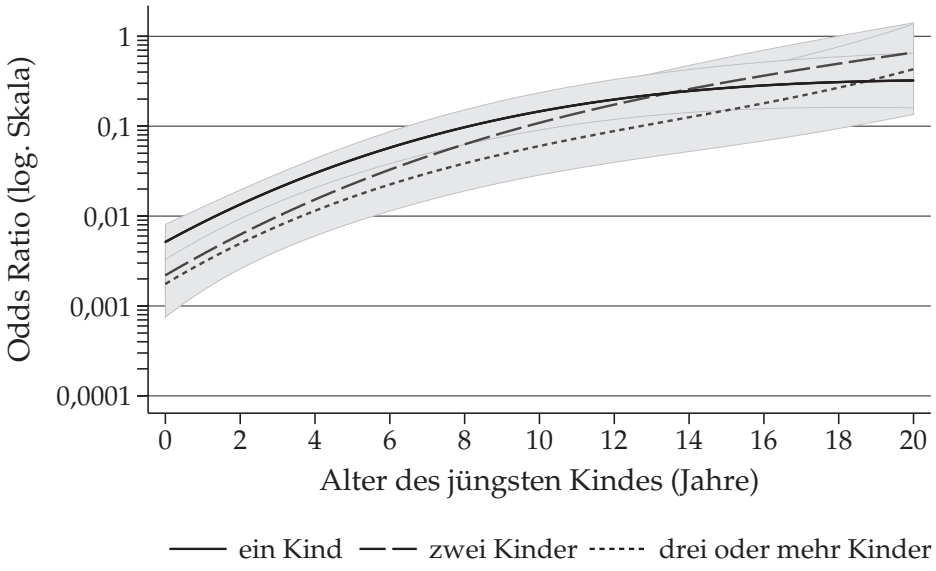
Tabelle 7.2: (Fortsetzung)

	Binäres Logit		Multinomiales Logit[a]	
	Erwerbs-tätigkeit	Vollzeiter-werbstätig	Teilzeit	Vollzeit
	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)	exp(β)/(SE)
verwitwet	0,558 (0,267)	0,274** (0,134)	2,767 (2,137)	0,341 (0,202)
Arbeitslosenquote	0,801** (0,063)	1,026 (0,028)	0,899 (0,107)	0,954 (0,096)
Arbeitslosenquote^2	1,013* (0,006)		1,004 (0,008)	1,004 (0,007)
Frauen	1 037	1 065	1 230	
Messzeitpunkte	7 114	7 393	8 428	
χ^2/(df)	2 690,0 (59)	3 442,1 (58)	4 639,1 (118)	
R^2_{MF}	0,459	0,583	0,558	

*: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. [a]: Referenzkategorie nicht erwerbstätig. Referenzkategorie: Kinderlos, nicht schwanger, Single. Nicht dargestellte Kontrollvariablen: jahresweise Altersdummies mit Alter 20 als Referenzkategorie.

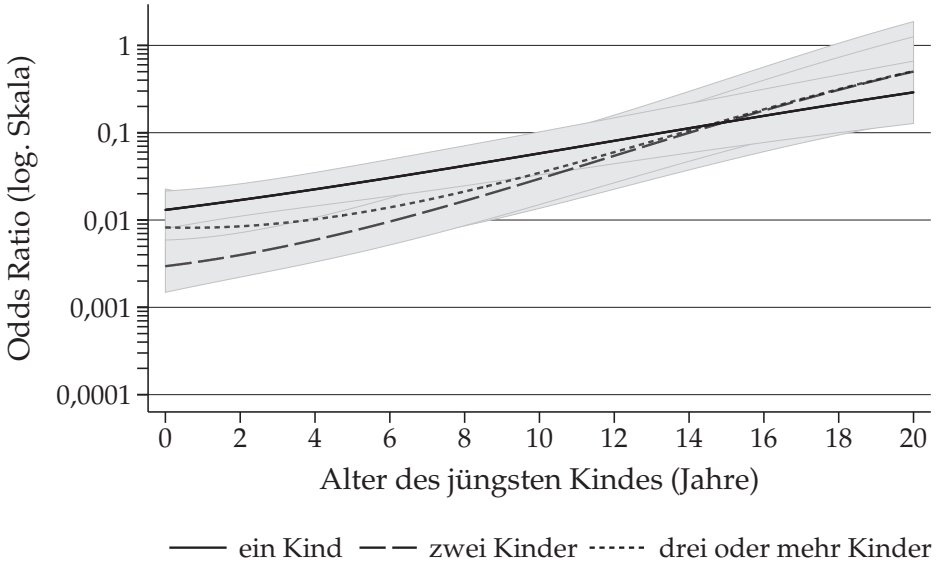
In Tabelle 7.2 sind die beiden binäre fixed-effects-Modelle wie bei Schröder und das gemeinsame multinomiale logistische Modell über die gekürzten Daten dargestellt. Zunächst ist zu erkennen, dass durch Beschränkung der Messzeitpunkte die Standardfehler der Schätzer steigen. Grundsätzlich behalten die Effekte aber die Richtung und bis auf einzelne Ausnahmen die Größenordnung der Effektstärke. Zur Verdeutlichung des Anstieg der Standardfehler der Schätzer sind die Odds-Ratio-Effekte der Kinderzahl und des Kinderalters für den gekürzten Datensatz in den Abbildungen 7.3 und 7.4 dargestellt.
Man sieht, dass die geschätzten Effekte auf Basis des gekürzten Datensatz insgesamt sehr ähnlich wie mit dem vollständigen Datensatz sind. Durch Beschränkung der Messzeitpunkte werden die Standardfehler der Effekte größer und damit die Konfidenzintervalle breiter. So unterscheiden sich die Odds, erwerbstätig bzw. vollzeiterwerbstätig zu sein, für Mütter mit mehr als einem Kind, wenn das jüngste Kind erwachsen ist, nicht mehr signifikant vom Odds der kinderlosen Frauen. Dies rührt hauptsächlich daher, dass die Stichprobe wenige Frauen enthält, die so viele Kinder in höherem Alter haben. Insgesamt zeigt sich, dass die Ergebnisse von Schröder (2010) auch mit den gekürzten Daten ausreichend gut repliziert werden.

Das eigentliche Interesse liegt auf den Effekten im femlogit-Modell. Wie bereits bei der Darstellung der Theorie erläutert wurde, ist zu erwarten, dass mit steigender Kinderzahl das Arbeitsangebot sinkt, und mit steigendem Alter des jüngsten Kindes das Angebot steigt. Damit erwarten wir also stärkere negative Effekte der Kinderzahl und stärkere



Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).

Abbildung 7.3: Odds Ratio von Erwerbstätigkeit mit Kindern vs. ohne Kinder nach Kinderalter aus dem fixed-effects-Modell mit gekürzten Daten



Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).

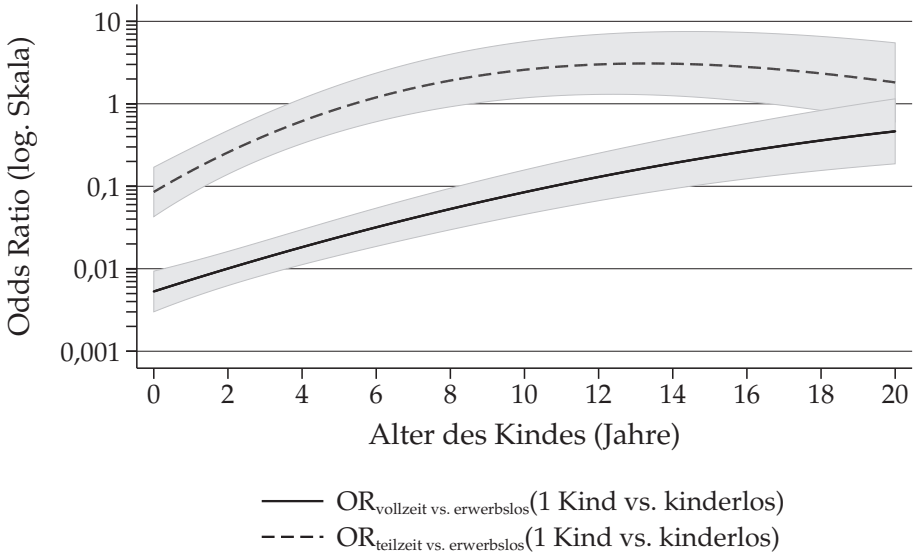
Abbildung 7.4: Odds Ratio von Vollzeiterwerbstätigkeit mit Kindern vs. ohne Kinder nach Kinderalter aus dem fixed-effects-Modell mit gekürzten Daten

positive Effekt des Kinderalters auf die Vollzeiterwerbstätigkeit gegenüber der Nichterwerbstätigkeit als auf den Kontrast zwischen Teilzeit- und Nichterwerbstätigkeit.

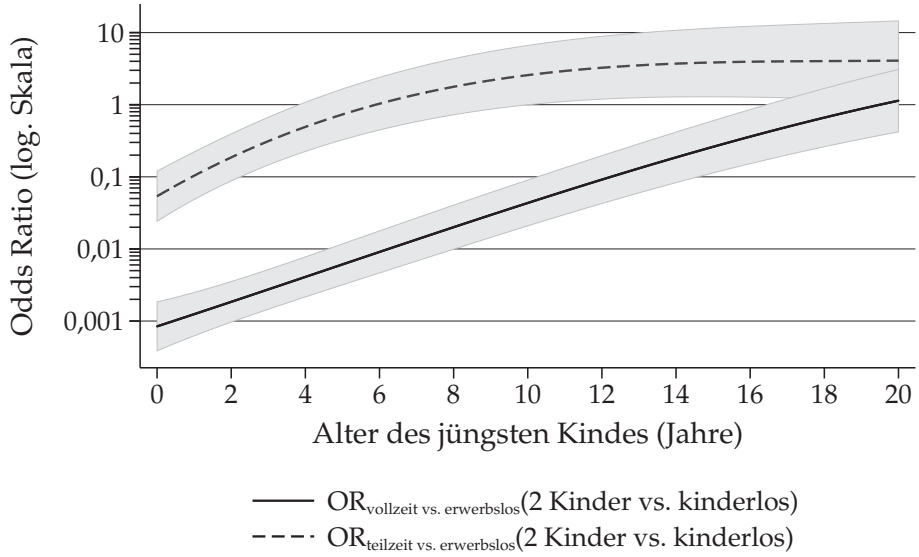
Wie die Ergebnisse in der dritten und vierten Spalte Tabelle 7.2 zeigen, werden diese Erwartungen überwiegend bestätigt. Das erste neugeborene Kind senkt das Odds zwischen Erwerbstätigkeit in Vollzeit gegenüber Erwerbslosigkeit um den Faktor 200. Das Odds zwischen Erwerbstätigkeit in Teilzeit gegenüber Erwerbslosigkeit sinkt nur um ungefähr den Faktor 12. In ähnlicher Weise ist das Odds vollzeiterwerbstätig statt erwerbslos zu sein für eine Mutter eines einjährigen Einzelkindes 1,4-mal höher als für eine Mutter eines Neugeborenen. D. h. mit steigendem Kinderalter nimmt der starke negative Effekt eines Kindes wieder ab. Demgegenüber steigt das Odds teilzeiterwerbstätig gegenüber nicht erwerbstätig zu für eine Mutter eines einjährigen Einzelkindes sogar um den Faktor von 1,8 gegenüber der Mutter eines Neugeborenen. Hier nimmt der negative Effekt der Kinderzahl mit dem Kinderalter also schneller ab.

Besser veranschaulicht findet man diese unterschiedlichen „Erholungs"-Effekte in den Abbildungen 7.5. In der ersten Teilabbildung ist der Odds-Ratio-Effekt für Frauen mit einem Kind gegenüber kinderlosen Frauen dargestellt, in der zweiten Teilabbildung der entsprechende Effekt für Frauen mit zwei Kindern gegenüber Kinderlosen und in der letzten Abbildung der Effekt für Frauen mit mehr als zwei Kindern. In den Abbildungen sind mehrere Dinge erkennbar. Zunächst einmal ist, wie schon aus Tabelle 7.2 leicht erkennbar war, der schwächere Effekt der Kinderzahl auf die Erwerbstätigkeit in Teilzeit in den Abbildungen offensichtlich. Wenn das jüngste Kind gerade neugeboren ist, ist das Odds-Ratio von Erwerbstätigkeit in Teilzeit gegenüber Nichterwerbstätigkeit zwischen Frauen mit einem Kind und Kinderlosen mehr als 10-mal größer als das Odds-Ratio von Vollzeiterwerbstätigkeit gegenüber Erwerbslosigkeit.
Das entsprechende Verhältnis der Odds ist bei Frauen mit zwei Kindern gegenüber Kinderlosen sogar mehr als 20-mal größer und das Verhältnis der Odds ist bei Frauen mit drei oder mehr Kinder gegenüber Kinderlosen ungefähr zehnmal größer. Mit steigenden Alter des jüngsten Kindes geht bei allen drei Vergleichspaarungen das Odds zwischen Erwerbstätigkeit in Teilzeit und Nichterwerbstätigkeit beschleunigt zurück. Wenn das jüngste Kind vier Jahre alt ist, ist das Odds zwischen Erwerbstätigkeit in Teilzeit und Nichterwerbstätigkeit für Frauen mit einem oder zwei Kindern nicht mehr signifikant verschieden vom entsprechenden Odds für Kinderlose. Bei Frauen mit mehr als zwei Kindern ist der Unterschied zu Kinderlosen erst, wenn das jüngste Kind sieben Jahre alt ist, nicht mehr signifikant.

Es ist besonders zu beachten, dass bei Frauen mit einem oder zwei Kindern das Odds, in Teilzeit zu arbeiten statt erwerbslos zu sein, wenn das jüngste Kind ungefähr zehn Jahre alt ist, sogar größer ist als das entsprechende Odds der Kinderlosen. Eine Ad-hoc-Erklärung für dieses Umschlagen des Effekts von Kindern auf die Erwerbstätigkeit in Teilzeit ist, dass sich die Bilanz des zeitlichen Betreuungsaufwands und der finanziellen Kosten mit zunehmendem Alter des Kindes umkehrt. Bei Säuglingen und Kleinkindern kann man davon ausgehen, dass der Betreuungsaufwand zeitlich so groß ist, dass die relative Erwerbsarbeitsproduktivität sehr hoch sein muss, um die Betreuung an den Partner oder eine dritte Person abzugeben. Mit zunehmendem Alter sinkt der zeitliche Betreuungsaufwand,

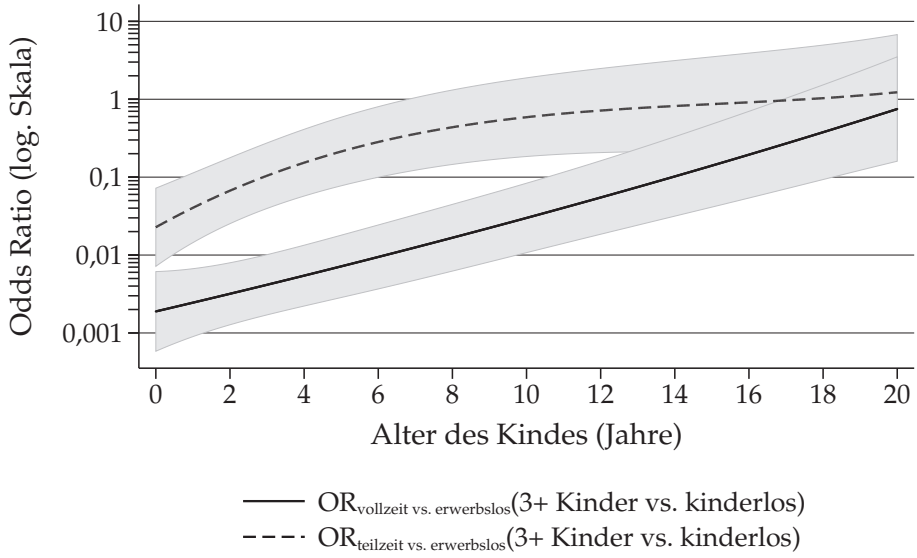


Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).



Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).

Abbildung 7.5: Odds Ratio von Vollzeit- und Teilzeiterwerbstätigkeit zwischen Frauen mit und ohne Kinder aus dem femlogit-Modell mit gekürzten Daten



Anmerkung: Konfidenzintervalle berechnet in Anlehnung an Brambor et al. (2006).

Abbildung 7.5: (Fortsetzung)

aber die finanziellen Kosten eines Kindes bleiben bestehen.[11] Da der Haushalt, also auch die Frau, die zusätzlichen finanziellen Kosten aufbringen muss, sollte der Nettobedarf für Erwerbsarbeit bei Haushalten mit Kindern höher sein als bei Haushalten ohne Kinder. Also wird mit zunehmendem Alter des Kindes Zeit „frei”, die auf Erwerbsarbeit umalloziiert wird. Da der zeitliche Betreuungsaufwand aber bis ins Jugendalter des Kinders bestehen bleibt, wird die Frau eher in Teilzeit arbeiten als in Vollzeit.

Neben dem Odds-Ratio zwischen Erwerbstätigkeit in Teilzeit und Erwerbslosigkeit ist in Abbildung 7.5 auch der Odds-Ratio-Effekt zwischen Vollzeiterwerbstätigkeit und Erwerbslosigkeit dargestellt. Hier erkennt man im Wesentlichen drei Dinge: Erstens ist das Odds-Ratio zwischen Vollzeiterwerbstätigkeit und Erwerbslosigkeit bei Frauen mit nur einem Kind gegenüber Kinderlosen betragsmäßig um knapp ein Fünftel schwächer als das entsprechende Odds-Ratio bei Frauen mit mehr als einem Kind gegenüber kinderlosen Frauen. Zweitens ist das Odds in Vollzeit zu arbeiten gegenüber erwerbslos zu sein für Mütter unabhängig von der Kinderzahl nicht mehr statistisch verschieden vom entsprechenden Odds für Kinderlose, wenn das jüngste Kind erwachsen ist. Drittens ist der Anstieg des Odds zwischen Vollzeiterwerbstätigkeit und Erwerbslosigkeit für Mütter gegenüber dem entsprechenden Odds bei Kinderlosen mit dem Alter des jüngsten Kindes unabhängig von der Kinderzahl in etwa linear. D. h. das Odds, in Vollzeit zu arbeiten statt nicht zu arbeiten, ist bei Müttern mit einem, zwei oder mehr als zwei Kindern nach

11 Die finanziellen Kosten können mit dem Lebensalter variieren. Man kann aber sicher annehmen, dass die Nettokosten erst dann auf Null oder darunter fallen, wenn das Kind finanziell unabhängig ist.

der Geburt 200 bis 1 000 mal kleiner als das entsprechende Odds von Kinderlosen. Die Odds unterscheiden sich aber nicht mehr signifikant voneinander, wenn das jüngste Kind 16–18 Jahre alt ist.

Wie bereits in Abschnitt 3.2.2 diskutiert wurde, ist man bei der Interpretation von binären und multinomialen logistischen Regressionen sehr stark eingeschränkt. Im Wesentlichen bleibt die Interpretation der Effekt auf die unanschaulichen Logits und Odds. Den Effekt auf die Odds haben wir in Anlehnung an Schröder (2010) oben ausführlich dargestellt und erläutert. Schröder unternimmt wegen der Unanschaulichkeit der Odds einen Versuch, die Effekte auf die *unbedingten* Wahrscheinlichkeiten zu interpretieren (s. o. Abschnitt 3.2.2). Hierzu nimmt sie konkrete Werte für die unbeobachtete Heterogenität an und simuliert für diese Zahlen die vorhergesagten Eintrittwahrscheinlichkeiten.

Im Folgenden soll diese Interpretationstechnik auf die Schätzer des femlogit-Modells angewendet werden, und die daraus folgenden Ergebnisse ausführlicher erläutert werden. Da Schröder nur dichotome Kontraste betrachtet, nimmt sie jeweils eine Heterogenität für jeden Kontrast an. Als Grundlage wählt sie eine prototypische Frau, und wählt für diese Frau die Heterogenität so, dass vorhergesagte Wahrscheinlichkeit der Erwerbstätigkeit bzw. Vollzeiterwerbstätigkeit bestimmte Werte annimmt. Als Prototyp wählt Schröder eine 40-jährige, verheiratete, kinderlose Frau, die einer Arbeitslosenquote von 8,4 % ausgesetzt ist. Schröder wählt dann in zwei Varianten die unbeobachtete „Neigung" zur Erwerbstätigkeit für so eine Frau so, dass sie mit einer Wahrscheinlichkeit von 99,5 % bzw. 60 % erwerbstätig ist. Bei der Vollzeiterwerbstätigkeit wählt sie die Heterogenität in drei Varianten so, dass die Wahrscheinlichkeit 60 %, 90 % und 98,5 % ist. Mit diesen Werten erzeugt sie Conditional-Effect-Plots über die vorhergesagten Wahrscheinlichkeiten in Abhängigkeit der Kinderzahl und des Alters des jüngsten Kindes. Aus Platzgründen werden diese Graphiken unter Verweis auf die Originalquelle hier nicht dargestellt.

Für das femlogit-Modell ist das Vorgehen grundsätzlich auch anwendbar. Wenn man für die Heterogenitäten für alle Kontraste bestimmte Werte annimmt, lassen sich entsprechend vorhergesagte Wahrscheinlichkeiten in Abhängigkeit von den beobachteten Variablen berechnen. Im vorliegenden Fall benötigen wir eine „Neigung" zur Erwerbstätigkeit in Vollzeit gegenüber der Erwerbslosigkeit und eine „Neigung" zur Teilzeiterwerbstätigkeit gegenüber der Erwerbslosigkeit. Da sich hier viele Variationsmöglichkeiten ergeben, beschränke ich mich auf einen prototypischen Fall, für den die Wahrscheinlichkeiten simuliert werden sollen.

Aus den Daten des ALLBUS 2000 (GESIS - Leibniz-Institut für Sozialwissenschaften 2003) wurden in Anlehnung an Schröder für das Jahr 2000 für alle westdeutschen kinderlosen Frauen im Alter von 15 bis 55 die Aggregatsanteile der Erwerbstätigkeit in Vollzeit und in Teilzeit und der Erwerbslosigkeit bestimmt. Diese bilden analog wie oben die Grundlage für die Heterogenitäten. Wir nehmen also eine Person an, die ohne Kinder mit der Wahrscheinlichkeit von ca. 0,68 vollzeiterwerbstätig ist, mit einer Wahrscheinlichkeit von ca. 0,09 teilzeiterwerbstätig ist und mit einer Wahrscheinlichkeit von ca. 0,22 nicht erwerbstätig ist.

Mit diesen Ausgangswahrscheinlichkeiten und den oben dargestellten Odds-Ratios lassen sich die vorhergesagten Wahrscheinlichkeiten in vollzeit und teilzeit oder nicht er-

werbstätig zu sein, die in Abbildung 7.6 dargestellt sind, bestimmen. In Teilabbildung (a) sind die vorhergesagten Wahrscheinlichkeiten für Frauen mit einem Kind, in Teilabbildung (b) die Wahrscheinlichkeiten für Frauen mit zwei Kindern und in Teilabbildung (c) die Wahrscheinlichkeiten für Frauen mit drei oder mehr Kindern dargestellt.

Man erkennt, dass die Effekte des Alters des jüngsten Kindes auf die Wahrscheinlichkeiten für Frauen mit einem Kind und zwei Kindern ungefähr gleich sind. Bei zwei Kindern sinkt die Wahrscheinlichkeit der Erwerbslosigkeit auch noch nach dem 14. Lebensjahr ab, während bei nur einem Kind die Wahrscheinlichkeit ab dem 14. Lebensjahr bei ca. 35 % stagniert.

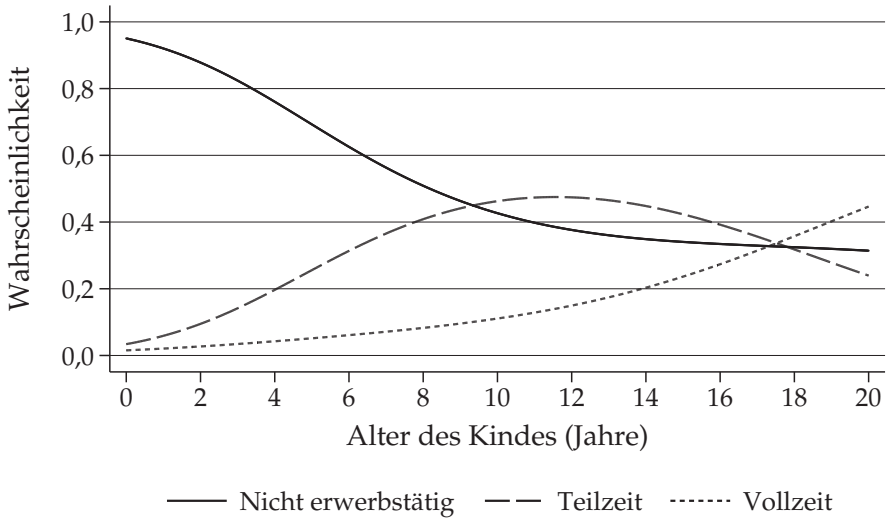
Das weitere Absinken der Wahrscheinlichkeit der Erwerbslosigkeit bei zwei Kindern geht vor allem mit einem stärkeren Anstieg der Wahrscheinlichkeit der Vollzeiterwerbstätigkeit einher. Bei mehr als zwei Kindern ist das Bild anders. Hier bleibt die Wahrscheinlichkeit nahezu unabhängig vom Alter des jüngsten Kindes auf dem niedrigen Niveau von höchstens 20 %. Dagegen bleibt die Wahrscheinlichkeit nicht erwerbstätig zu sein, bis das jüngste Kind 16 Jahre alt ist, bei über 50 %.

Bei den Abbildungen 7.6 sind bewusst keine Konfidenzintervalle dargestellt, da diese mit dem Ansatz von Schröder (2010) nicht bestimmbar sind. Durch Festsetzen der unbeobachteten Heterogenität auf einen bestimmten Wert können zwar für alle Ausprägungen der beobachteten Variablen vorhergesagte Wahrscheinlichkeiten berechnet werden. Für die Konfidenzintervalle müssen aber zusätzlich Korrelationen zwischen der Heterogenität und allen unabhängigen Variablen angenommen werden. Für eine sinnvolle Auswahl von konkreten Werten für diese Korrelationen fehlt aber jede Grundlage.
Außerdem können auch die einzelnen Plots über die vorhergesagten Auswahlwahrscheinlichkeit nur begrenzt sinnvoll interpretiert werden. So kann man zwar eine Festlegung der Wahrscheinlichkeiten, die für die Bestimmung der Heterogenitäten gewählt werden, ausreichend begründen. Da wir aber annehmen, dass die Heterogenität mit den unabhängigen Variablen in unbestimmter Weise korreliert ist, ist auch die bedingte Verteilung der unabhängigen Variable für einen bestimmten Wert unbekannt.

Insgesamt stellt daher der Interpretationsansatz von Schröder (2010) keine überzeugende Alternative gegenüber den bereits angesprochenen unanschaulichen Interpretationsformen dar.

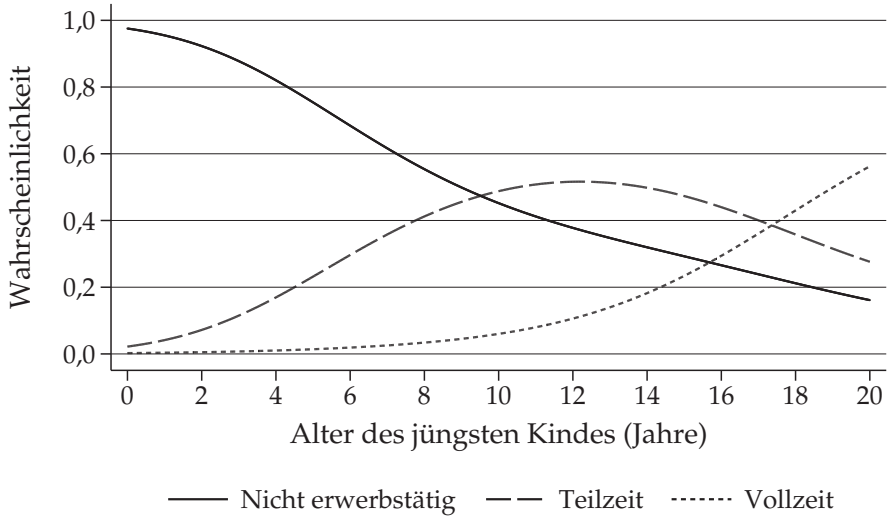
7.4 Zusammenfassung

In diesem Kapitel haben wir das im ersten Teil der Arbeit abgeleitete und implementierte femlogit-Modell angewendet, und die Analysen von Schröder (2010) zum Effekt der Fertilität auf die Erwerbstätigkeit der Frau zu replizieren. Da Schröder die Analysedaten zur Reanalyse zur Verfügung stellt, ist zunächst einmal eine exakte Replikation der Analysen von Schröder möglich. Wir finden in den dichotomen logistischen Regressionen mit fixed-effects einen starken negativen Effekt der Kinderzahl auf das Odds, erwerbstätig statt nicht erwerbstätig und auf das Odds, vollzeiterwerbstätig statt erwerbstätig in Teilzeit oder erwerbslos zu sein. Weiterhin finden wir einen positiven Effekt des Alters des



Anmerkung: Pr(vollzeit | kinderlos)=0,68 Pr(teilzeit | kinderlos)=0,09 Pr(nicht erwerbstätig | kinderlos)=0,22

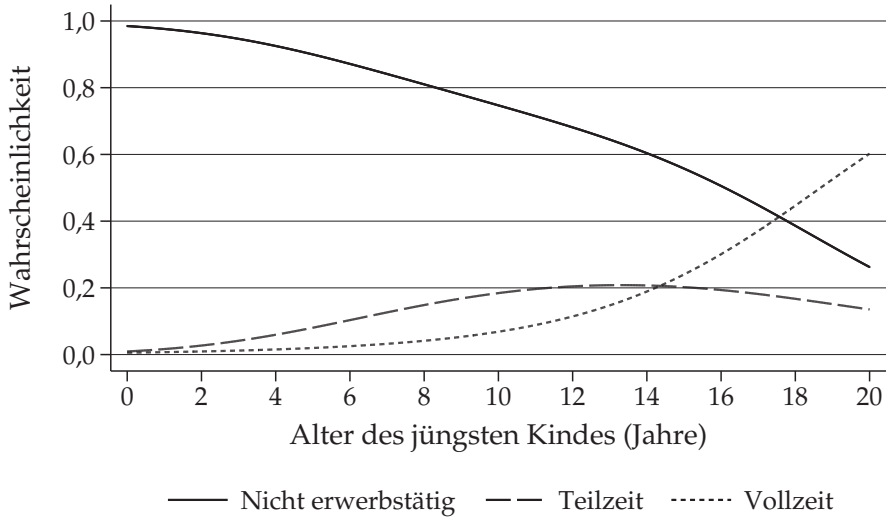
(a) Frauen mit einem Kind



Anmerkung: Pr(vollzeit | kinderlos)=0,68 Pr(teilzeit | kinderlos)=0,09 Pr(nicht erwerbstätig | kinderlos)=0,22

(b) Frauen mit zwei Kindern

Abbildung 7.6: Vorhergesagte Wahrscheinlichkeiten aus dem femlogit-Modell mit gekürzten Daten



Anmerkung: Pr(vollzeit | kinderlos)=0,68 Pr(teilzeit | kinderlos)=0,09
Pr(nicht erwerbstätig | kinderlos)=0,22

(a) Frauen mit mehr als zwei Kindern

Abbildung 7.6: (Fortsetzung)

jüngsten Kindes auf das Odds, erwerbstätig statt nicht erwerbstätig und auf das Odds, vollzeiterwerbstätig statt erwerbstätig in Teilzeit oder erwerbslos zu sein.

Im zweiten Schritt werden die Analysen von Schröder durch die Verwendung des femlogit-Modells dahingehend ergänzt, dass nun die Kontraste zwischen Erwerbstätigkeit in Vollzeit, Erwerbstätigkeit in Teilzeit und Erwerbslosigkeit unter Berücksichtigung von personenspezifischen fixed-effects gemeinsam analysiert werden. Für die Anwendung des femlogit-Modells muss zunächst die Anzahl der Messzeitpunkte, die Schröder verwendet, reduziert werden, um die Rechenzeit in einem handhabbaren Rahmen zu halten. Statt zwei Messzeitpunkte pro Jahr verwenden wir einen Messzeitpunkt alle drei Jahre.

Auf den ersten Blick zeigen die Schätzer des femlogit-Modells kein anderes Bild gegenüber den dichotomen fixed-effects-Modellen. Bei genauerer Betrachtung erkennt man mit dem multinomialen Modell gegenüber den Befunden aus den dichotomen Modellen, dass bei Müttern von einem oder zwei Kindern im Vergleich zu Kinderlosen mit steigendem Alter des jüngsten Kindes das zunächst stark reduzierte Odds der Teilzeiterwerbstätigkeit gegenüber Erwerbslosigkeit nicht nur wieder auf das Niveau der kinderlosen Frauen ansteigt. Wenn das jüngste Kind ungefähr zehn Jahre alt ist, ist das Odds, in Teilzeit statt nicht erwerbstätig zu sein, für Mütter von ein oder zwei Kindern sogar größer als das von Frauen ohne Kinder. Insoweit ergibt sich aus der Verwendung des femlogit-Modells ein

differenziertes Bild des Effekts der Kinderzahl und des Kinderalters, dass mit den aus der Theorie abgeleiteten Erwartung übereinstimmt.

Natürlich sind mit dem femlogit-Modell nicht alle Probleme bei der Untersuchung des Effekts der Fertilität auf die weibliche Erwerbstätigkeit gelöst. Zunächst einmal liegen beim binären wie multinomialen logit-Modell mit fixed-effects verschiedene Problem vor. Erstens muss bei diesen Modellen serielle Korrelation über die Messzeitpunkte und Korrelationen der Fehlerterme über die Alternativen (IIA) und per Annahme ausgeschlossen werden. Beide Annahmen kann man als in der Regel verletzt betrachten. Weiterhin ist bei den Ergebnisse keine Interpretation auf Ebene der unbedingten Wahrscheinlichkeiten möglich, so dass man auf die unanschaulicheren Odds-Effekte angewiesen ist.

Das Problem der Korrelation über die Messzeitpunkte und Alternativen lässt sich bei Beibehaltung des Modells nur dadurch abmildern, dass man Regressoren aufnimmt, die die „Ähnlichkeiten" in den Fehlertermen über die Messzeitpunkte und Alternativen hinweg abbilden. Bei der seriellen Korrelation ist die Aufnahme von Perioden-Effekt-Dummies ein Standardvorgehen, aber auch die Aufnahme von „Lags" und „Leads" der im Modell enthaltenen Kovariate kann das Problem der seriellen Korrelation abschwächen. [12] Bei Verletzung der IIA-Annahme kann man ebenso versuchen, durch Aufnahme alternativen-spezifischer Variablen die Unterschiede hinreichend gut in die beobachteten Variablen abzubilden. Beide Strategien scheitern, wenn die Ursache für die Korrelation über die Messzeitpunkte und die Alternative in unbeobachtbaren Variablen liegt.

Für das Problem der unanschaulichen Interpretierbarkeit gibt es nach den gegenwärtigen Kenntnisstand keine befriedigende Lösung. Für die unbeobachtete Heterogenität α_i liegt nur die suffiziente Statistik $\sum_t y_{it}$ vor, aber darüber hinaus lässt sich weder eine Aussage über die Verteilung von α treffen. D. h. eine Wahl eines bestimmten Wert für α_i ist willkürlich. Außerdem sind die Korrelationen zwischen einem so bestimmten α_i und den übrigen Kovariaten unbestimmt, was dazu führt, dass man keine Standardfehler berechnen kann.

Sowohl das Problem der Korrelation über die Messzeit und die Alternativen als auch die unanschauliche Interpretation kann stark reduziert werden, wenn man sich von den logistischen Regressionsmodellen löst, und dafür an anderer Stelle Abstriche macht. Beim felogit- und femlogit-Modell ist die Verteilung der Heterogenität bedingt auf die interessierenden Kovariate $f_{\alpha|X}$ völlig unbeschränkt. Das „correlated random effects"-Modell nach Mundlak (1978); Chamberlain (1980) beschränkt die Verteilung $f_{\alpha|X}$ so, dass zwar Korrelation zwischen den Kovariaten und der Heterogenität möglich ist, aber dennoch tiefere Aussagen als beim felogit- bzw. femlogit-Modell möglich sind. Mit diesem Modell kann man erstens Korrelationen zwischen den Messzeitpunkten und den Alternativen zulassen, und zweitens Effekte auf die Wahrscheinlichkeit interpretieren. (vgl. Wooldridge 2010, S. 615–617, 623).

12 Durch die Aufnahme von „Lags" und „Leads" reduziert sich die effektive Anzahl der Messzeitpunkte, was zu ineffizienteren Schätzern führt. Wenn man drei Messzeitpunkte hat und für eine Kovariate x den vorherigen Messzeitpunkt als Lag kontrolliert, bleiben effektiv zwei Messzeitpunkte: Statt $x_{\{t=1,2,3\}}$ hat man $(x_t, x_{t-1})_{\{t=2,3\}}$. Verwendet man zusätzlich den Folgemesszeitpunkt als Lead, bleibt nur ein effektiver Messzeitpunkt: $(x_t, x_{t-1}, x_{t+1})_{\{t=2\}}$.

Neben diesen beiden Problem liegt beim vorliegenden Untersuchungsgegenstand das Problem vor, dass die abhängige Variable auf die unabhängige Variable wirken kann, d. h. Endogenität im engen Sinn vorliegt. Hierfür schafft weder sowohl das fixed-effects- noch das correlated-random-effects-Modell Abhilfe. Wenn man annimmt, dass man die wechselseitige Abhängigkeit durch beobachtbare Variablen vollständig abbilden kann, könnte man dynamische Modelle wie bei Hyslop (1999) verwenden. Eine zweite Alternative ist der Instrument-Variablen-Ansatz wie bei z. B. bei Angrist und Evans (1998). Hier hängt der Erfolg stark davon ab, ob erstens das Instrument wirklich exogen ist und zweitens das Instrument trotzdem einen Effekt auf das Treatment hat.

8 Einfluss der Klassenzugehörigkeit auf die Parteipräferenz

In diesem Kapitel wird die Replikation der Arbeit von Kohler (2002, Kap. 6), bzw. Kohler (2005) zum Zusammenhang zwischen der sozialen Lage und der Parteipräferenz dargestellt. Kohler untersucht in der Arbeit von 2002 den Effekt der sozialen Lage, insbesondere der sozialen Klassenzugehörigkeit, auf die Parteipräferenz in Westdeutschland. Den Kern der Arbeit bildet die Analyse des Effekts verschiedener Wechsel soziostruktureller Merkmale über die Zeit auf die langfristige Parteipräferenz für die SPD, die CSU/CDU und FDP, und die Grünen mit den Daten des SOEP 1984–1997. Die zentralen unabhängigen Variablen sind Wechsel in der sozialen Klasse, Übergang in die Arbeitslosigkeit und ins Erwerbsleben, Abschluss eines weiteren Bildungsabschlusses, Hochzeit und Geburt von Kindern. Hierbei verwendet Kohler binäre logistische Regressionsmodelle mit fixed-effects. In der Arbeit von 2005 erweitert Kohler diese Analysen um die SOEP-Wellen bis 2000.

Im Folgenden wird zunächst die zugrundeliegende Theorie dargestellt. Da die betrachtete Arbeit von Kohler immer noch die einzige ist, die diesen Zusammenhang mit Längsschnittdaten und adäquaten Längsschnittsmodellen untersucht, fällt der Überblick über vergleichbare empirische Ergebnisse sehr kurz aus. Danach wird die Replikation des zentralen Modells von Kohler (2002) dargestellt. Dieses wird anschließend dahingehend erweitert, dass erstens auf die verwendeten Daten das femlogit-Modell angewendet wird, und zweitens das Modell um die mittlerweile verfügbaren SOEP-Wellen erweitert wird. Abschließend werden die neuen Ergebnisse aus dem femlogit-Modell und dem aktualisierten Datenbestand, und bestehende Defizite und mögliche Verbesserungen des Modells diskutiert.

8.1 Theorie

Kohler (2002) entwickelt für seine Untersuchung ein vereinheitlichendes Modell über die Parteipräferenz, in dem er versucht, die drei großen relevanten Theorieströnge zu integrieren. Diese sind der Interaktionsansatz aufbauend auf Lazarsfeld, Berelson, und Gaudet (1944), der Identifikationsansatz ausgehend von Campbell, Converse, E. Miller, und Stokes (1960) und die Cleavagetheorie von Lipset und Rokkan (1967). Den vereinheitlichenden Rahmen bildet Kohler mit einem Rational-Choice-Modell.

Der Ausgangspunkt für Kohlers Modell ist die Cleavagetheorie. Potentielle Wähler sind entlang sozialer Konfliktlinien geteilt. Kohler sieht für seine Fragestellung im Wesentlichen den Konflikt zwischen Arbeit und Kapital als zentral an. Daraus leitet er ab, dass bestimmte Klassenpositionen nach Erikson und Goldthorpe (1992) von bestimmten Parteien vertreten werden sollten. Kohler unterscheidet hier sechs Gruppen, mit teilweise gleichen politischen Interessenlagen, die im Weiteren noch ausführlicher erläutert werden.

Die Beiträge des Interaktionsansatzes und des Identifikationsansatzes sind im Modell von Kohler über die Cleavagetheorie nachrangig. Aus dem Interaktionsansatz folgt, dass die unmittelbaren Kontaktgruppen eines potentiellen Wählers durch die sozial differenzierte Gelegenheitsstruktur im Meeting-Mating-Prozess eine ähnliche Sozialstruktur aufweisen wie der Wähler selbst. Durch die wechselseitige Beeinflussung führt dies zu einer

Bestätigung und Verstärkung der Parteipräferenz. Für seine konkreten Analysen leitet Kohler hieraus aber nur ab, dass die Parteienpräferenz der Ehepartners einen Einfluss auf die eigene Parteienpräferenz hat.[1] Aus dem Identifikationsansatz kann Kohler nur ableiten, dass die Betrachtung der langfristigen Parteiidentifikation, wie sie im SOEP gemessen ist, als abhängige Variable zulässig ist. Nach diesem Ansatz sollte – kurz gesagt – der Wähler die Partei wählen, bei der die Bilanz kurzfristiger Kandidaten- und Programm-Faktoren und der langfristiger Parteiidentifikation gegenüber den anderen Parteien maximal ist. D. h. kurzfristig können Eigenschaften eines Kandidaten oder eines Wahlprogramms dazu führen, dass der Wähler kurzfristig eine Partei präferiert, die nach der Cleavagetheorie nicht seine Interessen vertritt.

Wie oben erwähnt, unterscheidet Kohler sechs Gruppen, die sich idealtypisch entlang der Konfliktlinie zwischen Arbeit und Kapital gegenüberstehen. Diese Gruppen ergeben sich im Wesentlichen aus dem Berufsklassenschema von Erikson und Goldthorpe (1992) und deren Erweiterung von Müller (1998): Selbstständige, Angehörige der administrativen Dienstklasse, Arbeiter, Beschäftige in den „soziale Diensten", Experten und Mischtypen. Arbeiter werden von der SPD vertreten, dagegen werden Selbständige und Angehörige der administrativen Dienstklasse[2] von der CDU/CSU und FDP vertreten. Kohler argumentiert weiter, dass die Gruppe der „Experten im sozialen und künstlerischen Bereich"[3], als Teilgruppe der Experten innerhalb der Dienstklasse, eher von der SPD und den Grünen vertreten werden sollten. Als Residualgruppe bildet Kohler die Gruppe der übrigen Experten und nicht eindeutig zuordnungsbaren Mischtypen.[4] Aufgrund der gleichen Interessenlage fasst Kohler die Gruppe der Arbeiter und die Gruppe der Sozialen Dienste zusammen.

Der Einfluss der sozialen Klasse auf die Parteipräferenz impliziert fast ausschließlich eine Präferenz für das linke oder rechte politische Lager. Darüber hinaus argumentiert Kohler, dass es neben dem Anreiz zu Macht und Wohlstand, die im Wesentlichen den Einfluss der sozialen Klassenlage auf die Präferenz des linken oder rechten politischen Lagers bestimmen, einen unabhängigen Anreiz zur Selbstverwirklichung gibt. Kohler stützt sich hierbei auf die Arbeiten von Inglehart (1977) und Schulze (1992). Er argumentiert davon ausgehend, dass mit steigender Bildung der Drang zur Selbstverwirklichung beim Wähler steigt, und dass weiterhin die Partei der Grünen diesem Drang eher gerecht wird.

Insgesamt leitet Kohler folgende Hypothese über den Effekt der Klassenzugehörigkeit auf die Parteipräferenz ab (vgl. Kohler 2002, S. 103):

---

1 Eine Überprüfung von differenzierten Hypothesen aus dem Interaktionsansatz sind mit den Daten des SOEP vor allem deshalb nicht möglich, da der Datensatz keine weitere Information über das übrige soziale Netzwerk enthält.

2 Mit Angehörigen der administrativen Dienstklasse sind umgangssprachlich prototypisch Manager gemeint, d. h. leitende Verwaltungsangestellte in der öffentlichen Dienst und leitende Angestellte und Geschäftsführer in der Privatwirtschaft.

3 Die sogenannten Experten im sozialen und künstlerischen Bereich werden auch als „Soziale Dienste" bezeichnet. Hiermit sind Mediziner, Angestellte im Gesundheitsweisen, Wissenschaftler aus den Geistes- und Gesellschaftswissenschaften, Lehrer, Schriftsteller, Künstler und Kirchenangestellte gemeint. Vgl. Müller (1998).

4 Die übrigen Experten sind Naturwissenschaftler, Mathematiker, Ingenieure, Architekten und Juristen. In die Gruppe der Mischtypen fallen im Wesentlichen nicht-manuelle Routineberufe, Heimberufe und Vorarbeiter. Vgl. Kohler (2002, S. 164).

- Arbeiter und Beschäftigte in den Sozialen Diensten haben eine sehr starke Neigung, die SPD zu wählen, eine schwächere Neigung, die Grünen zu wählen, und eine Abneigung gegen die CDU und die FDP.

- Selbstständige haben eine sehr starke Neigung, die CDU oder die FDP zu wählen, und eine starke Abneigung gegen die SDP und die Grünen.

- Angehörige der administrativen Dienstklasse haben eine stärkere Präferenz für die SPD als für die Grünen.

- Experten und Mischtypen haben eine moderate Präferenz für die SPD, eine schwache Präferenz für die Grünen und eine moderate Abneigung gegen die CDU und die FDP.

- Mit steigender Bildung steigt die Präferenz für die Grünen.

8.2 Forschungsstand

Die Menge der Arbeiten, die Effekt der Klassenzugehörigkeit auf die Parteipräferenz untersuchen, ist unüberschaubar. Daher soll an dieser Stelle auch kein Versuch unternommen werden, einen Überblick über diese Arbeiten zu geben. Kohler (2002) leistet bei der Untersuchung dieses Effekts einen entscheidenden Beitrag dadurch, dass er mit der Verwendung des dichotomen logistischen Regressionsmodells mit fixed-effects in der Lage ist, für zeitlich konstante Effekte zu kontrollieren. Nach dem Kenntnisstand des Autors ist die Arbeit von Kohler – abgesehen von seiner eigenen Replikation von 2005 – immer noch die einzige Arbeit, die für die Erklärung der Parteipräferenz bzw. der Wahlentscheidung ein fixed-effects-Modell verwendet. Für den deutschsprachigen Raum findet man zwar aktuelle Arbeiten wie Pappi und Brandenburg (2010) und Neundorf, Stegmüller, und Scotto (2011), die random-effects-Modelle auf Längsschnittdaten und im Mehrebenenkontext verwenden, aber fixed-effects-Modelle wurden bisher nicht verwendet. Dies lässt sich damit begründen, dass gerade in der deutschsprachigen Politikwissenschaft erstens „die Zahl der auf Deutschland bezogenen Studien, die Paneldaten genutzt haben, sehr begrenzt" ist (s. Steinbrecher 2009, S. 97–100). Weiterhin findet man bei Steinbrecher an derselben Stelle, dass die zwei zentralen Vorteile von Paneldaten die Analyse von intraindividuellen Veränderungen und die Erfassung der zeitlichen Reihenfolge sei, und eben nicht die Möglichkeit, durch die Verwendung von fixed-effects-Modellen, für unbeobachtete Drittvariable implizit kontrollieren zu können.

Die Arbeiten im englischsprachigen Raum sind zwar schwieriger zu überschauen, aber dem Autor ist keine Arbeit bekannt, die für die Untersuchung der Wahlentscheidung bzw. der Parteipräferenz mit Paneldaten fixed-effects-Modelle nutzt. In der amerikanischen Politikwissenschaft sind bei Mehrebenenanalysen zu anderen Fragestellungen vereinzelte Anwendungen von fixed-effects-Modellen zu finden.

8.3 Analysen

Im Weiteren soll nun eine Auswahl der Analysen von Kohler (2002) zum Effekt der Klassenzugehörigkeit auf die Parteipräferenz repliziert werden, und um eine gemeinsame Analyse über alle Parteien mit dem femlogit-Modell über alle verfügbaren Wellen bis 2010 erweitert werden.

Kohler (2002) schätzt für die in seiner Arbeit im Mittelpunkt stehenden Analysen binäre logistische Regressionsmodelle mit fixed-effects. Die einzelnen Modelle werden in drei Varianten jeweils auf die Parteipräferenz für die CDU oder FDP, für die SPD und für die Grünen gerechnet. In der ersten Variante berücksichtigt er als unabhängige Variablen, wie oben beschrieben, die Klassenzugehörigkeit, die Erwerbstätigkeit, Ausbildungsabschlüsse, Eintritt ins Erwerbsleben, Heirat und die Geburt von Kindern. In der zweiten Variante erweitert er das Modell dahingehend, dass er für alle vorherigen Kovariate eine Interaktion mit politischem Interesse modelliert. Im dritten Modell berücksichtigt er dann zusätzlich eine zeitverzögerte Wirkung aller Kovariaten. Da bei der Replikation die Anwendung des femlogit-Modells im Mittelpunkt steht, beschränken wir uns auf die einfachste, erste Variante der Modelle von Kohler. Aus inhaltlicher Perspektive ist zu beachten, dass Kohler gerade die Befunde der komplexeren Modelle im Fazit seiner Arbeit herausstellt. Die Interpretation, dass der Effekt der Klassenzugehörigkeit im einfachsten Modell „mit vollständiger Information ohne zeitverzögerte Wirkung" nicht nachzuweisen ist, teilen wir nicht. Die Effekte sind zwar schwächer, es gelingt Kohler auch mit dem einfachsten Modell, den Effekt der Klassenzugehörigkeit auf die Parteipräferenz unter Kontrolle von zeitlich konstanten individuellen Heterogenitäten zu schätzen.

8.3.1 Daten

Wie bereits erwähnt, verwendet Kohler (2002) für seine Analysen die Daten des Sozio-oekonomischen Panels (SOEP) 1984–1997. Das SOEP ist eine Längsschnittuntersuchung über Privathaushalte, in der seit 1984 jährlich alle Personen ab 16 Jahren im jeweiligen Haushalt befragt werden. Die Studie ist inhaltlich breit angelegt, so dass über alle Personen im Haushalt verschiedene sozio-strukturelle Merkmale erfasst werden (vgl. die ausführliche Darstellung bei Haisken-De New und Frick 1998). Die Replikation verwendet dieselbe Datengrundlage wie die Originalquelle. Danach wird die Analyse von Kohler aktualisiert und der aktuelle Datensatz SOEP v27 verwendet (vgl. Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2010, Version 27, SOEP 2011; Wagner, Frick, und Schupp 2007).

Analysestichprobe

Im SOEP laufen verschiedene Ausgangsstichproben parallel nebeneinander (Details wieder in Haisken-De New und Frick 1998 oder Wagner et al. 2007). Kohler (2002) verwendet die bis 1997 eingeführten vier Stichproben. Er schliesst aus dem Gesamtdatensatz zuerst Messzeitpunkte mit fehlenden Werten aus und dann Zeitreihen, bei denen weniger als zwei Messzeitpunkte übrigbleiben. Die besondere Beschränkung auf Zeitreihen mit min-

destens drei Beobachtungen ergibt sich aus der besonderen Kodierung der unabhängigen Variablen.[5]

Wie kurz angesprochen, analysiert Kohler für jede der drei Modellvarianten jeweils den Kontrast zwischen der SPD und den anderen Parteien, der CDU und der FDP gegenüber den anderen Parteien und den Grünen und den anderen Parteien. Die Restgruppe enthält jeweils auch „sonstige Parteien" und keine Parteineigung. Zu den oben dargestellten Einschränkungen kommt dann hinzu, dass für jede Beobachtungseinheit bei jedem Kontrast mindestens ein Wechsel von einem zum anderen Pol des Kontrastes vorliegen muss. Hieraus folgt, dass Kohler (2002) für das einfachste Modell für den Kontrast zwischen der SPD und den anderen Parteien 5 680 Zeitreihen mit durchschnittlich 8,7 Messzeitpunkten übrigbleiben. Für den Kontrast zwischen CDU und FDP und den anderen Parteien bleiben 4 842 Beobachtungen mit 8,6 Messungen und für den Kontrast zwischen den Grünen und den anderen Parteien 1 666 Beobachtungseinheiten mit 8,3 Messungen (s. Kohler 2002, S. 259).

Kohler verwendet bei seinen Analysen einen Gewichtungsfaktor, der erstens die unterschiedlichen Auswahlwahrscheinlichkeiten in den einzelnen Teilstichproben berücksichtigt. Zweitens soll der Gewichtungsfaktor die unterschiedlichen Attritionwahrscheinlichkeiten so berücksichtigen, dass der Faktor in etwa einem Querschnitts-Poststratifikations-Gewicht entspricht (vgl. Lynn 2005). Neben dem Gewicht verwendet Kohler ein Verfahren, dass für die Berechnung der Standardfehler der Schätzer die Schichtung der SOEP-Stichprobe berücksichtigt (vgl. Kohler 2002, S. 142-144).

Die erste Replikation versucht, die einfachsten Analysen von Kohler (2002) möglichst exakt nachzubilden. Beim DIW konnte der Originaldatensatz von 1998 beschafft werden, so dass die Ausgangsdaten identisch sein sollten. Da Kohler alle Syntaxdateien von der Datenaufbereitung über die Analyse bis zur Darstellung veröffentlicht,[6] konnte die Aufbereitung ebenso nah am Original nachvollzogen werden. Trotz der Transparenz der Aufbereitung gelingt es nicht, die Analyse des einfachsten Modells „mit vollständiger Information ohne zeitverzögerte Wirkung" zu replizieren. Schon die Fallzahl der Analysestichprobe weicht ab. Nach Ablauf der Aufbereitungssyntaxdateien bleiben für den Kontrast zwischen SPD und den anderen Parteien 5 750 Beobachtungen mit durchschnittlich 8,6 Messzeitpunkten. Für den Kontrast zwischen CDU und FDP und den anderen Parteien bleiben 4 926 Beobachtungen mit 8,5 Zeitpunkten und beim Kontrast zwischen den Grünen und den anderen Parteien bleiben 1 696 Beobachtungen mit 8,2 Zeitpunkten. Bei der Replikation wird im Zuge der Datenaufbereitung auch das Gewicht, wie es Kohler konstruiert, mitgebildet und verwendet. Das Verfahren zur Korrektur der Standardfehler, wie es Kohler verwendet, wird nicht angewendet, da das veröffentliche Ado `rgroup` nicht nachvollziehbare und behebbare Fehlermeldungen erzeugte.

Die Analysen von Kohler (2002) mit den Daten von 1984–1997 werden zunächst dadurch erweitert, dass die drei Modelle über die jeweiligen Kontraste mit dem femlogit-Modell in ein Modell vereinigt werden. Auch hierfür wird die Aufbereitungs-

5 Kohler (2005) erweitert seine Analyse auf die 1998 und 2000 hinzugefügten Stichproben. Außerdem beschränkt er seine Analyse auf Befragte, die in den Ausgangsstichproben in Westdeutschland gelebt haben.

6 Siehe die Dokumentation bei `http://www.wzb.eu/de/personen/ulrich-kohler?s=12918`.

syntax von Kohler (2002) verwendet. Die Analysestichprobe für dieses Modell enthält 10 387 Beobachtungseinheiten mit durchschnittlich 8,2 Messzeitpunkten.

Diese Analysen werden dann durch die SOEP-Daten von 1984–2010 aktualisiert. Hierfür werden zunächst die drei Modelle über die drei Kontraste aktualisiert, und im zweiten Schritt das gemeinsame femlogit-Modell. Die Datenaufbereitung folgt der aktualisierten Arbeit von Kohler (2005). Hier verwendet Kohler die SOEP-Daten von 1984–2000. Auch für diese Arbeit ist die Datenaufbereitung vollständig transparent und öffentlich zugänglich. Für die Daten von 1984–2010 kann die Aufbereitung bis auf einzelne Anpassungen fast vollständig übernommen werden. Nach der Aufbereitung bleiben für das aktualisierte Modell über den Kontrast zwischen der SPD und den anderen Parteien 11 796 Beobachtungen mit im Mittel 12,7 Messzeitpunkten. Für den Kontrast zwischen CDU/CSU und FDP und den anderen Parteien bleiben 11 094 Beobachtungen mit durchschnittlichen 12,3 Zeitpunkten und für den Kontrast zwischen den Grünen und den übrigen Parteien bleiben 3 575 Beobachtungen mit 12,6 Zeitpunkten.

Zuletzt werden die aktualisierten Modelle über die drei Kontraste mit dem femlogit-Modell vereinigt. Ähnlich wie bei der Replikation der Analysen von Schröder (2010) oben in Abschnitt 7.3.2 tritt auch hier das Problem auf, dass durch die hohe Anzahl der Messzeitpunkte im vollständigen Datensatz zu sehr lange Rechenzeiten führt. Daher wird auch hier der Datensatz gekürzt, indem nur die ungeraden Jahre, also 1985, 1987, usw. analysiert werden. So bleiben nach der angepassten Aufbereitungssyntax von Kohler (2005) und der Kürzung 22 195 Beobachtungseinheiten mit durchschnittlich 5,8 Messzeitpunkten.

Operationalisierung

Die Operationalisierung der Analysevariablen ist vollständig transparent aus den veröffentlichten Aufbereitungssyntaxdateien nachvollziehbar. Die abhängige Variable ist bei allen hier dargestellten Modellen allgemein gesprochen die Parteipräferenz. Sie wird aus der Frage nach der langfristigen Parteipräferenz gebildet: „Viele Leute in der Bundesrepublik neigen längere Zeit einer bestimmten Partei zu, obwohl Sie auch ab und zu eine andere Partei wählen. Wie ist das bei Ihnen: Neigen Sie einer bestimmten Partei in Deutschland zu? Falls ‚ja', welcher Partei?" Diese Frage ist über alle SOEP-Wellen hinweg gleich abgefragt worden, wobei sich die Kodierung der Antworten über die Wellen hinweg leicht, jedoch unproblematisch, verändert. Die exakte Formulierung ist insoweit von Bedeutung, als es sich hier um die typische Abfrage der langfristigen Parteiidentifikation nach Frank Dishaw handelt (vgl. Kaase und Klingemann 1994, S. 346f.) und nicht, wie bei der Darstellung des theoretischen Modells bereits erwähnt wurde, um eine Frage nach der Parteipräferenz.

Grundsätzlich werden für alle Analysen die Angabe, keine Präferenz für irgendeine Partei zu haben, als getrennte Antwortkategorie kodiert. Weiterhin wird die Präferenz für die CDU und die CSU zusammengefasst, und schließlich werden alle anderen Angaben als eine eindeutige Präferenz für die SPD, CDU/CSU, FDP oder die Grünen als eine Präferenz für eine andere Partei kodiert. Für die dichotomen Analysen werden aus dieser

kategorialen Variable drei Kontraste gebildet: (1) SPD vs. {keine Partei, CDU/CSU, FDP, Grüne, andere Partei}, (2) {CDU/CSU, FDP} vs. {keine Partei, SPD, Grüne, andere Partei}, (3) Grüne vs. {keine Partei, SPD, CDU/CSU, FDP, andere Partei}.[7] Bei den gemeinsamen Analysen mit dem femlogit-Modell stellen die sechs Möglichkeiten getrennte Alternativen dar, d. h. die Alternativenmengen ist {keine Partei, SPD, CDU/CSU, FDP, Grüne, andere Partei}.

Die besondere Operationalisierung der unabhängigen Variablen nimmt in der Arbeit von Kohler einen wichtigen Stellenwert ein. Da diese natürlich auch in der Replikation und die Aktualisierung übernommen wird, wird die Operationalisierung hier etwas ausführlicher in zwei Teilen erläutert. Zunächst werden die inhaltlichen Bestandteile der Analysevariable erläutert und im zweiten Teil wird die besondere Kodierung der unabhängigen Variable als Ereignisse dargestellt.

Die zentrale unabhängige Variable ist die Klassenzugehörigkeit. Bei Kohler (2002) werden zunächst aus dem in den SOEP-Daten von 1984–1997 verfügbaren Dreisteller-ISCO68 und der beruflichen Stellung das Klassenschema nach Erikson und Goldthorpe (1992) nachgebildet. Aus diesem selbst kodierten Klassenschema werden unter Bezugnahme auf Müller (1998) unter Berücksichtigung der beruflichen Stellung sechs Gruppen der sozialen Klassenlage gebildet: Selbstständige, Angehörige der administrativen Dienstklasse, Arbeiter, Beschäftige in den „Soziale Diensten", Experten und Mischtypen. Von diesen sechs Gruppen fasst Kohler erstens die Gruppe der Mischtypen und der Experten und zweitens die Gruppe der Arbeiter und der Beschäftigten in den „Soziale Diensten" zusammen (vgl. die Erläuterungen oben). Diese Aufbereitung wird für die Replikation und das integrierende femlogit-Modell über die SOEP-Daten von 1984–1997 direkt übernommen.

Für die aktualisierten Analysen mit den SOEP-Daten 1984–2010 wird die Aufbereitung von Kohler (2005) herangezogen. Hier bildet Kohler die Erikson-Goldthorpe-Klassifikation nicht selbst, sondern verwendet die im Datensatz bereitgestellte Klassifikation. Daraus erstellt er mit Hilfe des in diesen Daten verfügbaren Dreisteller-ISCO68 wie oben die sechs Gruppen über die soziale Klasse, und fasst diese wieder zu vier Gruppen zusammen. Bei der Übertragung dieser Aufbereitung auf die aktuellen Daten müssen verschiedene Anpassungen vorgenommen werden. Erstens fehlt in den aktuellen Daten der ISCO-Code des vorherigen Berufs für die erste Panelwelle von 1984. Da diese fehlende Angabe nur eine geringfügige Bedeutung hat, wird darauf verzichtet, so dass die Differenz durch den Periodeneffekt ausgeglichen werden muss. Zweitens geht die Aufbereitung von Kohler (2005) vom Dreisteller-ISCO68-Code aus, der in den aktuellen Daten durch den Viersteller-ISCO88-Code ersetzt wurde. Daher wurde die Aufbereitung der sechs Gruppen entsprechend angepasst, wobei die ausführliche Beschreibung von Müller (1998) herangezogen wurde.

Die übrigen unabhängigen Variablen sind von nachrangiger Bedeutung. Dennoch soll hier kurz auf die Operationalisierung eingegangen werden. Kohler kontrolliert zunächst neben dem eigentlichen Einfluss der Klassenlage für den Effekt der Arbeitslosigkeit. Diese wird sowohl in der Arbeit von 2002 als auch in der Arbeit von 2005 direkt aus der regelmäßig erhobenen Abfrage über den gegenwärtigen Erwerbsstatus kodiert. Bei der

7 Die Auswahl dieser Kontraste folgt bei Kohler (2002) vollständig aus der Theorie.

aktualisierten Analyse mit den SOEP-Daten 1984–2010 ausgehend von der Aufbereitung von 2005 ergibt sich das Problem, dass in den Aufbereitungssyntaxdateien die Aufbereitung nur unvollständig dokumentiert ist. Hier wird die Analyse an die vollständig dokumentierte Aufbereitung von 2002 übernommen und sinngemäß angepasst.

Weiterhin wird für Bildung kontrolliert. Auch hier wird auf die in allen SOEP-Wellen ab 1985 vorliegende Information über den Abschluss einer Schul-, Hochschul- und Berufsausbildung im jeweils zurückliegenden Jahr verwendet. Hieraus werden für alle hier dargestellten Analysen in gleicher Form Indikatoren über das Ende der Schul-, Hochschul- und Berufsausbildung gebildet. Es ist darauf hinzuweisen, dass, wie bei der Darstellung des theoretischen Modells bereits erwähnt wurde, der Effekt der Bildung auf die Parteipräferenz nach der Theorie von Kohler (2002) nicht nur als Kontrollvariable zu verstehen ist. Da erst mit steigender Bildung unabhängig von den anderen Faktoren eine Neigung zu den Grünen vorhergesagt wird, ist die Bildung im Sinne der Argumentation von Kohler der wesentliche Grund für den Bruch der Parteienlandschaft über die einfache ökonomische Links-rechts-Dimension hinaus.

Neben der Erwerbstätigkeit und der Bildung wird ferner für Hochzeit und die Geburt von Kindern kontrolliert. Die Hochzeit wird als Dummyvariable konstruiert, die den Zeitpunkt der ersten Heirat markiert. Für die Erstellung wird die in allen SOEP-Wellen verfügbare, bereitgestellte Information über den Familienstand in jeder Welle genutzt. Die Variable wird entsprechend so kodiert, dass der Dummy die Ausprägung in der Welle von null auf eins wechselt, in der die Beobachtungseinheit verheiratet ist und in der direkt vorherigen Welle ledig ist. Für die Geburt von Kindern werden zwei Dummyvariablen konstruiert, die den Zeitpunkt markieren, in der die Beobachtungseinheit ihr erstes und in der sie das zweite Kind bekommt. Die Information wird aus dem für alle Panelwellen verfügbaren, vorab bereitgestellten Zusatzdatensatz „biobirth" abgeleitet.

Bei den bisherigen Ausführungen zur Operationalisierung der unabhängigen Variable wurde nur die inhaltliche Zusammensetzung der unabhängigen Variablen dargestellt. Ein wesentlicher Beitrag der Arbeit von Kohler ist die besondere Operationalisierung der unabhängigen Variable als Ereignisse (vgl. Kohler 2002, S. 240–244). Dies soll im Folgenden ausführlicher erläutert werden. Kohler kodiert die unabhängigen Variablen in allen Modellen nicht als einfache Zustandsvariablen, sondern als Ereignisindikatoren. D. h. die unabhängigen Variablen sind nicht als Indikatoren über den jeweiligen *Zustand* in einer Welle kodiert, sondern als Indikatoren über einen *Zustandswechsel*. Zur Veranschaulichung betrachten wir die einzelnen Variablen: Beim Effekt der Klassenzugehörigkeit betrachtet Kohler nicht die Parteipräferenz einer Beobachtungseinheit, die einer bestimmten sozialen Klasse angehört, sondern die Parteipräferenz einer Beobachtungseinheit, die von einer Klasse in eine andere Klasse gewechselt ist. Weiterhin mit den Kontrollvariablen nicht einfach der Effekt der Arbeitslosigkeit gegenüber der Erwerbstätigkeit betrachtet, sondern der Übergang von der Erwerbstätigkeit in die Arbeitslosigkeit und umgekehrt. Auch die Bildungeffekte sind Übergangs- statt Zustandseffekte, d. h. man betrachtet *nicht* den Effekt eines Bildungsniveaus, sondern den Effekt des Wechsels in einen höheren Abschluss. Analog sind auch der Effekt der Hochzeit und der Effekt der Geburt des ersten und zweiten Kindes Übergangseffekte. Wieder wird nicht der Einfluss der Ehe oder der

Kinderzahl, sondern der Einfluss des Übergangs in die Ehe, bzw. des Wechsels von der Kinderlosigkeit in die Elternschaft betrachtet.

Formal heißt das, dass Kohler (2002) beim Effekt der Klassenzugehörigkeit – in vereinfachter Darstellung – *nicht* dieses statistische Modell betrachtet:

$$\ln \frac{\Pr(\text{Parteipräferenz}_{it} = \text{„SPD”})}{\Pr(\text{Parteipräferenz}_{it} \neq \text{„SPD”})} = \alpha_{i\text{„SPD”}} + \beta_{\text{„SPD”}}\text{Klasse}_{it} + \epsilon_{it\text{„SPD”}}.$$

Stattdessen geht Kohler mit der Ereigniskodierung von diesem Modell aus:

$$\ln \frac{\Pr(\text{Parteipräferenz}_{it} = \text{„SPD”})}{\Pr(\text{Parteipräferenz}_{it} \neq \text{„SPD”})}$$
$$= \alpha_{i\text{„SPD”}} + \beta_{\text{„SPD”}}(\text{Klasse}_{it} - \text{Klasse}_{i,t-1}) + \epsilon_{it\text{„SPD”}}.$$

Man beachte, dass die Klasse hier vereinfacht als eine metrische Variable dargestellt ist. Wie oben beschrieben, wird bei Kohler diese soziale Klasse als vierstufige kategoriale Variable modelliert.

Diese Ereignisoperationalisierung führt aber zu einer inhaltlich schwierigen Interpretierbarkeit der Effekte. Durch diese Kodierung werden die Zeitpunkte vor und nach einem Wechsel gleich behandelt, da die Differenz $\text{Klasse}_{it} - \text{Klasse}_{i,t-1}$ zu diesem Zeitpunkt null ist. Man betrachtet also den Unterschied der Neigung für eine bestimmte Partei zum Zeitpunkt der Änderung der sozialen Klasse gegenüber der Neigung für diese Partei zu den Zeitpunkten, in denen die Klassenzugehörigkeit konstant ist.

Diese Modellierung ist insoweit problematisch, als damit der Effekt der Klassenzugehörigkeit selbst nicht identifiziert werden kann. Zum Verständnis betrachten wir wieder die formale, leicht reduzierte Darstellung des Beispiels von oben. Nehmen wir zunächst an, dass es keinen zeitverzögerten Effekt der Klassenzugehörigkeit des vorherigen Messzeitpunkts gibt:

$$\ln \frac{\Pr(y_{it} = j)}{\Pr(y_{it} \neq j)}$$
$$= \alpha_{ij} + \beta_{j1}x_{it} + \epsilon_{itj}$$

Mit β_{j1} sei der eigentlich interessierende Effekt der betrachteten unabhängigen Variablen x_t bezeichnet. Man kann den Ausdruck nun so umformen, dass man einen eigenen Ausdruck für die Differenz, wie sie Kohler operationalisiert, erhält. Dafür addiert man zuerst die triviale Differenz $(-\beta_{j1}x_{i,t-1} + \beta_{j1}x_{i,t-1})$:

$$= \alpha_{ij} + \beta_{j1}x_{it} - \beta_{j1}x_{i,t-1} + \beta_{j1}x_{i,t-1} + \epsilon_{itj}$$

Hieraus erhält man durch Zusammenfassen:

$$= \alpha_{ij} + \beta_{j1}(x_{it} - x_{i,t-1}) + \beta_{j1}x_{i,t-1} + \epsilon_{itj}.$$

Wenn man annimmt, dass es einen zeitverzögerten Effekt der Klassenzugehörigkeit des vorherigen Messzeitpunkts gibt, geht man von diesem Ausdruck aus:

$$\ln \frac{\Pr(y_{it} = j)}{\Pr(y_{it} \neq j)} = \alpha_{ij} + \beta_{j1} x_{it} + \beta_{j2} x_{i,t-1} + \epsilon_{itj}$$

Mit β_{j2} ist nun der zeitverzögerte Effekt der unabhängigen Variable des vorherigen Zeitpunktes x_{t-1} bezeichnet. Auch hier lässt sich der Ausdruck so umformen, dass man einen Ausdruck für die Ereignisoperationalisierung erhält. Wir ersetzen dafür zuerst den Koeffizienten β_{j1} mit dem Ausdruck $\beta_j^* - \beta_{j2}$:

$$= \alpha_{ij} + (\beta_j^* - \beta_{j2}) x_{it} + \beta_{j2} x_{i,t-1} + \epsilon_{itj}$$

Dann kann man die erste Differenz distributiv aufteilen.

$$= \alpha_{ij} + \beta_j^* x_{it} - \beta_{j2} x_{it} + \beta_{j2} x_{i,t-1} + \epsilon_{itj}$$

Dann kann man die Faktoren zum Koeffizienten β_{j2} zusammenfassen.

$$= \alpha_{ij} + \beta_j^* x_{it} + (-\beta_{j2})(x_{it} - x_{i,t-1}) + \epsilon_{itj}.$$

Wie man sieht, bleibt im ersten Fall als zusätzliches Restglied $\beta_{j1} x_{i,t-1}$. Im zweiten Fall bleibt entsprechend $\beta_j^* x_{it}$. Die Operationalisierung von Kohler ist nur dann gültig, wenn diese Restglieder entweder dadurch verschieden, dass der Regressionskoeffizient null ist, oder diese unabhängig von den anderen Kovariaten sind, und somit im Fehlerterm ϵ aufgenommen werden können. Für den ersten Fall heißt das, dass entweder das Ereignis keinen Effekt hat ($\beta_{j1} = 0$), oder dass der zeitverzögerte Zustand vom Ereignis unabhängig sein muss ($(x_{it} - x_{i,t-1}) \perp x_{i,t-1}$). Für den zweiten Fall heißt das, dass es entweder keinen eigenen Effekt des gegenwärtigen Zustands gibt ($\beta_j^* = 0$), oder dass der gegenwärtige Zustand vom Ereignis unabhängig sein muss ($(x_{it} - x_{i,t-1}) \perp x_{it}$). Im ersten Fall scheinen beide Alternativen unplausibel, im zweiten Fall ist am ehesten davon auszugehen, das der gegenwärtige Zustand keinen Effekt hat. Wenn diese Annahme aber verletzt ist, folgt aus der Operationalisierung von Kohler (2002), dass der entsprechende Effekt inkonsistent geschätzt wird. Man beachte erstens, dass dies sowohl für die Schätzer für den eigentlichen Effekt des Zustands, als auch den Effekt des Ereignisses gilt. Zweitens gelten diese Ausführungen für alle oben dargestellten Variablen, weil Kohler für alle unabhängigen Variablen eine Ereignisoperationalisierung vornimmt.

Selbst wenn man diese Ausführung außer acht lässt und nur von einem eigenständigen Effekt des Übergangs selbst ausgeht, ist zu beachten, dass die Übergangsereignisse grundsätzlich seltenere Ereignisse sind, d. h. die unabhängigen Variablen immer sehr schief verteilt sind. Dadurch sinkt im Allgemeinen die Vorhersagekraft, dass heißt die Standardfehler der Effekte sind größer.

Ohne Anbindung an die Theorie, sondern nur zur Berücksichtigung von Kalendereffekten, kontrolliert Kohler in allen Modellen für die Erhebungswellen. Hierfür nimmt er für jede Welle einen Dummy auf, wobei die jeweils letzte Welle die Referenzkategorie

bildet. Sowohl in der Replikation als auch in der Aktualisierung wird dieses Vorgehen beibehalten. Bei der Aktualisierung wird durch die Kürzung der Daten entsprechend nur für die betrachteten Wellen mit ungeraden Jahreszahlen betrachtet.

Soweit in den obigen Ausführungen nicht anders dargestellt wurde, werden für die Replikation und die Aktualisierung die Operationalisierungen direkt übernommen. Dies gilt insbesondere auch für die kritisierte Ereignisoperationalisierung, da das Ziel der vorliegenden Anwendung primär die Replikation der Arbeit von Kohler (2002) ist.

8.3.2 Ergebnisse

Wie bereits erwähnt, rechnet Kohler (2002) verschiedene Analysemodelle über den Effekt der Klassenzugehörigkeit auf die Parteipräferenz. Da in der vorliegenden Arbeit die Anwendung des femlogit-Modells im Vordergrund steht, beschränken wir uns auf die Darstellung und Replikation des einfachsten Modells von Kohler. Konkret handelt es sich hier um drei dichotome logistische Regressionsmodelle mit fixed-effects auf die Parteipräferenz. Im ersten Modell wird der Kontrast der Präferenz für die SPD gegenüber den anderen Parteien (oder keiner Partei) betrachtet, im zweiten Modell die Präferenz für die CDU/CSU oder FDP gegenüber den anderen Parteien und im dritten Modell die Präferenz für die Grünen gegenüber den anderen Parteien. In jedem dieser Modelle werden die unabhängigen Variablen verwendet, die oben ausführlich dargestellt wurden. Im Weiteren sollen nun die Ergebnisse von Kohler (2002) und die Replikation dieser Analysen vorgestellt werden. Danach werden die Analysen von Kohler dahingehend erweitert, dass die Analysen unter Verwendung des femlogit-Modells in einer gemeinsamen Analyse gerechnet werden. Abschließend werden diese Analysen aktualisiert, d. h. mit den SOEP-Daten von 1984–2010 neu gerechnet.

Zuerst betrachten wir die Analysen von Kohler (2002). Wie gesagt, rechnet Kohler dichotome logistische Regressionen mit fixed-effects auf die Kontraste der Parteipräferenz für die SPD, die CDU/CSU oder FDP und die Grünen – jeweils gegenüber allen anderen Parteien. Kohler legt großen Wert darauf, dass sowohl die Aufbereitung als auch die Analysen transparent und replizierbar sind, so dass er den gesamten Syntax öffentlich zu Reanalyse zur Verfügung stellt. Um eine möglichst exakte Replikation der Ergebnisse von Kohler zu erzielen, werden die Original-SOEP-Daten von 1984–1997 verwendet. Trotz dieser Maßnahmen gelingt es nicht, die Analysen von Kohler exakt wiederzugeben. In Tabelle 8.1 sind daher zunächst die Originalanalysen von Kohler (2002) wiedergegeben. Zunächst einmal ist zu beachten, dass bei diesen Ergebnissen nicht wie im vorherigen Abschnitt zur Replikation von Schröder (2010) die Odds-Ratio-Effekte dargestellt sind, sondern Logit-Effekte.

Tabelle 8.1: Binäre fixed-effects-logit-Modelle von Kohler (2002)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Administrative Dienstklasse → Selbständig	−0,81 (−0,14)	−0,09 (−0,24)	−0,02 -[a]
Administrative Dienstklasse → Mischt./Experte	0,06 (0,80)	−0,07 (−0,57)	0,22 (0,40)
Administrative Dienstklasse → Arbeiter/Soz. Dienste	0,22 (0,74)	0,08 (0,17)	−1,11 (−0,18)
Selbständig → Admin. Dienstkl.	0,06 (0,13)	−0,04 (−0,10)	0,56 -[a]
Selbständig → Mischt./Experte	0,71^{+} (1,74)	−0,07 (−0,15)	0,16 (0,03)
Selbständig → Arbeiter/Soz. Dienste	−0,28 (−1,35)	0,38 (0,09)	−0,73 (−0,16)
Mischtyp/Experte → Admin. Dienstkl.	−0,27 (−1,40)	0,23 (1,12)	−0,19 (−0,51)
Mischtyp/Experte → Selbständig	−0,65 (−0,80)	−0,09 (−0,47)	0,10 (0,02)
Mischtyp/Experte → Arbeiter/Soz. Dienste	−0,21** (−3,19)	−0,32^{+} (−1,65)	0,59 (1,36)
Arbeiter/Soziale Dienste → Admin. Dienstkl.	0,03 (0,11)	−0,14 (−0,29)	0,49 (0,64)
Arbeiter/Soziale Dienste → Selbständig	0,00 (0,00)	0,45 (1,29)	0,13 (0,03)
Arbeiter/Soziale Dienste → Mischt./Experte	0,04 (0,26)	0,15 (0,81)	−0,59** (−2,83)
Eintritt in Arbeitslosigkeit	0,17^{+} (1,73)	−0,26 (−1,41)	−0,04 (−0,22)
Austritt aus Arbeitslosigkeit	−0,21 (−1,43)	0,20 (0,94)	0,01 (0,06)
Schulabschluss	0,08 (0,32)	−0,11 (−0,54)	0,47* (2,26)
Hochschulabschluss	0,20 (0,74)	0,06 (0,30)	−0,41 (−1,32)
Berufsausbildungs-abschluss	0,11 (0,61)	0,03 (0,26)	−0,02 (−0,10)
Beginn Erwerbsleben	0,40^{+} (1,87)	0,09 (0,70)	0,28 (1,38)
Erste Hochzeit	0,42** (2,62)	0,14 (0,93)	−0,06 (−0,26)
Geburt des ersten Kindes	0,15 (0,91)	0,18 (1,30)	0,30 (1,01)

Tabelle 8.1: (Fortsetzung)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Geburt des zweiten Kindes	−0,02 (−0,12)	−0,11 (−0,71)	0,07 (0,28)
Personen	5 680	4 842	1 666
Messzeitpunkte/Personen	8,7	8,6	8,3
χ^2/(df)	777 (33)	850 (33)	217 (33)
R^2_{MF}	0,02	0,03	0,02

$^{+}: p < 0{,}1$, $^{*}: p < 0{,}05$, $^{**}: p < 0{,}01$. [a]: Korrigierte Standardfehler wegen zu kleiner Fallzahlen nicht berechenbar. Nicht dargestellte Kontrollvariablen: jahresweise Wellendummies. Standardfehler sind korrigiert für geschichtete Stichprobe. Schätzung unter Berücksichtigung von Design-, Nonresponse- und Attritionsgewichten.
Quelle: Kohler (2002, S. 304f.).

Weiterhin sind hier nicht die Standardfehler der Schätzer, sondern die entsprechenden z-Werte dargestellt. Der Grund hierfür ist, dass Kohler (2002) nur diese Werte mit sehr wenigen Nachkommastellen berichtet. Dadurch kann eine Rückumrechnung der Odds-Ratio-Effekte und der Standardfehler keine zufriedenstellenden Ergebnisse liefern. Weiterhin verwendet Kohler bei seinen Analysen erstens eine Kombination aus Design-, Ausfall- und Attrition-Gewichten und zweitens korrigiert er die Standardfehler für Probleme, die sich aus der Schichtung der Einzelstichproben ergeben. Durch die Korrektur der Standardfehler kann in zwei Fällen der Standfehler nicht berechnet werden.

Allgemein betrachtet fällt auf, dass Kohler wenig signifikante Effekte findet. Weiterhin sind die Effekte in Bezug auf die Hypothesen auch betragsmäßig sehr uneinheitlich. Bei den Effekten der Klassenzugehörigkeit sind selbst bei den stärksten Kontrasten die Befunde uneindeutig. Gemäß der Theorie sollte bei einem Wechsel aus der Selbständigkeit in die Gruppe der Arbeiter und der Beschäftigten in den „Sozialen Diensten" und bei einem Wechsel aus der Selbstständigkeit in die Gruppe der Mischtypen bzw. Experten die Neigung für die SPD und für die Grünen stark ansteigen und die Neigung für die CDU/CSU und für die FDP stark zurückgehen. Bei einem Wechsel in die umgekehrte Richtung sollte der entsprechend umgekehrte Effekt auftreten. Diese Effekte findet man nur sehr bedingt in den geschätzten Modellen. Beim Wechsel aus der Selbstständigkeit in die Gruppe der Mischtypen bzw. Experten steigt die Neigung zur SPD schwach signifikant, zumindest betragsmäßig schwach die Neigung zu den Grünen. Beim umgekehrten Übergang sinkt die Neigung zu den SPD, wenn auch nicht signifikant. Beim Übergang aus der Gruppe der Arbeiter und der Beschäftigten in den „Sozialen Diensten" in die Selbständigkeit steigt die Neigung zur CDU/CSU und zur FDP, wobei dieser wiederum nicht signifikant ist. Dagegen sinkt aber beim Wechsel in die umgekehrte Richtung entgegen der theoretischen Erwartung die Neigung zur SPD und zu den Grünen, und die Neigung zur CDU/CSU und zur FDP steigt.

Die Effekte der Kontrollvariablen sind ebenso uneindeutig. Bei Abschluss der Schulabschluss steigt zwar die Neigung zur den Grünen, bei Abschluss des Hochschulabschluss sinkt diese aber. Bei der Hochzeit und Geburt eines Kindes wird eine stärkere Neigung konservativen Parteien erwartet. Beide Effekte finden sich nicht in den geschätzten Modellen. Weiterhin wird bei Eintritt in die Arbeitslosigkeit ein Anstieg der Neigung zur SPD und zu den Grünen und umgekehrt bei Rückkehr aus der Arbeitslosigkeit eine Anstieg der Neigung zu den CDU/CSU und FDP. Diese Effekte tauchen auch in den geschätzten Modellen – wenn auch nicht signifikant – auf.

Die Replikation der Modelle von Kohler (2002) ist in Tabelle 8.2 dargestellt. Hier sind die gleichen Schätzmodelle dargestellt wie bei Kohler, jedoch wurde hier auf die Verwendung von Gewichten und das Verfahren zur Korrektur der Standardfehler verzichtet (s. o.).[8] Zunächst sieht man an diesen Ergebnissen, dass sich – wie bereits oben erwähnt wurde – die Fallzahlen der einzelnen Modelle von den Analysen bei Kohler (2002) unterscheiden.

Tabelle 8.2: Replikation der binären fixed-effects-logit-Modelle von Kohler (2002) (ungewichtet)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Administrative Dienstklasse → Selbständig	$-0{,}67^{**}$ $(-2{,}83)$	$-0{,}02$ $(-0{,}12)$	$0{,}00$ $(0{,}01)$
Administrative Dienstklasse → Mischt./Experte	$-0{,}13$ $(-1{,}34)$	$-0{,}08$ $(-0{,}95)$	$0{,}08$ $(0{,}48)$
Administrative Dienstklasse → Arbeiter/Soz. Dienste	$0{,}47^{*}$ $(2{,}53)$	$-0{,}16$ $(-0{,}82)$	$-0{,}66^{*}$ $(-2{,}12)$
Selbständig → Admin. Dienstkl.	$0{,}14$ $(0{,}60)$	$-0{,}01$ $(-0{,}05)$	$0{,}09$ $(0{,}24)$
Selbständig → Mischt./Experte	$0{,}60^{**}$ $(2{,}84)$	$-0{,}34^{*}$ $(-2{,}27)$	$0{,}48$ $(1{,}50)$
Selbständig → Arbeiter/Soz. Dienste	$0{,}05$ $(0{,}28)$	$0{,}28$ $(1{,}52)$	$-0{,}09$ $(-0{,}34)$
Mischtyp/Experte → Admin. Dienstkl.	$-0{,}11$ $(-1{,}15)$	$0{,}07$ $(0{,}80)$	$-0{,}05$ $(-0{,}30)$
Mischtyp/Experte → Selbständig	$-0{,}24$ $(-1{,}27)$	$0{,}30^{+}$ $(1{,}86)$	$0{,}14$ $(0{,}46)$
Mischtyp/Experte → Arbeiter/Soz. Dienste	$-0{,}09$ $(-1{,}41)$	$-0{,}11$ $(-1{,}50)$	$0{,}30^{*}$ $(2{,}30)$
Arbeiter/Soziale Dienste → Admin. Dienstkl.	$-0{,}13$ $(-0{,}74)$	$0{,}15$ $(0{,}73)$	$0{,}66^{*}$ $(2{,}01)$
Arbeiter/Soziale Dienste → Selbständig	$-0{,}27^{+}$ $(-1{,}70)$	$0{,}42^{**}$ $(2{,}79)$	$-0{,}10$ $(-0{,}37)$
Arbeiter/Soziale Dienste → Mischt./Experte	$0{,}04$ $(0{,}59)$	$-0{,}07$ $(-0{,}99)$	$-0{,}42^{***}$ $(-3{,}59)$

8 Die Replikation der Modelle von Kohler (2002) unter Verwendung von Design-, Ausfall- und Attritiongewichten findet sich im Anhang B in Tabelle B.1.

Tabelle 8.2: (Fortsetzung)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Eintritt in Arbeitslosigkeit	$0{,}16^{**}$ $(2{,}73)$	$-0{,}19^{*}$ $(-2{,}39)$	$-0{,}12$ $(-1{,}11)$
Austritt aus Arbeitslosigkeit	$-0{,}12^{*}$ $(-1{,}98)$	$0{,}10$ $(1{,}25)$	$0{,}10$ $(0{,}85)$
Schulabschluss	$0{,}30^{***}$ $(3{,}56)$	$0{,}23^{*}$ $(2{,}33)$	$0{,}37^{**}$ $(3{,}08)$
Hochschulabschluss	$-0{,}04$ $(-0{,}43)$	$0{,}09$ $(0{,}80)$	$-0{,}10$ $(-0{,}83)$
Berufsausbildungs-abschluss	$0{,}03$ $(0{,}46)$	$-0{,}02$ $(-0{,}25)$	$0{,}14$ $(1{,}64)$
Beginn Erwerbsleben	$0{,}18^{**}$ $(2{,}92)$	$0{,}27^{***}$ $(3{,}83)$	$0{,}17^{+}$ $(1{,}78)$
Erste Hochzeit	$0{,}17^{*}$ $(2{,}22)$	$0{,}15$ $(1{,}60)$	$-0{,}19$ $(-1{,}53)$
Geburt des ersten Kindes	$0{,}35^{***}$ $(4{,}70)$	$0{,}03$ $(0{,}29)$	$0{,}28^{*}$ $(2{,}30)$
Geburt des zweiten Kindes	$-0{,}02$ $(-0{,}30)$	$-0{,}04$ $(-0{,}58)$	$0{,}05$ $(0{,}53)$
Personen	5 750	4 926	1 696
Messzeitpunkte	49 711	41 742	13 978
χ^2/(df)	785,2 (33)	851,7 (33)	214,9 (33)
R^2_{MF}	0,021	0,027	0,022

$^{+}$: $p < 0{,}1$, *: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. Nicht dargestellte Kontrollvariablen: jahresweise Wellendummies.

Auch die Effekte unterscheiden sich, aber das Gesamtbild ist weniger uneinheitlich als bei Kohlers Analysen. Beim Effekt der Klassenzugehörigkeit findet man beim Wechsel aus der Selbständigkeit in die Gruppe der Mischtypen bzw. Experten erwartungsgemäß einen Anstieg der Neigung für die SPD und für die Grünen und einen Rückgang der Neigung für die CDU/CSU und für die FDP. Weiterhin findet man auch den analogen umgekehrten Effekt beim Übergang in die entgegengesetzte Richtung. Der theoretisch erwartete Anstieg der Neigung für die SPD und für die Grünen und ein Rückgang der Neigung für die CDU/CSU und für die FDP beim Übergang aus der Selbstständigkeit in die Gruppe der Arbeiter und der Beschäftigten in den „Sozialen Diensten” ist zwar nicht zu finden, aber der umgekehrte Effekte beim Wechsel in die entgegensetzte Richtung. Jedoch findet man auch Effekte, die der Erwartung widersprechen. Beim Wechsel aus der administrativen Dienstklasse in die Gruppe der Arbeiter und der Beschäftigen in den „Sozialen Diensten” sinkt die Präferenz für die Grünen signifikant. Einen analog umgekehrten signifikanten

Effekt findet man auch beim Übergang in die andere Richtung. Weiterhin steigt die Präferenz für die Grünen beim Wechsel aus der Gruppen der Mischtypen bzw. Experten in die Gruppe der Arbeiter und der der Beschäftigen in den „Sozialen Diensten". Auch hier findet man einen signifikanten umgekehrten Effekt beim Wechsel in die andere Richtung. Beide Effekte widersprechen den Hypothesen von Kohler. Insgesamt findet sich bei der Replikation also auch ein uneinheitliches Gesamtbild beim Einfluss der Klassenzugehörigkeit, wenn auch die Befunde zumindest teilweise den Erwartungen entsprechen.

Die Effekte der Kontrollvariablen verlaufen wie bei den Analysen von Kohler (2002) insgesamt entgegen der theoretischen Erwartungen. Entgegen der Erwartung eines einheitlichen Effekts der Bildung auf die Präferenz für die Grünen, findet man nur beim Abschluss der Schulausbildung einen Anstieg für alle Parteien. Weiterhin werden bei der Hochzeit und der Geburt von Kindern eine Tendenz zu konservativen Parteien erwartet. Beides findet sich nicht in den Ergebnissen. Insgesamt findet man starke Hinweise auf Alterseffekte auf die allgemeine Präferenz für alle Parteien. Mit Abschluss der Schulausbildung und Eintritt ins Erwerbsleben, also als junger Wähler, ist die Neigung zu *allen* Parteien höher.

Es lässt sich zusammenfassen, dass die Analysen von Kohler (2002) nur bedingt repliziert werden können. Trotz der transparenten Datenaufbereitung kann die Analysestichprobe nicht exakt reproduziert werden. Die Analyse kann zwar größtenteils wie im Original nachvollzogen werden, aber das Verfahren zur Korrektur der Standardfehler, dass Kohler verwendet, kann nicht erfolgreich repliziert werden. Die Ergebnisse der Replikation bieten zwar wie das Original ein uneinheitliches Bild, es findet sich aber insgesamt eine größere Übereinstimmung mit der theoretischen Erwartung als bei den Analysen von Kohler.

Im nächsten Schritt sollen Kohlers Analysen nun zunächst dahingehend erweitert werden, dass die Modelle über Kontraste zwischen den einzelnen Parteien mit dem femlogit-Modell in einem gemeinsamen Modell geschätzt werden. Hieraus sollten sich erstens eine effizientere Schätzung ergeben, d. h. die theoretisch erwarteten Effekte sollten deutlicher zutage treten. Zweitens sollten darüber hinaus theoretisch erwartete Unterschiede über die Präferenz über die einzelnen Parteien differenzierter erkennbar werden. In Tabelle 8.3 sind die Ergebnisse des femlogit-Modells auf die SOEP-Daten 1984–1997 dargestellt.

Tabelle 8.3: Anwendung des femlogit-Modells auf Kohler (2002) mit SOEP-Daten 1984–1997

	SPD	CDU/CSU	FDP	Grüne	Andere Parteien
	β/(z)	β/(z)	β/(z)	β/(z)	β/(z)
Admin. Dienstkl.	$-0{,}64^{**}$	$-0{,}15$	$0{,}41$	$-0{,}07$	$-0{,}07$
→ Selbständig	$(-2{,}62)$	$(-0{,}84)$	$(1{,}07)$	$(-0{,}19)$	$(-0{,}15)$
Admin. Dienstkl.	$-0{,}13$	$-0{,}06$	$-0{,}42^{+}$	$0{,}06$	$-0{,}09$
→ Mischt./Experte	$(-1{,}33)$	$(-0{,}61)$	$(-1{,}74)$	$(0{,}33)$	$(-0{,}36)$
Admin. Dienstkl.	$0{,}37^{+}$	$-0{,}11$	$-0{,}11$	$-0{,}52$	$0{,}54$
→ Arbeiter/Soz. Dienste	$(1{,}92)$	$(-0{,}52)$	$(-0{,}23)$	$(-1{,}63)$	$(0{,}91)$

Tabelle 8.3: (Fortsetzung)

	SPD	CDU/CSU	FDP	Grüne	Andere Parteien
	β/(z)	β/(z)	β/(z)	β/(z)	β/(z)
Selbständig → Admin. Dienstkl.	0,09 (0,36)	0,09 (0,52)	−0,47 (−1,11)	0,03 (0,08)	−0,09 (−0,17)
Selbständig → Mischt./Experte	0,64** (2,90)	−0,27^{+} (−1,71)	−0,23 (−0,64)	0,59^{+} (1,79)	−0,19 (−0,29)
Selbständig → Arbeiter/Soz. Dienste	−0,02 (−0,13)	0,32^{+} (1,67)	−0,36 (−0,73)	−0,21 (−0,71)	−0,37 (−0,67)
Mischt./Experte → Admin. Dienstkl.	−0,14 (−1,45)	−0,03 (−0,31)	0,56* (2,45)	−0,09 (−0,51)	−0,55* (−2,29)
Mischt./Experte → Selbständig	−0,22 (−1,14)	0,21 (1,21)	0,71^{+} (1,88)	0,04 (0,14)	−0,20 (−0,37)
Mischt./Experte → Arbeiter/Soz. Dienste	−0,06 (−0,95)	−0,11 (−1,37)	−0,09 (−0,38)	0,27* (2,06)	0,17 (0,99)
Arbeiter/Soz. Dienste → Admin. Dienstkl.	−0,04 (−0,22)	0,19 (0,89)	−0,40 (−0,77)	0,67* (2,00)	−0,60 (−1,15)
Arbeiter/Soz. Dienste → Selbständig	−0,29^{+} (−1,70)	0,44** (2,69)	0,88* (1,97)	−0,14 (−0,53)	0,66 (1,52)
Arbeiter/Soz. Dienste → Mischt./Experte	0,01 (0,14)	−0,09 (−1,25)	0,13 (0,58)	−0,41*** (−3,34)	0,13 (0,77)
Eintritt in Arbeitslosigkeit	0,16** (2,58)	−0,12 (−1,45)	−0,50* (−2,04)	−0,05 (−0,48)	0,44** (2,94)
Austritt aus Arbeitslosigkeit	−0,11^{+} (−1,78)	0,05 (0,63)	0,10 (0,40)	0,05 (0,45)	−0,29^{+} (−1,90)
Schulabschluss	0,39*** (4,50)	0,31** (2,82)	0,30 (1,38)	0,53*** (4,28)	0,27 (1,18)
Hochschulabschluss	−0,10 (−0,90)	0,05 (0,44)	0,25 (1,00)	−0,15 (−1,16)	−0,67 (−1,64)
Berufsausbildungsabschluss	0,05 (0,81)	−0,02 (−0,25)	0,29 (1,38)	0,18* (2,03)	0,14 (1,00)
Beginn Erwerbsleben	0,24*** (3,73)	0,31*** (4,13)	0,42* (2,46)	0,26** (2,65)	−0,01 (−0,06)
Erste Hochzeit	0,16* (2,03)	0,18^{+} (1,85)	−0,28 (−0,99)	−0,12 (−0,94)	−0,30 (−1,23)
Geburt des ersten Kindes	0,42*** (5,45)	0,08 (0,89)	0,37 (1,36)	0,45*** (3,59)	0,36 (1,51)

Tabelle 8.3: (Fortsetzung)

	SPD	CDU/CSU	FDP	Grüne	Andere Parteien
	β/(z)	β/(z)	β/(z)	β/(z)	β/(z)
Geburt des zweiten Kindes	$-0{,}01$ $(-0{,}14)$	$-0{,}09$ $(-1{,}33)$	$0{,}54^{**}$ $(2{,}72)$	$0{,}07$ $(0{,}79)$	$0{,}34^{+}$ $(1{,}90)$
Personen			10 387		
Messzeitpunkte			85 114		
χ^2/(df)			2 660,8 (165)		
R^2_{MF}			0,034		

$^{+}$: $p < 0{,}1$, *: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. Nicht dargestellte Kontrollvariablen: jahresweise Wellendummies.

Zunächst einmal ist bei diesem Modell zu beachten, dass wir hier die Logiteffekte auf die Wahrscheinlichkeit, eine bestimmte Partei zu präferieren, betrachten. Die Referenzkategorie ist hier die Präferenz für keine Partei. Im Gegensatz zu den dichotomen Modellen muss man hier bei der Vorzeicheninterpretation Vorsicht walten lassen. Ein positiver Effekt bedeutet, dass das Odds, eine bestimmte Partei statt keine Partei zu präferieren, steigt. Es bedeutet aber nicht, dass die Wahrscheinlichkeit, dass die Partei präferiert wird, steigt.

Bei der inhaltlichen Interpretation des Effekts der Klassenzugehörigkeit auf die Parteipräferenz ist nun eine etwas differenziertere Aussage möglich. Wir beschränken uns wieder nur auf die beiden stärksten Kontraste. Gegenüber den dichotomen Modellen erwarten wir nun beim Wechsel aus der Selbständigkeit in die Gruppe der Mischtypen und Experten nicht einfach einen uniformen Anstieg der Präferenz für die CDU/CSU und die FDP. Hier sollte die Präferenz für die FDP stärker ansteigen als die Präferenz für die CDU/CSU. Entsprechend sollte bei einem Wechsel in die umgekehrte Richtung die Präferenz für die FDP stärker sinken als die Präferenz für die CDU/CSU. Analog sollte dann dasselbe für den Wechsel aus der Selbständigkeit in die Gruppe der Arbeiter und Beschäftigten in den „Sozialen Diensten", bzw. für den Wechsel in die umgekehrte Richtung.

Jedoch sind auch hier die Effekte der Klassenzugehörigkeit uneinheitlich. Einerseits findet man bestätigende Effekte: Beim Übergang aus der Selbständigkeit in die Gruppe der Mischtypen und der Experten steigt das Odds, die SPD gegenüber keine Partei zu wählen, signifikant um den Faktor 1,89 ($= \exp(0{,}64)$). Das Odds, die Grünen statt keine Partei zu wählen, steigt um den Faktor 1,8. Das Odds, die CDU/CSU zu wählen statt keine Partei, sinkt um den Faktor 1,31 ($= 1/\exp(-0{,}27)$). Das Odds, die FDP statt keine Partei zu wählen, sinkt gemäß der Theorie auch, aber nur um den Faktor 1,26. Weiterhin sinkt beim Wechsel aus der Gruppe der Arbeiter und der Beschäftigten in den „Sozialen Diensten" in die Selbständigkeit das Odds, die SPD gegenüber keiner Partei zu wählen, um den Faktor 1,34. Das Odds, die Grünen statt keine Partei zu wählen, steigt um das 1,15-fache. Demgegenüber sinkt das Odds, die CDU/CSU statt keiner Partei zu wählen, signifikant

um den Faktor 1,55, und das Odds, die FDP gegenüber keiner Partei zu wählen, sinkt signifikant um den Faktor 2,41.

Neben diesen prinzipiell theoriekonformen Befunden in Bezug auf den Effekt der Klassenzugehörigkeit finden wir aber auch, dass beim Wechsel aus der Selbständigkeit in die Gruppe der Arbeiter und Beschäftigen in den „Sozialen Diensten” das Odds, die SPD statt keiner Partei zu wählen, minimal sinkt. Weiterhin sinkt das Odds entgegen der Erwartung, die Grünen gegenüber keiner Partei zu wählen, um den Faktor 1,23, und das Odds, die CDU statt keine Partei zu wählen, steigt um den Faktor 1,38.

Auch bei den Effekten der Kontrollvariablen bleiben manche Hypothesen wie bei den dichotomen Modellen bestehen, und wie oben finden wir auch hier ein uneinheitliches Bild. Mit Eintritt in die Arbeitslosigkeit sollte die Präferenz für die Oppositionsparteien, also die SPD und die Grünen, steigen und mit Rückkehr in die Erwerbstätigkeit sollte die Präferenz für die Regierungsparteien, also die CDU/CSU und die FDP steigen. Beide Hypothesen werden mit den Daten nicht bestätigt. Weiterhin sollte mit steigender Bildung, also mit Abschluss höherer Bildungsgrade die Präferenz für die Grünen steigen. Diese Hypothese findet insgesamt auch keine Bestätigung in den Ergebnissen. Neben diesen Hypothesen sind beim gemeinsamen femlogit-Modell etwas differenzierte Aussagen möglich. Nach der ersten Hochzeit und mit Geburt von Kindern sollte die Präferenz für die konservativen Parteien, also die CDU/CSU, steigen. Auch diese Hypothesen kann man in Anbetracht der Ergebnisse als widerlegt betrachten. Jedoch findet man bei den Effekten der Kontrollvariable insgesamt Hinweise auf einen Alterseffekt auf die Präferenz, überhaupt zu wählen. Bei Abschluss der Schulbildung steigt das Odds, irgendeine Partei zu wählen, mindestens um den Faktor 1,31. Beim Eintritt ins Erwerbsleben steigt das Odds, eine der größeren Partei statt nicht zu wählen oder eine andere Partei zu wählen, um mindestens um den Faktor 1,27. Beide Ereignisse treten eher bei jüngeren Wählern auf als bei älteren. Daneben findet man auch Hinweise darauf, dass Arbeitslosigkeit dazu führt, dass extreme Parteien gewählt werden. Beim Übergang in die Arbeitslosigkeit steigt das Odds, andere Parteien statt keine Partei zu wählen, signifikant um den Faktor 1,55, und umgekehrt sinkt das Odds, andere Parteien statt nicht zu wählen, bei Rückkehr in die Erwerbstätigkeit um den Faktor 1,33.

Im zweiten Schritt erweitern wir die Analysen von Kohler (2002) um die aktuellen SOEP-Daten 1984–2010. Durch die vergrößerte Datenbasis ist einerseits zu erwarten, dass die in den bisher dargestellten Modellen sehr schwachen Effekte deutlicher erkennbar werden. Andererseits steht dieser Erwartung entgegen, dass in der Literatur diskutiert wird, dass der Effekt von soziostrukturellen Merkmalen auf die Parteibindung über die Zeit hinweg zurückgegangen ist (vgl. z. B. Pappi und Brandenburg 2010). Bei der Ausdehnung auf die aktuellen Daten kann man wiederum zuerst nur die dichotomen Modelle von Kohler betrachten, bevor man die aktuelle Datenbasis auf das vereinheitlichte femlogit-Modell anwendet. Aus Platzgründen beschränken wir uns hier auf die Darstellung des gemeinsamen Modells.[9] Die Ergebnisse der Schätzung des femlogit-Modells mit den Daten des SOEP von 1984–2010 sind in Tabelle 8.4 dargestellt. Bei diesen Modellen ist zunächst zu

9 Die Ergebnisse der dichotomen Modelle sind im Anhang B in Tabelle B.2 zu finden.

beachten, dass, wie bereits oben geschildert wurde, die Anzahl der Messzeitpunkte verringert wurde, um die Rechenzeit in einem realistischen Rahmen zu halten. Hierfür werden vom vollständig aufbereiteten Analysedatensatz nur die ungeraden Jahre verwendet.[10] Diese Beschränkung hat natürlich zur Folge, dass die Standardfehler der Effekte steigen, und damit der Präzisionszugewinn durch die längeren Zeitreihen abgeschwächt wird.

Tabelle 8.4: Anwendung des femlogit-Modells auf Kohler (2002) mit SOEP-Daten 1984–2010

	SPD	CDU/CSU	FDP	Grüne	Andere Parteien
	β/(z)	β/(z)	β/(z)	β/(z)	β/(z)
Admin. Dienstkl. → Selbständig	−0,03	−0,06	−0,00	0,05	0,01
	(−0,38)	(−0,79)	(−0,01)	(0,41)	(0,06)
Admin. Dienstkl. → Mischt./Experte	0,08	0,16	0,19	0,08	0,07
	(0,72)	(1,40)	(0,82)	(0,44)	(0,31)
Admin. Dienstkl. → Arbeiter/Soz. Dienste	0,01	0,05	−0,33	−0,04	0,03
	(0,15)	(0,65)	(−1,39)	(−0,25)	(0,18)
Selbständig → Admin. Dienstkl.	0,05	0,11	0,08	0,10	−0,06
	(0,62)	(1,45)	(0,52)	(0,85)	(−0,42)
Selbständig → Mischt./Experte	0,16	0,06	−0,00	−0,42	−0,62*
	(1,06)	(0,36)	(−0,01)	(−1,53)	(−2,50)
Selbständig → Arbeiter/Soz. Dienste	0,09	−0,03	0,12	−0,42*	−0,07
	(1,01)	(−0,38)	(0,44)	(−2,17)	(−0,44)
Mischt./Experte → Admin. Dienstkl.	−0,08	0,11	−0,28	0,21	0,16
	(−0,79)	(0,96)	(−1,25)	(1,30)	(0,81)
Mischt./Experte → Selbständig	0,11	−0,07	0,05	−0,37	−0,38
	(0,72)	(−0,45)	(0,15)	(−1,45)	(−1,38)
Mischt./Experte → Arbeiter/Soz. Dienste	−0,07	0,09	0,23	−0,13	0,02
	(−0,52)	(0,66)	(0,59)	(−0,55)	(0,07)
Arbeiter/Soz. Dienste → Admin. Dienstkl.	−0,01	−0,21**	0,10	0,11	0,01
	(−0,13)	(−2,72)	(0,46)	(0,83)	(0,08)
Arbeiter/Soz. Dienste → Selbständig	−0,16*	0,02	0,22	−0,24	−0,19
	(−2,04)	(0,20)	(0,84)	(−1,33)	(−1,25)
Arbeiter/Soz. Dienste → Mischt./Experte	0,03	−0,26*	−0,31	0,13	0,08
	(0,23)	(−2,05)	(−0,88)	(0,66)	(0,38)
Eintritt in Arbeitslosigkeit	0,10***	−0,03	−0,30***	0,24***	0,11*
	(3,43)	(−0,91)	(−3,95)	(4,60)	(1,96)
Austritt aus Arbeitslosigkeit	−0,10*	−0,11*	0,31*	−0,11	−0,10
	(−2,57)	(−2,37)	(2,53)	(−1,54)	(−1,21)
Schulabschluss	0,55***	0,43***	0,59**	0,50***	0,17
	(6,95)	(4,62)	(3,10)	(4,65)	(1,19)
Hochschulabschluss	0,16*	0,20*	0,08	0,20*	−0,23
	(2,05)	(2,42)	(0,52)	(2,28)	(−1,51)

10 Man beachte, dass die Auswahl der ungeraden Jahre nicht bedeutet, dass die Zeitspannen der Übergänge größer sind, sondern dass man im Mittel nur die Hälfte der Übergänge betrachtet.

Tabelle 8.4: (Fortsetzung)

	SPD	CDU/CSU	FDP	Grüne	Andere Parteien
	β/(z)	β/(z)	β/(z)	β/(z)	β/(z)
Berufsausbildungs-abschluss	−0,02 (−0,50)	0,15** (3,16)	0,38** (3,21)	0,15* (2,15)	0,00 (0,00)
Beginn Erwerbsleben	0,30*** (3,50)	0,08 (0,85)	0,16 (0,82)	0,13 (1,18)	0,04 (0,27)
Erste Hochzeit	0,11 (1,53)	0,12 (1,47)	0,25 (1,21)	0,09 (0,77)	−0,08 (−0,59)
Geburt des ersten Kindes	0,34*** (5,18)	0,16* (2,16)	0,15 (0,79)	0,31** (2,91)	−0,13 (−1,05)
Geburt des zweiten Kindes	−0,05 (−1,03)	0,02 (0,43)	0,15 (1,16)	0,07 (1,00)	−0,06 (−0,59)
Personen	22 195				
Messzeitpunkte	107 850				
χ^2/(df)	3 882,3 (230)				
R^2_{MF}	0,040				

$^{+}$: $p < 0{,}1$, *: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. Nicht dargestellte Kontrollvariablen: zweijahresweise Wellendummies.

Betrachten wir die Schätzergebnisse in Tabelle 8.4 wieder zuerst in Bezug auf den Effekt der Klassenzugehörigkeit und dann für die Kontrollvariablen. Beim Effekt der Klassenzugehörigkeit erkennt man, dass keiner der theoretischen erwarteten Effekte mehr zu erkennen ist. Selbst bei den vier Übergängen, bei denen die stärksten Effekte auf die Präferenzen zu erwarten sind, entsprechen nur noch die Effekte beim Wechsel aus der Gruppe der Arbeiter und der Beschäftigten in den „Sozialen Diensten" in die Selbständigkeit in etwa der theoretischen Erwartung: Das Odds für die SPD statt Nichtwahl und für die Grünen statt Nichtwahl gehen zurück, und das Odds für die FDP gegenüber Nichtwahl steigt.

Auch die Effekte der Kontrollvariablen entsprechen nicht der theoretischen Erwartung. Mit steigender Bildung, d. h. nach dem Schul-, Hochschul- oder Berufsausbildungsabschluss steigen bis auf Ausnahmen die Odds, eine der vier bedeutenden Parteien statt nicht zu wählen, signifikant an. Weiterhin wird auch die Erwartung über den Effekt der Hochzeit und der Kindergeburt nicht bestätigt. Die Hochzeit und die Geburt des zweiten Kindes hat keinen signifikanten Effekt. Bei Geburt des ersten Kindes steigen sowohl das Odds, die konservative CDU/CSU statt nicht zu wählen, als auch die entsprechenden Odds, die SPD oder die Grünen zu wählen, signifikant. Die Effekte der Arbeitslosigkeit entsprechen in etwa der Erwartung aus der Theorie. Bei Eintritt in die Arbeitslosigkeit steigen die Odds, die Oppositionsparteien SPD und Grüne zu wählen, und zumindest das Odds, die

FDP zu wählen, sinkt. Umgekehrt sinkt bei Rückkehr in die Erwerbstätigkeit das Odds, die SPD und die Grünen statt nicht zu wählen, und das Odds, die FDP statt nicht zu wählen, steigt. Jedoch sinkt bei diesem Wechsel entgegen der theoretischen Erwartung das Odds, die CDU zu wählen, signifikant.

Insgesamt ist festzuhalten, dass mit der Erweiterung des gemeinsamen Modells über die Analysen von Kohler (2002) um die aktuellen SOEP-Daten kein einheitlicher Effekt der Klassenzugehörigkeit auf die Parteipräferenz mehr zu erkennen ist. Auch die übrigen von Kohler erwarteten Effekte der soziostrukturellen Merkmale auf die Parteipräferenz treten nur sehr uneinheitlich auf.

8.4 Zusammenfassung

In diesem Abschnitt wurden die Analysen zum Effekt der sozialen Klassenzugehörigkeit auf die Parteipräferenz aus der Arbeit von Kohler (2002) repliziert und erweitert. Ausgangspunkt bilden die dichotomen logistischen Regressionen mit fixed-effects über die Präferenz für eine bestimmte Partei gegenüber allen anderen Parteien, wie sie Kohler in seiner Arbeit verwendet.

Im ersten Schritt wird versucht, diese Analysen möglichst exakt zu replizieren. Obwohl Kohler seine Aufbereitungs- und Auswertungssyntax vollständig transparent der Öffentlichkeit zur Verfügung stellt, und weiterhin sogar die Originaldaten beschafft werden konnten, gelang es nicht, die Analysen von Kohler (2002) vollständig zu replizieren. Erstens kann nicht die gleiche Analysestichprobe erzeugt werden, und zweitens kann das von Kohler verwendete Verfahren zur Korrektur der Standardfehler nicht nachvollzogen werden. Trotz dieser Mängel unterscheiden sich die Effekte der Nachbildung der Analysen von Kohler substantiell wenig vom Original. Im zweiten Schritt werden die Analysemodelle von Kohler dadurch erweitert, dass die getrennten dichotomen logistischen Regressionsmodelle bei Kohler mit dem femlogit-Modell als gemeinsames Modell über die Präferenz für alle Parteien bzw. Präferenz für keine Partei geschätzt werden. Im dritten Schritt werden die Modelle von Kohler durch die Verwendung der aktuellen SOEP-Daten 1984–2010 ergänzt.

Der zusätzliche Beitrag des femlogit-Modells und der Verwendung der aktuellen Daten lässt sich weniger eindeutig bewerten als im vorherigen Abschnitt über die Replikation der Arbeit von Schröder (2010). Zunächst einmal findet man schon im Ausgangsmodell von Kohler nur eine uneinheitliche Bestätigung der Hypothesen von Kohler über den Einfluss der Klassenzugehörigkeit auf die Parteipräferenz und über den Einfluss der Kontrollvariablen. Es ist darauf hinzuweisen, dass in der vorliegenden Arbeit, um eine einfachere Darstellung und Interpretation der Ergebnisse zu gewährleisten, das einfachste Modell von Kohler (2002) zur Replikation ausgewählt wurde. In den komplexeren Modellen bei Kohler findet man teilweise eine deutlichere Bestätigung für seine Hypothesen.

In der Replikation von Kohlers einfachstem Modell mit dem gleichen Datenbestand zeigt sich eine deutlichere Bestätigung der Vorhersagen über den Einfluss der Klassenzugehörigkeit. Zumindest beim Wechsel zwischen den Klassengruppen, bei denen der stärkste Effekt auf die Parteipräferenz erwartet wird, entsprechen bei der Replikation von

Kohlers Analysen die Ergebnisse der theoretischen Erwartung. Wie in Tabelle 8.2 zu sehen ist, steigt beim Übergang aus der Selbständigkeit in die Gruppe der Mischtypen und Experten die Präferenz für die SPD und die Grünen und die Präferenz für die CDU/CSU und die FDP sinkt. Beim umgekehrten Wechsel aus der Gruppe der Mischtypen und Experten in die Selbständigkeit und beim Wechsel aus der Gruppe der Arbeiter und Beschäftigten in den „Sozialen Diensten” in die Selbständigkeit sinkt die Präferenz für die SPD und die Grünen und die Präferenz für die CDU/CSU und die FDP steigt. Die Effekte der Kontrollvariablen entsprechen nicht der theoretischen Erwartung. Dagegen ist ein Alterseffekt auf die Wahlteilnahme erkennbar. Bei Abschluss der Schulausbildung und beim Eintritt ins Erwerbsleben, also bei Ereignissen, die eher am Anfang eines „Wählerlebens” stehen, steigen die Präferenzen für alle bedeutenden Parteien.

Insgesamt sind die Befunde in den Ausgangsmodellen von Kohler in Bezug auf den Effekt der Klassenzugehörigkeit auf die Parteipräferenz also bestenfalls uneinheitlich. Die gemeinsame Schätzung mit dem femlogit-Modell sollte dazu führen, dass die vorhergesagten Effekte deutlicher zu Tage treten. Erstens sind durch die Trennung der Präferenzen für die einzelnen Parteien und die Präferenz für keine Partei die Kontraste besser und trennschärfer modelliert als bei den dichotomen Modellen. Drittens ermöglicht die getrennte Analyse die Prüfung der spezielleren Hypothese über die unterschiedlichen Grade der Präferenzen für die CDU/CSU und die FDP. Viertens wird das femlogit-Modell mit einer größeren Fallzahl als die einzelnen dichotomen Modelle geschätzt, was wiederum dazu führen sollte, dass die erwarteten Effekte deutlicher erkennbar sein sollten. Die Ergebnisse der Analyse mit dem femlogit-Modell mit den SOEP-Daten von 1984–1997 zeigen eine schwache Bestätigung der Hypothesen von Kohler über den Effekt der Klassenzugehörigkeit auf die Parteipräferenz. Vereinfacht ausgedrückt, steigt beim Wechsel aus der Selbständigkeit in Klassengruppen, die der politischen Linken zugeordnet werden, die Präferenz für linke Parteien und es sinkt die Präferenz für die rechten Parteien. Beim Wechsel in die umgekehrte Richtung sind die Effekte umgekehrt. Neben diesem Befund, den wir schon im Replikationsmodell gefunden haben, finden wir hier nun eine schwache Bestätigung für die stärkere Präferenz für die FDP als für die CDU/CSU beim Wechsel aus einer linksstehenden Klasse in die Selbständigkeit. Bei den Kontrollvariablen bleibt mit der gemeinsamen Schätzung die uneinheitliche Bestätigung der erwarteten Effekte bestehen. Wie bei der Replikation finden wir auch hier Hinweise auf Alterseffekte auf die allgemeine politische Partizipation. Hinzu kommen aber Hinweise auf einen Effekt der Erwerbstätigkeit auf die Präferenz für extreme Parteien. Durch die differenzierte Modellierung der „anderen Parteien” neben den großen Parteien können wir messen, dass mit dem Wechsel in die Arbeitslosigkeit die Präferenz für die „anderen Parteien” ansteigt.

Zuletzt werden die Analysen von Kohler mit den aktuellen SOEP-Daten 1984-2010 geschätzt. Hier ist einerseits zu erwarten, dass mit der größeren Datenbasis die gemäß der Theorie zu erwartenden Effekte deutlicher werden. Andererseits wird in der relevanten Literatur auch diskutiert, dass der Einfluss soziostruktureller Merkmale auf die Parteibindung gesunken ist und auch weiterhin sinken wird. Alleine aus diesen Grund ist unklar, ob wir mit den zusätzlichen Daten eher präzisere Schätzer erhalten sollten oder nicht. Hinzu kommt, dass die Zeitreihen zusätzlich gekürzt werden, um die Rechenzeit in ei-

nem handhabbaren Rahmen zu halten. Die Schätzergebnisse zeigen keine einheitlichen Effekte der Klassenzugehörigkeit, und bei den Kontrollvariablen zeigt sich keiner der erwarteten Effekte. Jedoch zeigt sich hier, dass mit steigender Bildung die Präferenz für die etablierten Parteien steigt.

Insgesamt ist der Zugewinn durch die Verwendung des femlogit-Modells unklar. Es ist aber zu vermuten, dass dies nicht dem Modell geschuldet ist, sondern vor allem der Ereignisoperationalisierung der unabhängigen Variablen. Kohler hebt in seiner Arbeit hervor, dass die unabhängigen Variablen nicht als Zustandsvariablen sondern als Ereignisindikatoren kodiert werden müssen. Dem ist zu entgegnen, dass man in der Regel am Effekt der eigentlichen Ausprägung einer unabhängigen Variablen, und nicht am Effekt der Änderung interessiert ist, und dass man diesen mit Kohlers Operationalisierung nicht schätzen kann.[11] Wenn man aber wirklich am Effekt der Änderung interessiert ist, sollte man für den Ausgangszustand kontrollieren, um konsistente Schätzer zu erhalten. Schließlich hat Kohlers Operationalisierung den Nachteil, dass Ereignisse typischerweise selten sind, d. h. die unabhängigen Variablen sehr schief verteilt sind. Selbst wenn Kohlers theoretische Ableitungen hinsichtlich des Effekts der Änderung der Klassenzugehörigkeit und der anderen soziostrukturellen Merkmale richtig ist, kann alleine die Seltenheit der Änderungen der Hauptgrund dafür sein, dass man in allen Modellen bestenfalls sehr schwache Effekte der soziostrukturellen Merkmale findet. Darüber hinaus verfehlt Kohler aber gerade beim Einfluss der sozialen Klassen mit der Ereigniskodierung seine eigene theoretische Implikation. Kohler selbst argumentiert, dass soziale Klassen von bestimmten Parteien vertreten werden (vgl. ebd., S. 68–79). Durch die Ereignisoperationalisierung impliziert Kohler dagegen aber, dass Personen zum Zeitpunkt der Änderung der Klassenlage eine bestimmte Parteipräferenz haben. Zu den Zeitpunkten, in denen sich die Klassenzugehörigkeit gegenüber dem Vorjahr nicht ändert, haben alle Personen, *gleich welcher Klasse sie angehören*, die gleichen Parteipräferenzen. Nach Ansicht des Autors sollte man mit einer Operationalisierung der unabhängigen Variablen als Zustandsvariablen statt als Ereignisvariablen deutlichere Effekte der Klassenzugehörigkeit auf die Parteipräferenz erhalten. Es ist hinzuzufügen, dass hierdurch nicht ausgeschlossen werden muss, dass die Effekte zeitverzögert wirken, so wie Kohler in seinen komplexeren Modellen nachzuweisen versucht.

11 Ein Ausnahme findet sich z. B. bei Beck, Brüderl, und Woywode (2008).

9 Zusammenfassung

In diesem Kapitel wurde das im ersten Teil der Arbeit entwickelte und implementierte femlogit-Modell beispielhaft für die Replikationen und Erweiterungen von Analysen aus den Arbeiten von Schröder (2010) und Kohler (2002) anwendet. Dabei konnten wir gerade bei der Replikation der Analysen von Schröder den Zugewinn durch das femlogit-Modell zeigen. Erst mit dem multinomialen Modell kann gezeigt werden, dass mit steigendem Alter des jüngsten Kindes die Neigung zur Erwerbstätigkeit in Teilzeit gegenüber der Arbeitslosigkeit im Vergleich zu den kinderlosen Frauen nicht nur nach der Geburt wieder ansteigt und das gleiche Niveau wie das der Kinderlosen erreicht. Darüber hinaus kann man zeigen, dass ungefähr ab dem Einschulungsalter die Neigung zur Teilzeiterwerbstätigkeit gegenüber der Erwerbslosigkeit bei den Müttern sogar das der Kinderlosen übertrifft, d. h. solche Frauen dann sogar eher in Teilzeit arbeiten statt erwerblos zu sein als kinderlose Frauen. Mit dem femlogit-Modell gelingt es also, unterschiedliche Effekte auf die einzelnen Alternativen unter Kontrolle der zeitlichen konstanten, unbeobachteten Heterogenität zu schätzen.

Das Verfahren ist aber auch nicht unproblematisch. Erstens kann das Versprechen, dass die Implementation auf alle Anwendungen uneingeschränkt anwendbar ist, nicht vollständig eingehalten werden. Da die Rechenzeit proportional zur Anzahl der Permutationen der Zeitreihe der abhängige Variable ist, steigt die Rechenzeit in der Regel faktoriell mit der Anzahl der Messzeitpunkte. D. h. bei realen Anwendungen, wie bei beiden in diesem Abschnitt dargestellten Beispielen, ist die Anzahl der Messzeitpunkte ohne weitere Einschränkungen so groß, dass die Rechenzeit schnell selbst mit Hochleistungsrechnern mehrere hundert Jahre dauern kann. So mussten bei beiden Replikationen die Anzahl der Messzeitpunkte gekürzt werden, um die Rechenzeit in einem realisierbaren Rahmen zu halten. Bei der Replikation der Analysen von Schröder wurde gegenüber der Vorlage ein Sechstel der Zeitpunkte verwendet, bei der Replikation von Kohler die Hälfte. Grundsätzlich sinkt mit der Reduktion der Messzeitpunkte, wenn diese geeignet zufällig ist, nur die Effizienz der Schätzer, die Konsistenz ist aber weiterhin gegeben. Gerade bei fixed-effects-Schätzern ist das Problem der Ineffizienz von besonderer Schwere. Da man sich bei fixed-effects-Modellen auf die Information, die in der Varianz der unabhängigen und abhängigen Variablen *innerhalb* einer Person über die Messzeit steckt, beschränkt, sinkt die relative Effizienz der Schätzer ohnehin schon, so dass durch die zusätzliche Kürzung der Messzeitpunkte das Problem auftreten kann, dass schwache Effekte nicht mehr identifizierbar sind. Es ist aber darauf hinzuweisen, dass die lange Rechenzeit zwar bei diesen fixed-effects-Modellen grundsätzlich stark von der Anzahl der Messzeitpunkte abhängt, aber die Implementation in der vorliegenden Version auch noch Optimierungpotential enthält. D. h. es erscheint durchaus möglich, dass bei beiden Anwendungen eine Verdopplung der Messzeitpunkte in einem realistischen Zeitraum geschätzt werden können.

Zweitens zeigt sich bei den vorliegenden Replikationen das Problem des femlogit-Modells, dass der Verzicht auf die Annahme der Unabhängigkeit der unabhängigen Variablen von der unbeobachteten Heterogenität bei den nicht linearen Modellen dadurch erkauft wird, dass kein Effekt auf die Auswahlwahrscheinlichkeiten interpretiert werden kann.

D. h. man ist bei den Interpretation auf Odds-Ratio- und Logit-Effekt-Interpretation angewiesen. Wie schon im ersten Teil der Arbeit diskutiert wurde, haben beide Interpretationen den Nachteil, dass sie erstens sehr unanschaulich sind, und dass sie zweitens direkt von der arbiträren Annahme über die Varianz der idiosynkratischen Fehlerterme abhängen, was das Problem der „neglected heterogeneity" zur Folge haben kann. Zum ersten Problem ist zu entgegnen, dass Odds-Ratio-Effekte zwar unanschaulicher als Wahrscheinlichkeitseffekte sind, aber bei entsprechender Sorgfalt grundsätzlich nachvollziehbare Interpretationen möglich sind. Das zweite Problem ist schwerlich ganz von der Hand zu weisen, jedoch sollte das Problem bei fixed-effects-Modellen dadurch abgeschwächt werden, dass bei diesen Modellen für alle zeitlich konstanten Variablen kontrolliert wird. Wie bereits im Abschnitt zu Interpretation der fixed-effects-Modelle auf Seite 47 dargestellt wurde, skizzieren Cameron und Trivedi (2010) eine Möglichkeit, den Effekt auf die bedingte Wahrscheinlichkeit zu interpretieren. Weiterhin schlägt, wie in Abschnitt 7.3.2 erläutert wurde, Schröder eine Interpretation der Effekte auf die unbedingten Wahrscheinlichkeit vor. Nach Ansicht des Autors stellen beide Vorgehensweisen keine gangbaren Alternativen dar. Im ersten Fall müssen für die Beobachtungseinheiten bestimmte Zeitreihen gewählt werden, auf die die Wahrscheinlichkeiten bedingt werden. Im zweiten Fall müssen Heterogenitäten bestimmt werden, aus denen die unbedingten Wahrscheinlichkeiten berechnet werden. Beide Entscheidungen sind in etwa gleichwertig und auf den ersten Blick unproblematisch, da sich diese scheinbar aus der für jede Beobachtungseinheit vorliegenden Zeitreihe der abhängigen Variablen ergeben. Beim Vorschlag von Cameron und Trivedi würde man auf die gegebene Zeitreihe bedingen, bei Schröder substituiert man die Heterogenität mit der suffizienten Statistik, also der Anzahl der Nennungen jeder Alternativen in der gegebenen Zeitreihe. Das entscheidende Problem bei beiden Vorgehensweisen ist aber, dass man für eine Interpretation der Effekte auf die Wahrscheinlichkeiten die Kovarianzen der Heterogenitäten mit den beobachteten unabhängigen Variablen benötigt. Gerade hierfür hat man keine ausreichenden Hinweise, um eine sinnvolle Werte festzulegen. Wenn man dagegen naiverweise keine Korrelation annimmt, könnte man das Modell auch mit einem pooled- oder random-effects-Modell schätzen, und so direkt Effekte auf die Wahrscheinlichkeit schätzen.

Insgesamt besteht also bei der Anwendung des femlogit-Modells das Dilemma, dass man zwar für die unbeobachtete Heterogenität implizit kontrollieren kann, aber die Interpretation der Ergebnisse erschwert wird. Deshalb sollte man eine Anwendung des femlogit-Modells immer mit pooled- oder random-effects-Modellen untermauern.

Abschließend ist noch zu erwähnen, dass beide in dieser Arbeit dargestellten Anwendungen in gleicher Weise durch das femlogit-Modell erweitert werden. Sowohl in der Arbeit von Schröder (2010) als auch in der Arbeit von Kohler (2002) werden dichotome logistische Regressionsmodelle mit fixed-effects so verwendet, dass über eine mehrstufige kategoriale Variable verschiedene dichotome Kontraste gebildet werden. Mit dem femlogit-Modell werden diese Kontraste in einem gemeinsamen Modell geschätzt, so dass der Zugewinn hauptsächlich darin besteht, dass die Effekte effizienter geschätzt werden. Teilweise können, wie schon gesagt, erst damit aber auch bestimmte Alternativeneffekte richtig identifiziert werden. Eine weitere interessante Anwendung besteht darin, beste-

hende Analysen mit einem multinomialen Modell ohne fixed-effects mit dem femlogit-Modell zu schätzen. So analysiert z. B. Thurner (1998) den Einfluss der wahrgenommenen Positionen der Parteien auf verschiedenen politischen Dimensionen auf die Parteipräferenz. Hierbei verwendet er Daten einer Panelstudie (Forschungsgruppe Wahlen, Mannheim, Kaase, Klingemann, Küchler, Pappi, und Semetko 1991), entscheidet sich aber für eine reine Querschnittsanalyse. Einer der wesentlichen Aspekte seiner Analysen ist die Kontrolle für die langfristige Parteiidentifikation. Dabei ist Thurner natürlich auf die explizit gemessene Parteiidentifikation angewiesen. Dieses Vorgehen wird aus verschiedenen Gründen kritisiert (vgl. z. B. Jagodzinski und Kühnel 1990). Durch die Verwendung des femlogit-Modells auf diese Analyse kann man den Einfluss der subjektiven Policypositionen der Parteien auf die Parteipräferenz unter impliziter Kontrolle der langfristigen Parteibindung schätzen, ohne auf eine mangelhafte Messung des Konstrukts angewiesen zu sein. Zukünftige Arbeiten sollten sich Analysen dieser Art zuwenden.

10 Schluss

Das Hauptziel der Arbeit, nämlich die Umsetzung des multinomialen logistischen Regressionsmodells mit fixed-effects in Stata, konnte zunächst einmal erreicht werden. Im ersten Teil der Arbeit wurde das statistische Modell ausführlich in Gegenüberstellung mit den pooled-Modellen erläutert, und die Implementation des Modells in Stata dargestellt.

Der entscheidende Vorteil des femlogit-Modells ist, dass man keine Annahme über den Zusammenhang von unbeobachteten, zeitlich konstanten Variablen und den im Modell berücksichtigten unabhängigen Variablen getroffen werden müssen. Der Nachteil dieses Modell besteht darin, dass man zur Schätzung erstens die Annahme der seriellen Unabhängigkeit der Fehler über die Messzeitpunkte benötigt, die gerade bei langen Zeitreihen häufig verletzt ist. Zweitens benötigt man die IIA-Annahme, d. h. die Fehler müssen über die Alternativen hinweg unkorreliert sein. Darüber hinaus besteht bei diesem Modell das Problem, dass die Interpretation der Effekte gegenüber den Querschnitts-, pooled- und random-effects-Modellen eingeschränkt ist. Beim femlogit-Modell kann man keine Effekte auf die Wahrscheinlichkeiten interpretieren und ist auf die Interpretation auf die Odds-Ratio-Effekte beschränkt. Diese Interpretation ist erstens unanschaulich, und zweitens sind die Odds-Ratio-Effekte direkt von der willkürlich angenommen Varianz der Fehlerterme abhängig, was zum Problem der „neglected heterogeneity” führt.

Die Implementation hat, abgesehen von den Vor- und Nachteilen, die dem statistischen Modell selbst innewohnen, in der vorliegenden Fassung die Fähigkeit, dass sie als allgemein anwendbares Kommando analog wie das bereits vorliegende Kommando `clogit` für dichotome abhängige Variable mit fixed-effects anwendbar ist. Ferner ist die Spezifikation von linearen Beschränkungen voll integriert, so dass es möglich ist, neben den üblichen zeitlich veränderlichen Variablen über die Beobachtungseinheiten Variablen zu berücksichtigen, die über die Alternativen hinweg variieren. Daneben besteht bei der gegenwärtigen Fassung aber das Problem, dass die Schätzung sehr rechenintensiv ist. Die Rechenzeit steigt überexponentiell mit der Anzahl der Messzeitpunkte, so dass in der vorliegenden Umsetzung lange Zeitreihen gekürzt werden müssen, um die Rechenzeit in realisierbaren Rahmen zu halten. Gerade hier besteht noch Optimierungspotential für zukünftige Verbesserungen an der Implementation.

Das femlogit-Modell wurde im zweiten Teil der Arbeit auf zwei Replikationen angewendet. In der ersten Anwendung werden Analysen von Schröder (2010) zum Effekt der Fertilität auf das Erwerbstätigkeitsverhältnis von Frauen repliziert und durch die Verwendung des femlogit-Modells erweitert. Zunächst lassen sich die Befunde von Schröder vollständig replizieren. Bei Geburt eines Kindes sinkt das Odds, erwerbstätig statt nicht erwerbstätig zu sein, stark ab, unabhängig von Kinderzahl. Das stark gesunkene Odds der Erwerbstätigkeit gegenüber der Erwerbslosigkeit steigt mit steigendem Alter der jüngsten Kinder wieder an. Durch die Verwendung des femlogit-Modells erkennt man, dass die Effekte der Kinderzahl und des Alters des jüngsten Kindes über die Kontraste „erwerbstätig in Vollzeit vs. erwerbslos” und „erwerbstätig in Teilzeit vs. erwerbslos” unterschiedlich sind. Diese Unterschiede können mit den dichotomen Analysen nicht getrennt werden. Mit der differenzierten Analyse mit dem femlogit-Modell wird deutlich, dass das Odds der

Teilzeiterwerbstätigkeit gegenüber der Erwerbslosigkeit ebenso nach einer Geburt stark sinkt, aber mit dem Alter des jüngsten schneller ansteigt als das entsprechende Odds der Vollzeiterwerbstätigkeit gegenüber Erwerbslosigkeit.

Die zweite Anwendung ist die Replikation der Analyse von Kohler (2002) zum Effekt der Klassenzugehörigkeit auf die Parteipräferenz. Hier lassen sich weniger eindeutige Befunde identifizieren wie bei der der Replikation von Schröder. Zunächst einmal können die Analysen von Kohler trotz der gleichen Datenquelle und der Verwendung der von Kohler zur Verfügung gestellten Aufbereitungs- und Analysesyntax nicht vollständig reproduziert werden. Weiterhin sind die Befunde schon in den Originalanalysen sehr schwach und im Verhältnis zur theoretischen Erwartung uneinheitlich. In der Replikation treten die erwarteten Effekte teilweise etwas deutlicher zu Tage. Durch die Verwendung des femlogit-Modells können teilweise Unterschiede zwischen den Effekten auf die konservativen Parteien identifiziert werden, die der theoretischen Erwartung entsprechen. Insgesamt ist aber bei diesen Analysen kein großer Zugewinn durch das femlogit-Modell erkennbar. Schließlich werden die Analysen von Kohler mit den aktualisierten Daten gerechnet, jedoch erhält man auch hier nur sehr uneinheitliche Befunde. Der Hauptgrund der sehr schwachen Effekte in allen Analysen ist in der von Kohler gewählten Operationalisierung der unabhängigen Variablen zu sehen. Bei Kohler und damit auch bei den hier dargestellten Replikationen und erweiterten Analysen werden die unabhängigen Variablen als Übergangsereignisse und nicht als Zustände betrachtet. D. h. Kohler betrachtet nicht den Unterschied in der Parteipräferenz zwischen zwei Klassen, sondern zwischen den Zeitpunkten, in denen man die Klasse wechselt gegenüber den Zeitpunkten, in der man derselben Klasse angehört. Da erstens Wechselereignisse insgesamt selten sind, d. h. alle unabhängigen Variablen sehr schief verteilt sind, und da zweitens aus inhaltlicher Perspektive die eigentliche Wirkung nicht aus den Klassenübergängen, sondern aus der Klassenzugehörigkeit selbst folgen sollte, sind die Effekte insgesamt stark abgeschwächt.

Mit den Replikationen kann zwar gezeigt werden, dass das femlogit-Modell grundsätzlich anwendbar ist, und es können zumindest bei der Replikation von Schröder neue Erkenntnisse gewonnen werden, jedoch bleiben selbstverständlich viele Probleme bestehen. Erstens müssen bei beiden Analysen die Daten gekürzt werden, um die Rechenzeit in einem realisierbaren Rahmen zu halten. Zweitens bleiben selbst bei der Replikation von Schröder, wo die Ergebnisse grundsätzlich den theoretischen Erwartungen entsprechen, Zweifel übrig, da man keine Effekte auf die Wahrscheinlichkeiten, sondern nur Odds-Ratio-Effekte interpretieren kann. Drittens werden bei beiden Analysen lange Zeitreihen verwendet, so dass davon auszugehen ist, dass die serielle Korrelation nicht vernachlässigt werden kann. Viertens ist bei der Replikation von Kohlers Analysen zur Parteienpräferenz davon auszugehen, dass die IIA-Annahme verletzt ist.

Für zukünftige Arbeiten gibt es verschiedene Verbesserungsmöglichkeiten. Zunächst besteht bei der hier dargestellten Implementation des femlogit-Modells noch Optimierungspotential, wodurch die Rechenzeit kürzer werden sollte, so dass die Beschränkung der Messzeitpunkte verringert werden kann. Ob man in zukünftigen Versionen der Implementation aber die anderen Probleme des femlogit-Modells beheben kann, ist gegenwärtig nicht absehbar.

Die konkreten Analysen, die in dieser Arbeit gezeigt wurden, lassen sich aber unabhängig davon noch verbessern. Bei der Replikation von Schröder (2010) erscheint es sinnvoll, zusätzlich auch die Eigenschaften des Partners stärker zu berücksichtigen. Dies kann mit den hier verwendeten Daten des Familiensurvey 2000 nicht geleistet werden. Hinsichtlich der Informationen über den Partner stellt der SOEP eine mögliche Alternative dar, jedoch müssen hier hinsichtlich der Messgenauigkeit der biographischen Informationen Abstriche gemacht werden. Eine weitere Alternative sind prinzipiell die Daten des Deutschen Familienpanels pairfam (vgl. Huinink, Brüderl, Nauck, Walper, Castiglioni, und Feldhaus 2011), wobei hier jedoch erst zwei Wellen verfügbar sind. Die Replikationsanalysen zur Arbeit von Kohler (2002) lassen sich vor allem dadurch verbessern, dass man statt der Ereignisoperationalisierung die konventionelle Zustandsoperationalisierung verwendet. Weiterhin sollte man die von Kohler vermutete zeitliche verzögerte Wirkung der soziostrukturellen Variablen durch die Hinzunahme der zeitlich vorgelagerten Wellen wie üblich analysieren.

Schließlich besteht bei zukünftigen Analysen über multinomiale kategoriale abhängige Variablen prinzipiell die Möglichkeit, mit anderen statistischen Modellen manche Defizite des femlogit-Modells abzuschwächen, ohne gänzlich auf die Vorteile des Modells verzichten zu müssen. Beim femlogit-Modell erkauft man die Möglichkeit, auf Annahmen über den Zusammenhang zwischen der unbeobachteten Heterogenität *vollständig* verzichten zu können, hauptsächlich damit, dass man die Effekte nur noch sehr schwer interpretieren kann. Weiterhin benötigt man bei diesem Modell die Annahmen der seriellen Unabhängigkeit der Fehlerterme und der IIA-Annahme. Wie Wooldridge (2010, S. 608–625) zeigt, bietet das „correlated random effects"-Modell nach Mundlak (1978) bzw. Chamberlain (1980) unter Umständen einen günstigeren Kompromiss. Bei diesem Modell ist der Zusammenhang zwischen der unbeobachteten Heterogenität nicht völlig unbeschränkt, sondern auf eine lineare Korrelation beschränkt. Mit dieser Einschränkung gegenüber dem fixed-effects-Modell ist es möglich, dass man erstens die gemittelten AME- und ADE-Effektinterpretation verwenden kann. D. h. damit ist eine anschaulichere Interpretation auf der Ebene der Wahrscheinlichkeiten möglich, und weiterhin ist damit auch das Problem der „neglected heterogeneity" abgeschwächt (vgl. Mood 2010). Zweitens kann man mit diesem Modell die Annahme der seriellen Unabhängigkeit der Fehlerterme aufgeben. Es ist aber anzumerken, dass für dieses Modell für multinomiale kategoriale abhängige Variablen bisher noch keine Implementation in gängigen Softwarepaketen vorliegen. Zukünftige Arbeiten sollten sich daher zum Ziel setzen, dieses Modell erstens für ein geläufiges Softwarepaket umzusetzen und zweitens das Modell für Fragestellungen, wie sie in dieser Arbeit bearbeitet wurden, anzuwenden.

Teil III

Appendix

A Code

Die hier dargestellten Codezeilen für den Statabefehl `femlogit` und das Evaluator-Unterprogramm `femlogit_eval_gf2()` sind hier ungekürzt und im Original dargestellt. Das Evaluator-Unterprogramm ist als do-file dargestellt, mit die kompilierte Fassung der Funktion erzeugt wird.

A.1 Befehl `femlogit.ado`

femlogit.ado

```
*-------------------------------------------------------------------------------
*! version 0.7.2 06feb2012 17:14

program femlogit, eclass properties(svyb)
   version 11.0
   syntax varlist [if] [in], GRoup(varlist) [Baseoutcome(passthru) /*
   */ CONSTraints(numlist) DIFFicult]

   // sortpreserve
   tempvar sortsq
   quietly gen double `sortsq'=_n

   // missings
   marksample touse
   markout `touse' `varlist' `group', strok

   // remove collinear variables, handle basecategory (following mlogit.ado)
   quietly _rmcoll `varlist' if `touse', mlogit `baseoutcome'

   // Results stored in locals and matrix
   local varlist `"`r(varlist)'"'
   // outcome vector (baseoutcome inclusive)
   tempname out
   matrix `out' = r(out)
   // vector position of baseoutcome in vector out
   local ibase = r(ibaseout)
   // Number of outcomes (baseoutcome inclusive)
   local nout = r(k_out)
   // Error if only one outcome
   if (`nout' == 1) {
      error 148
   }

   // split depvar from varlist
   gettoken lhs rhs : varlist
   local nindeps `:list sizeof rhs'

   // matsize-check: Error if matsize to small
   if `nout'*`nindeps' > c(matsize) {
      error 908
   }

   // matrix out2eq for Mata
   tempname out2eq
   matrix `out2eq'=J(`nout',2,0)
   local j=1
   forvalues i=1/`nout' {
      matrix `out2eq'[`i',1]=`out'[1,`i']
      if `i'!=`ibase' {
         matrix `out2eq'[`i',2]= `j'
         local j=`j'+1
      }
   }

   // Ignore offending observations/groups.
```

```
56    // Taken from and adjusted clogit.ado version 1.6.14 19may2010
57    /* `vv' /// */
58    cap noi CheckGroups `lhs' `group' `touse' `rhs' /*`wgt', `offopt'*/
59    if _rc {
60       exit _rc
61    }
62    local rhs `"`r(varlist)'"'
63    local n `r(N)'
64    local ng `r(ng)'
65    local n_drop `r(n_drop)'
66    local ng_drop `r(ng_drop)'
67    // not supported
68    // local multiple `r(multiple)'
69    /* not supported
70    if !`r(useoffset)' {
71       local offopt
72    }
73    */
74
75    // "estimation" of baseline log.likelihood (=ll0 (inverse of number of
76    // permutations))
77    // handling model w/o indep. var.'s / baseline for LR-test
78    tempname ll0 v1 v2
79    sort `touse' `group' `lhs'
80    // Number of measurements $T_i$ for each observation unit $i$
81    quietly by `touse' `group': gen double `v1'=_n if `touse'
82    quietly by `touse' `group': replace `v1'=`v1'[_N] if `touse' & _n==1
83    quietly by `touse' `group': replace `v1'=. if `touse' & _n!=1
84    // $ln((T_i)!)$
85    quietly replace `v1'=lnfactorial(`v1') if `touse'
86    // Number of chosen outcomes $\delta_{y_{it}=j}$
87    quietly by `touse' `group' `lhs': gen double `v2'=_n if `touse'
88    quietly by `touse' `group' `lhs': replace `v2'=`v2'[_N] if `touse' & _n==1
89    quietly by `touse' `group' `lhs': replace `v2'=. if `touse' & _n!=1
90    // $ln((k_ij)!)$
91    quietly replace `v2'=lnfactorial(`v2') if `touse'
92    // $ln((k_i1)!)+...+ln((k_iJ)!)=ln((k_i1)!*...*(k_iJ)!)$
93    quietly by `touse' `group': replace `v2'=sum(`v2') if `touse'
94    quietly by `touse' `group': replace `v2'=`v2'[_N] if `touse' & _n==1
95    quietly by `touse' `group': replace `v2'=. if `touse' & _n!=1
96    // Number of permutations for each i = $|\Delta_i|$
97    quietly by `touse': replace `v1'=sum(`v1'-`v2') if `touse'
98    scalar `ll0'=-`v1'[_N]
99    drop `v1' `v2'
100
101   // init values (only if indep.vars given)
102   // inspired by clogit's "Check for initial values" (init values are coef's
103   // from pooled mlogit)
104   if `nindeps'>0 {
105      quietly mlogit `lhs' `rhs' if `touse', b(`=`out'[1,`ibase']')
106      tempname aux2 init
107      // matrix 1 row, (#indep.vars + constant) x (#outcomes) cols
108      matrix `aux2'=e(b)
109      // Cols of init matrix (1 row, #indep.vars x (outcomes-1) cols
110      local aux1=((colsof(`aux2')/`nout')-1)*(`nout'-1)
111      matrix `init'=J(1,`aux1',0)
112      // build init matrix
113      local i2=1
114      forvalues i=1/`nout' {
115         if `i'!=`ibase' { /* loop over outcomes except base category */
116            local j2=1
117            forvalues j=1/`=colsof(`aux2')/`nout'' {
118                /* loop over indep. variables + constant except const */
119               if `j'!=`=colsof(`aux2')/`nout'' {
120                  matrix `init'[1,`=((`i2'-1)*(colsof(`aux2')/`nout'-1))+ /*
121                  */ `j2'']=`aux2'[1,`=((`i'-1)*colsof(`aux2')/`nout')+`j'']
122                  local j2=`j2'+1
123               }
124            }
```

```
            local i2='i2'+1
         }
      }
      matrix drop 'aux2'
   }

   // Constraint handling for moptimize (supposedly not necessary with ml)
   // (only if indep. vars given)
   if 'nindeps'>0 & '"'constraints'"'!="" {
      // dummy beta-vector & V-matrix
      tempname b V T a C cnsmat
      matrix 'b'=J(1,'=('nout'-1)*'nindeps'',0)
      // Column names necessary for constraint matrix creation
      matrix roweq 'b'="_"
      matrix rownames 'b'="y1"
      forvalues i=1/'nout' {
         if 'i'!='ibase' {
            foreach var in 'rhs' {
               local eqlist 'eqlist' "'=abbrev(strtoname("':label ('lhs') /*
               */ '='out'[1,'i']'",1),12)'"
               local namelist 'namelist' 'var'
            }
         }
      }
      matrix coleq 'b'='eqlist'
      matrix colnames 'b'='namelist'
      matrix 'V'='b''*'b'

      // ereturn post
      quietly ereturn post 'b' 'V'

      // create matrices for transformation of reduced-form to unreduced-form
      // vectors/matrices
      noisily _makecns 'constraints'
      capture quietly local i=r(clist)==""
      // _rc==0 -> r(clist) string -> at least 1 constr correctly specified
      // _rc==109 -> r(clist) not string -> no constr correctly specified
      if _rc==0 { /* all fine, at least one constraint survived */
         matrix 'cnsmat'=e(Cns)
         quietly matcproc 'T' 'a' 'C'
      }
      if _rc!=0 { /* all constraints dropped */
         di as txt "(note: all constraints are dropped,"
         di as txt "       switched to unconstrained estimation,"
         di as txt "       check constraint specification.)"
         matrix 'cnsmat'=J(1,1,.)
         matrix 'T'=J(1,1,.)
         matrix 'a'=J(1,1,.)
         matrix 'C'=J(1,1,.)
         local constraints=""
      }
   }
   if 'nindeps'>0 & '"'constraints'"'==="" {
      tempname cnsmat T a C
      matrix 'cnsmat'=J(1,1,.)
      matrix 'T'=J(1,1,.)
      matrix 'a'=J(1,1,.)
      matrix 'C'=J(1,1,.)
   }

   // Sorting before evaluator call (if indep.vars given)
   if 'nindeps'>0 {
      sort 'touse' 'group'
   }

   // moptimize-call with mata-evaluator (if indep.vars given)
   if 'nindeps'>0 {
      // Most parts in mata
      di ""
```

```
194         mata: moptcall()
195
196         // Rest if results
197         // Scalars
198         ereturn scalar k_eq_model=`nout'-1
199         if "`constraints'"=="" {
200            ereturn scalar ll_0=`ll0'
201         }
202         ereturn scalar df_m=`=(`nout'-1)*`nindeps'-cond("`constraints'"!="", /*
203         */ colsof(`T'),0)'
204         if "`constraints'"=="" {
205            ereturn scalar chi2=2*(e(ll)-`ll0') /* LR chi2 */
206            ereturn scalar p=chi2tail(e(df_m),e(chi2))
207         }
208         if "`constraints'"!="" {
209            tempname chi2temp
210            mata: st_matrix(st_local("chi2temp"),st_matrix("e(b)")* /*
211            */ pinv(st_matrix("e(V)"))*st_matrix("e(b)")') /* Wald chi2 */
212            ereturn scalar chi2=`chi2temp'[1,1]
213            ereturn scalar p=chi2tail(e(df_m),e(chi2))
214         }
215         if "`n_drop'"!="" {
216            ereturn scalar N_drop=`n_drop'
217            ereturn scalar N_group_drop=`ng_drop'
218         }
219         if "`constraints'"=="" {
220            ereturn scalar r2_p=1-e(ll)/e(ll_0)
221         }
222         if "`constraints'"!="" {
223            ereturn scalar r2_p=.
224         }
225         ereturn scalar ibaseout=`ibase'
226         ereturn scalar baseout=`=`out'[1,`ibase']'
227         ereturn scalar k_out=`nout'
228         // Macros
229         ereturn local crittype="log likelihood"
230         ereturn local title="Fixed-effects multinomial logistic regression"
231         ereturn local vce="oim"
232         if "`constraints'"=="" {
233            ereturn local chi2type="LR"
234         }
235         if "`constraints'"!="" {
236            ereturn local chi2type="Wald"
237         }
238         ereturn local group="`group'"
239         local t1 "`:coleq e(b)'"
240         forvalues i=1/`=(`nout'-1)*`nindeps'' {
241            if mod(`i',`nindeps')==1 {
242               local t2 `t2' `:word `i' of `t1''
243            }
244         }
245         ereturn local eqnames="`t2'"
246         ereturn local cmd="femlogit"
247         ereturn local cmdline="femlogit `0'"
248         // Matrices
249         ereturn matrix out=`out', copy
250
251         // Display results table (adapted from mlogit)
252         noisily _coef_table_header
253         if "`constraints'"=="" { /* single line skip for aesthetic reasons */
254            noisily di ""
255         }
256         tempname T
257         quietly .`T'   = ._b_table.new
258         noisily .`T'.display_titles, depname("`lhs'") cnsreport
259         local j=1
260         forval i = 1/`nout' {
261            if `i' == `ibase' {
262               .`T'.sep
```

```
            .`T'.display_comment "`=abbrev(strtoname("`:label (`lhs') /*
            */ `=`out'[1,`ibase']'",1),12)'", comment("  (base outcome)")
         }
         else {
            .`T'.display_eq #`j',
            local j=`j'+1
         }
      }
      .`T'.finish
      if (!missing(e(rc)) & e(rc) != 0) error e(rc)

      _prefix_footnote
   }


   // display of results if NO INDEP.VARS GIVEN
   if `nindeps'==0 {
      // Fake ml step
      noisily di "" /* single line skip */
      noisily display as text /*
         */ "Iteration 0:   log likelihood = " as result `ll0'

      // silent clogit (this can be simplified)
      quietly tempvar av
      quietly gen `av'=`lhs'!=`=`out'[1,`ibase']' if `touse'
      quietly clogit `av', group(`group')

      // fill up estimates
      ereturn scalar N=`n'
      ereturn scalar ll_0=`ll0'
      ereturn scalar ll=`ll0'
      ereturn scalar df_m=0
      ereturn scalar chi2=0
      ereturn scalar r2_p=0
      ereturn local chi2type="LR"
      ereturn local group="`group'"
      ereturn local properties="b V"
      ereturn local predict="predict"
      ereturn local crittype="log likelihood"
      ereturn local cmd="femlogit"
      ereturn local depvar="`lhs'"
      ereturn local cmdline="femlogit `0'"
      ereturn repost, esample(`touse')

      // Display results table
      noisily _coef_table_header, /*
      */ title(Fixed-effects multinomial logistic regression)
      noisily di "" /* single line skip */
      tempname T
      quietly .`T'   = ._b_table.new
      noisily .`T'.display_titles, depname("`lhs'")
      noisily .`T'.sep
      noisily .`T'.finish
   }

   // Reorder
   sort `sortsq'
end

// Taken from clogit.ado version 1.6.14  19may2010
program CheckGroups, rclass
   local caller = _caller()
   version 8.2, missing
   syntax varlist(fv) [fw iw] [, offset(varname)]
   gettoken y varlist : varlist
   gettoken group varlist : varlist
   gettoken touse xvars : varlist

   sort `touse' `group'
```

```
332
333     // Check weights.
334     if "`weight'" != "" {
335        tempvar w
336        qui gen double `w'`exp' if `touse'
337        cap by `touse' `group':assert `w'==`w'[1] if `touse'
338        if _rc {
339           error 407
340        }
341        if "`weight'"=="fweight" {
342           local freq `w'
343        }
344     }
345
346     // Check at least one good group.
347     cap by `touse' `group': assert `y'==`y'[1] if `touse'
348     if !_rc {
349        di as txt "outcome does not vary in any group"
350        exit 2000
351     }
352
353     /* does not work directly with femlogit, "sum(`y')" works only with dummy
354     // Check for multiple positive outcomes within groups.
355
356     tempvar sumy
357     qui by `touse' `group': gen double `sumy' = cond(_n==_N, ///
358        sum(`y'), .) if `touse'
359     qui count if `sumy' > 1 & `sumy' < .
360     if `r(N)' {
361        di as txt "note: multiple positive outcomes within " _c
362        di as txt "groups encountered."
363        local multiple multiple
364     }
365     */
366
367     // Delete groups where outcome doesn't vary.
368     CountObsGroups `touse' `group' `freq'
369     local n_orig = r(n)
370     local ng_orig = r(ng)
371
372     tempvar varies rtouse
373     qui by `touse' `group': gen byte `varies' = cond(_n==_N, ///
374        sum(`y'!=`y'[1]), .) if `touse'
375     qui by `touse' `group': gen byte `rtouse' = (`varies'[_N]>0) & `touse'
376     qui replace `touse' = `rtouse'
377     sort `touse' `group'
378
379     CountObsGroups `touse' `group' `freq'
380     local n = r(n)
381     local ng = r(ng)
382
383     if `n' < `n_orig' {
384        if `ng_orig'-`ng' > 1 {
385           local s s
386        }
387        di as txt "note: " `ng_orig'-`ng' " group`s' (" _c
388        di as txt `n_orig'-`n' _c
389        di as txt " obs) dropped because of all positive or"
390        di as txt "      all negative outcomes."
391        local ng_drop   = `ng_orig' - `ng'
392        local n_drop    = `n_orig' - `n'
393     }
394
395     // Check that each xvar varies in at least 1 group.
396     if `""`xvars'"' != "" {
397        fvexpand `xvars'
398        local xvars "`r(varlist)'"
399        foreach v of local xvars {
400           _ms_parse_parts `v'
```

```
401            if r(type) == "variable" & !r(omit) {
402                cap by `touse' `group':   ///
403               assert `v'==`v'[1] if `touse'
404                if !_rc {
405               di as txt "note: `v' omitted because of no "_c
406               di as txt "within-group variance."
407               if `caller' < 11 {
408                  local v
409               }
410               else   local v o.`v'
411                }
412            }
413            local xs `xs' `v'
414         }
415      }
416
417      // Check that offset varies in at least 1 group.
418      local useoffset 0
419      if "`offset'" != "" {
420         cap by `touse' `group': assert `offset'==`offset'[1] if `touse'
421         if !_rc {
422            di as txt "note: offset `offset' omitted " _c
423            di as txt "because of no " _c
424            di as txt "within-group variance."
425         }
426         else {
427            local useoffset 1
428         }
429      }
430
431      return local multiple `multiple'
432      return local varlist `xs'
433      return local useoffset `useoffset'
434      return scalar N = `n'
435      return scalar ng = `ng'
436      if `:length local n_drop' {
437         return scalar n_drop = `n_drop'
438         return scalar ng_drop = `ng_drop'
439      }
440   end
441
442   program CountObsGroups, rclass
443      args touse group freq
444
445      tempvar i
446      if "`freq'" == "" {
447         qui count if `touse'
448         return scalar n = r(N)
449         qui by `touse' `group': gen byte `i' = _n==1 & `touse'
450         qui count if `i'
451         return scalar ng = r(N)
452      }
453      else {
454         qui summ `freq' if `touse', meanonly
455         return scalar n = r(sum)
456         qui by `touse' `group': gen double `i' = (_n==1&`touse')*`freq'
457         qui summ `i' if `touse', meanonly
458         return scalar ng = r(sum)
459      }
460   end
461
462   mata:
463   mata set matastrict on
464   void moptcall(){
465      // declare variables
466      transmorphic matrix M
467      real scalar i,j
468
469      // Initialize moptimize
```

```
    M=moptimize_init()

    // Definition of problem
    moptimize_init_touse(M,st_local("touse")) /* sample-indicator */
    moptimize_init_depvar(M,1,st_local("lhs")) /* dep. var. */
    // # of eq's (w/o ref.cat.!)
    moptimize_init_eq_n(M,strtoreal(st_local("nout"))-1)
    j=1
    /* indep. vars for each eq (loop over all outcomes) */
    for(i=1;i<=strtoreal(st_local("nout"));i++) {
       if (i!=strtoreal(st_local("ibase"))) { /* take out ref.cat.*/
          // 'rhs' for each equation!
          moptimize_init_eq_indepvars(M,j,st_local("rhs"))
          moptimize_init_eq_cons(M,j,"off") /* no constant */
          // dep.var has value label
          if (st_varvaluelabel(st_local("lhs"))!="") {
             moptimize_init_eq_name(M,j,strtoname(st_vlmap( /*
             */ st_varvaluelabel(st_local("lhs")),/*
             */ st_matrix(st_local("out"))[1,i])!="" ? /*
             */ st_vlmap(st_varvaluelabel(st_local("lhs")), /*
             */ st_matrix(st_local("out"))[1,i]) : /*
             */ strofreal(st_matrix(st_local("out"))[1,i])))
          }
          // dep.var has no value label
          if (st_varvaluelabel(st_local("lhs"))=="") {
             moptimize_init_eq_name(M,j,strtoname(strofreal( /*
             */ st_matrix(st_local("out"))[1,i])))
          }
          j=j+1
       }
    }
    if (st_local("constraints")!="") { /* constraint-matrix */
       moptimize_init_constraints(M,st_matrix(st_local("cnsmat")))
    }
    moptimize_init_technique(M,"nr")
    if (st_local("difficult")!="") {
       moptimize_init_singularHmethod(M, "hybrid")
    }
    // init-value
    // indep. vars for each eq (loop over all outcomes except ref.cat.)
    for(i=1;i<=strtoreal(st_local("nout"))-1;i++) {
       moptimize_init_eq_coefs(M,i,rowshape(st_matrix(st_local("init")), /*
       */ strtoreal(st_local("nout"))-1)[i,.])
    }
    moptimize_init_search(M,"off")
    moptimize_init_valueid(M,"log likelihood")

    // Evaluator
    moptimize_init_evaluator(M,&femlogit_eval_gf2())
    moptimize_init_evaluatortype(M,"gf2")

    // Maximize
    moptimize(M)

    // Post results to Stata in e() macros
    st_eclear()
    moptimize_result_post(M)

    // Preparation for posting of all results in e() macros
    st_numscalar("e(ll)",moptimize_result_value(M))
}
end
```

A.2 Evaluator-Unterprogramm `femlogit_eval_gf2()`

femlogit_eval_gf2()

```
*-------------------------------------------------------------------------------
*! version 0.7.2 06feb2012 11:14

mata set matastrict on
// gf2-type
void femlogit_eval_gf2(transmorphic scalar ML, real scalar todo, /*
   */ real rowvector b, real colvector lnfj, real matrix S, /*
   */ real matrix H) {
   // necessary things to be done before in Stata
   // sort touse group
   // matrix out2eq=(...)
   // matrix rvm=(...)

   // declare variables
   real colvector lnli, touse, id, yi, d // touse kann potentiell raus
   real matrix gi, Hi, panelinfo, Xi, out2eq, X, E
   real scalar N, M, J, i, A, B, j, m, Z1
   real rowvector C, D, permuteinfo, Z2

   // get things from Stata
   st_view(touse=.,.,st_local("touse"))
   st_view(id=.,.,st_local("group"),st_local("touse")) // sort id!
   st_view(X=.,.,st_local("rhs"),st_local("touse"))

   // moptimize_util_depvar(M,1)
   out2eq=st_matrix(st_local("out2eq"))

   // derived information
   J=rows(out2eq)
   M=cols(X)
   panelinfo=panelsetup(id,1) // sort id!
   N=panelstats(panelinfo)[1]

   // init lnli, gi, Hi
   lnli=J(N,1,0)
   if (todo>0) {
      gi=J(N,(J-1)*M,0)
      if (todo==2) {
         Hi=J(N,((J-1)*M)^2,0) // panel-wise Hessian component in wide-format
      }
   }

   // calculate lnli, gi, Hi
   for(i=1;i<=N;i++) { // loop over panels
      // create panel-wise variables (only one call per panel)
      yi=moptimize_util_depvar(ML,1)[|panelinfo[i,1]\panelinfo[i,2]|]
      Xi=X[|panelinfo[i,1],.\panelinfo[i,2],.|]

      // init major auxiliary variables (A,B,C,D,E)
      A=0
      B=0
      if (todo>0) {
         C=J(1,(J-1)*M,0)
         D=J(1,(J-1)*M,0)
         if (todo==2) {
            E=J((J-1)*M,(J-1)*M,0)
         }
      }

      // calculate A,C
      for(j=1;j<=J;j++) { // loop over outcomes
         if (out2eq[j,2]!=0) { // exclude base outcome
            A=A+quadcolsum((yi:==out2eq[j,1]):* /*
            */ ((Xi*(colshape(b,M)'))[.,out2eq[j,2]]))
            if (todo>0) {
```

```
                for(m=1;m<=M;m++) { // loop over indep. vars
                   C[1,(out2eq[j,2]-1)*M+m]=quadcolsum((yi:==out2eq[j,1]):* /*
                   */ (Xi[.,m]))
                }
             }
          }
       }

       // calculate B,D,E
       // generate Delta_i=Set of permutations of y_i
       permuteinfo=cvpermutesetup(yi)
       // loop over permutations of y_i (d_i in Delta_i)
       while((d=cvpermute(permuteinfo))!=J(0,1,.)) {
          // init minor auxiliary variables
          Z1=0
          if (todo>0) {
             Z2=J(1,(J-1)*M,0)
          }

          // calculate Z1,Z2
          for(j=1;j<=J;j++) { // loop over outcomes
             if (out2eq[j,2]!=0) { // exclude base outcome
                Z1=Z1+quadcolsum((d:==out2eq[j,1]):*((Xi* /*
                */ (colshape(b,M)'))[.,out2eq[j,2]]))
                if (todo>0) {
                   for(m=1;m<=M;m++) {
                      Z2[1,(out2eq[j,2]-1)*M+m]= /*
                      */ quadcolsum((d:==out2eq[j,1]):*(Xi[.,m]))
                   }
                }
             }
          }
          Z1=exp(Z1)

          // fill up B,D,E with minor aux. var's
          B=B+Z1
          if (todo>0) {
             D=D+Z2:*Z1
             if (todo==2) {
                E=E+(quadcross(Z2,Z2)):*Z1
             }
          }
       }

       // fill up lnli,gi,Hi with major aux. var's A,B,C,D,E
       lnli[i]=A-ln(B)
       if (todo>0) {
          gi[i,.]=C-D:/B
          if (todo==2) {
             // fill up Hi in wide format
             Hi[i,.]=rowshape(((quadcross(D,D)):/(B^2))-(E:/B),1)
          }
       }
    }

    // fill up lnf,h,H with panel-wise lnli,gi,Hi
    lnfj=lnli
    if (todo>0) {
       S=gi
       if (todo==2) {
          H=colshape(quadcolsum(Hi),((J-1)*M))
       }
    }
}
```

B Einfluss der Klassenzugehörigkeit auf die Parteipräferenz

Tabelle B.1: Replikation der binären fixed-effects-logit-Modelle von Kohler (2002) (gewichtet)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Administrative Dienstklasse → Selbständig	$-0{,}80^{***}$ $(-278{,}70)$	$-0{,}02^{***}$ $(-12{,}36)$	$0{,}08^{***}$ $(20{,}40)$
Administrative Dienstklasse → Mischt./Experte	$-0{,}07^{***}$ $(-69{,}19)$	$0{,}15^{***}$ $(162{,}29)$	$0{,}20^{***}$ $(119{,}35)$
Administrative Dienstklasse → Arbeiter/Soz. Dienste	$0{,}26^{***}$ $(165{,}57)$	$0{,}29^{***}$ $(146{,}65)$	$-0{,}99^{***}$ $(-306{,}29)$
Selbständig → Admin. Dienstkl.	$0{,}17^{***}$ $(69{,}42)$	$-0{,}03^{***}$ $(-15{,}38)$	$0{,}16^{***}$ $(39{,}46)$
Selbständig → Mischt./Experte	$0{,}95^{***}$ $(371{,}90)$	$-0{,}23^{***}$ $(-149{,}89)$	$0{,}30^{***}$ $(90{,}30)$
Selbständig → Arbeiter/Soz. Dienste	$-0{,}32^{***}$ $(-175{,}34)$	$0{,}53^{***}$ $(292{,}30)$	$-0{,}81^{***}$ $(-319{,}94)$
Mischtyp/Experte → Admin. Dienstkl.	$-0{,}15^{***}$ $(-148{,}34)$	$0{,}02^{***}$ $(17{,}46)$	$-0{,}15^{***}$ $(-87{,}94)$
Mischtyp/Experte → Selbständig	$-0{,}77^{***}$ $(-352{,}22)$	$-0{,}00^{+}$ $(-1{,}65)$	$0{,}11^{***}$ $(34{,}08)$
Mischtyp/Experte → Arbeiter/Soz. Dienste	$-0{,}17^{***}$ $(-240{,}05)$	$-0{,}19^{***}$ $(-242{,}62)$	$0{,}32^{***}$ $(236{,}82)$
Arbeiter/Soziale Dienste → Admin. Dienstkl.	$-0{,}02^{***}$ $(-11{,}37)$	$-0{,}20^{***}$ $(-89{,}24)$	$0{,}79^{***}$ $(195{,}18)$
Arbeiter/Soziale Dienste → Selbständig	$0{,}04^{***}$ $(23{,}14)$	$0{,}34^{***}$ $(200{,}90)$	$0{,}27^{***}$ $(109{,}18)$
Arbeiter/Soziale Dienste → Mischt./Experte	$0{,}12^{***}$ $(167{,}71)$	$0{,}10^{***}$ $(132{,}32)$	$-0{,}51^{***}$ $(-413{,}82)$
Eintritt in Arbeitslosigkeit	$0{,}16^{***}$ $(222{,}01)$	$-0{,}26^{***}$ $(-295{,}41)$	$-0{,}01^{***}$ $(-6{,}84)$
Austritt aus Arbeitslosigkeit	$-0{,}18^{***}$ $(-258{,}90)$	$0{,}20^{***}$ $(222{,}95)$	$-0{,}02^{***}$ $(-19{,}14)$
Schulabschluss	$0{,}12^{***}$ $(122{,}52)$	$-0{,}06^{***}$ $(-64{,}20)$	$0{,}46^{***}$ $(366{,}61)$
Hochschulabschluss	$0{,}17^{***}$ $(175{,}96)$	$0{,}05^{***}$ $(39{,}84)$	$-0{,}38^{***}$ $(-392{,}58)$
Berufsausbildungs-abschluss	$0{,}11^{***}$ $(182{,}49)$	$0{,}02^{***}$ $(30{,}45)$	$-0{,}04^{***}$ $(-48{,}92)$
Beginn Erwerbsleben	$0{,}36^{***}$ $(629{,}15)$	$0{,}15^{***}$ $(248{,}92)$	$0{,}29^{***}$ $(325{,}70)$
Erste Hochzeit	$0{,}39^{***}$ $(472{,}09)$	$0{,}10^{***}$ $(110{,}02)$	$-0{,}03^{***}$ $(-22{,}99)$
Geburt des ersten Kindes	$0{,}21^{***}$ $(259{,}29)$	$0{,}12^{***}$ $(132{,}45)$	$0{,}31^{***}$ $(245{,}98)$

Tabelle B.1: (Fortsetzung)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Geburt des zweiten Kindes	$-0{,}03^{***}$ $(-47{,}84)$	$-0{,}11^{***}$ $(-169{,}11)$	$0{,}07^{***}$ $(77{,}04)$
Personen	5 750	4 926	1 696
Messzeitpunkte[a]	49 472	41 442	13 880
χ^2/(df)[a]	6 885 489,8 (33)	8 886 658,2 (33)	3 328 470,1 (33)
R^2_{MF}	0,023	0,030	0,036

*: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. Nicht dargestellte Kontrollvariablen: jahresweise Wellendummies. Schätzung unter Berücksichtigung von Design-, Nonresponse- und Attritiongewichten.

Tabelle B.2: Replikation der binären fixed-effects-logit-Modelle von Kohler (2002) mit SOEP-Daten 1984–2010 (ungewichtet)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Administrative Dienstklasse → Selbständig	−0,04 (−0,71)	0,03 (0,62)	−0,02 (−0,19)
Administrative Dienstklasse → Mischt./Experte	0,12 (1,56)	0,19* (2,52)	−0,04 (−0,38)
Administrative Dienstklasse → Arbeiter/Soz. Dienste	0,01 (0,24)	−0,09 (−1,55)	−0,11 (−1,07)
Selbständig → Admin. Dienstkl.	0,09+ (1,69)	0,07 (1,38)	0,04 (0,45)
Selbständig → Mischt./Experte	0,07 (0,67)	0,12 (1,29)	−0,29+ (−1,65)
Selbständig → Arbeiter/Soz. Dienste	0,21*** (3,68)	−0,11+ (−1,87)	−0,30* (−2,26)
Mischtyp/Experte → Admin. Dienstkl.	−0,16* (−2,27)	0,01 (0,18)	0,25* (2,28)
Mischtyp/Experte → Selbständig	0,09 (0,83)	−0,20* (−2,08)	−0,40* (−2,30)
Mischtyp/Experte → Arbeiter/Soz. Dienste	0,01 (0,13)	−0,03 (−0,32)	−0,05 (−0,35)
Arbeiter/Soziale Dienste → Admin. Dienstkl.	−0,02 (−0,41)	−0,10+ (−1,93)	0,05 (0,52)
Arbeiter/Soziale Dienste → Selbständig	−0,17** (−3,26)	0,17** (2,84)	−0,34** (−2,83)
Arbeiter/Soziale Dienste → Mischt./Experte	0,07 (0,85)	−0,10 (−1,20)	−0,01 (−0,06)
Eintritt in Arbeitslosigkeit	0,06*** (3,35)	−0,09*** (−4,65)	0,15*** (4,44)
Austritt aus Arbeitslosigkeit	−0,09*** (−3,59)	−0,02 (−0,61)	−0,12** (−2,58)
Schulabschluss	0,53*** (9,80)	0,47*** (7,78)	0,38*** (5,06)
Hochschulabschluss	0,14** (2,87)	0,16** (3,02)	0,14* (2,35)
Berufsausbildungs-abschluss	0,06+ (1,88)	0,20*** (6,38)	0,08 (1,59)
Beginn Erwerbsleben	0,20*** (3,58)	0,09 (1,60)	0,10 (1,38)
Erste Hochzeit	0,07 (1,42)	0,18*** (3,31)	−0,11 (−1,40)
Geburt des ersten Kindes	0,27*** (6,18)	0,07 (1,46)	0,36*** (4,92)

Tabelle B.2: (Fortsetzung)

	SPD	CDU/CSU, FDP	Grüne
	β/(z)	β/(z)	β/(z)
Geburt des zweiten Kindes	$-0{,}02$ $(-0{,}51)$	$0{,}06^{+}$ $(1{,}74)$	$0{,}09^{+}$ $(1{,}77)$
Personen	11 796	11 094	3 575
Messzeitpunkte	149 792	136 592	45 072
χ^2/(df)	3 275,4 (46)	1 767,1 (46)	924,2 (46)
R^2_{MF}	0,029	0,018	0,031

*: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. Nicht dargestellte Kontrollvariablen: jahresweise Wellendummies.

Literaturverzeichnis

Agresti, Alan. 1990. *Categorical data analysis*. New York, NY: Wiley.

Aisenbrey, Silke, Marie Evertsson, und Daniela Grunow. 2009. Is There a Career Penalty for Mothers' Time Out? A Comparison of Germany, Sweden and the United States. *Social Forces* 88: 573–605.

Allison, Paul D. 1994. Using Panel Data to Estimate the Effects of Events. *Sociological Methods & Research* 23: 174–199.

Allison, Paul D. 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 28: 186–208.

Allison, Paul D. 2009. *Fixed effects regression models*. Thousand Oaks, CA et al.: Sage Publications.

Amemiya, Takeshi. 1981. Qualitative response models: A survey. *Journal of Economic Literature* 19: 1483–1536.

Amemiya, Takeshi. 1985. *Advanced econometrics*. Oxford: Blackwell.

Andrew, Mark. 2004. A Permanent Change in the Route to Owner Occupation? *Scottish Journal of Political Economy* 51: 24–48. Andrew, M.

Angrist, Joshua D., und William N. Evans. 1998. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review* 88: 450–477.

Apps, Patricia, und Ray Rees. 2005. Gender, Time Use, and Public Policy over the Life Cycle. *Oxford Review of Economic Policy* 21: 439–461.

Arellano, Manuel. 2003. *Panel data econometrics*. Oxford, New York, NY: Oxford University Press.

Baltagi, Badi H. 1995. *Econometric analysis of panel data*. Chichester, New York, NY: Wiley Press.

Baltagi, Badi H. 2001. *Econometric analysis of panel data*. 2. Aufl.. Chichester, New York, NY: Wiley Press.

Baltagi, Badi H. 2005. *Econometric analysis of panel data*. 3. Aufl.. Chichester, New York, NY: Wiley Press.

Baum, Christopher F. 2009. *An introdution to Stata programming*. College Station, TX: Stata Press.

Beck, Nicklaus, Josef Brüderl, und Michael Woywode. 2008. Momentum or deceleration? Theoretical and methodological reflections on the analysis of organizational change. *Academy of Management Journal* 51: 413–435.

Becker, Gary S. 1994. *A Treatise on the Family*. 3. Aufl.. Cambridge, MA: Harvard University Press.

Ben-Akiva, Moshe E., und Stephen R. Lerman. 1994. *Discrete choice analysis: Theory and application to travel demand*. 6. Aufl.. Cambridge, MA, London: MIT Press.

Berger, Lawrence M., und Jane Waldfogel. 2004. Maternity leave and the employment of new mothers in the United States. *Journal of Population Economics* 17: 331–349.

Berninger, Ina. 2009. Which family policies support the labour market participation of mothers? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61: 355–385.

Best, Henning, und Christof Wolf. 2010. Logistische Regression. In *Handbuch der sozialwissenschaftlichen Datenanalyse*, Hrsg. Christof Wolf, und Henning Best. Wiesbaden: VS Verlag, 827–854.

Bien, Walter, und Jan H. Marbach, Hrsg.. 2003. *Partnerschaft und Familiengründung: Ergebnisse der dritten Welle des Familien-Survey*. Opladen: Leske + Budrich.

Bien, Walter, und Richard Rathgeber, Hrsg.. 2000. *Die Familie in der Sozialberichterstattung: Ein europäischer Vergleich*. Opladen: Leske + Budrich.

Börsch-Supan, Axel. 1987. *Econometric analysis of discrete choice: With applications on the demand for housing in the U.S. and West-Germany*. Berlin et al.: Springer Verlag.

Börsch-Supan, Axel. 1990. Panel data analysis of housing choices. *Regional science and urban economics* 20: 65–82.

Börsch-Supan, Axel, und Henry O. Pollakowski. 1990. Estimating housing consumption adjustments from panel data. *Journal of urban economics* 27: 131–150.

Brambor, Thomas, William Roberts Clark, und Matt Golder. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14: 63–82.

Bronars, Stephen G., und Jeff Grogger. 1994. The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment. *The American Economic Review* 84: 1141–1156.

Bronstein, Ilja N., Konstantin A. Semendjajew, Gerhard Musiol, und Heiner Mühlig. 2001. *Taschenbuch der Mathematik*. 5. Aufl.. Thun, Frankfurt am Main: Harri Deutsch Verlag.

Brüderl, Josef. 2000. Regressionsverfahren in der Bevölkerungswissenschaft. In *Handbuch der Demographie 1: Modelle und Methoden*, Hrsg. Ulrich Müller, Bernhard Nauck, und Andreas Diekmann. Berlin et al.: Springer Verlag, 589–642.

Brüderl, Josef. 2010. Kausalanalyse mit Paneldaten. In *Handbuch der sozialwissenschaftlichen Datenanalyse*, Hrsg. Christof Wolf, und Henning Best. Wiesbaden: VS Verlag, 963–994.

Buchholz, Sandra, und Daniela Grunow. 2006. Women's employment in West Germany. In *Globalization, uncertainty and women's careers*. Edward Elgar, 61–83.

Budig, Michelle J. 2003. Are women's employment and fertility histories interdependent? An examination of causal order using event history analysis. *Social Science Research* 32: 376–401.

Cáceres-Delpiano, Julio. 2012. Can We Still Learn Something From the Relationship Between Fertility and Mother's Employment? Evidence From Developing Countries. *Demography* 49: 151–174.

Calhoun, Charles A. 1994. The Impact of Children on the Labor Supply of Married-Women: Comparative Estimates from European and United-States Data. *European Journal of Population* 10: 293–318.

Cameron, A. Colin, und Pravin K. Trivedi. 2009. *Microeconometrics: Methods and applications*. 8. Aufl.. Cambridge et al.: Cambridge University Press.

Cameron, A. Colin, und Pravin K. Trivedi. 2010. *Microeconometrics using Stata*. 2. Aufl.. College Station, TX: Stata Press.

Campbell, Angus, Philip E. Converse, Warren E. Miller, und Donald E. Stokes. 1960. *The American Voter*. New York, NY: Wiley.

Chamberlain, Gary. 1980. Analysis of Covariance with Qualitative Data. *Review of Economic Studies* 57: 225–238.

Charles, Maria, Marlis Buchmann, Susan Halebsky, Jeanne M. Powers, und Marisa M. Smith. 2001. The Context of Women's Market Careers: A Cross-National Study. *Work and Occupations: An International Sociological Journal* 28: 371–396.

Cochran, William G. 1977. *Sampling techniques*. 3. Aufl.. New York, NY et al.: John Wiley & Sons.

Desai, Sonalde, und Linda J. Waite. 1991. Women's Employment During Pregnancy and After the First Birth: Occupational Characteristics and Work Commitment. *American Sociological Review* 56: 551–566.

Deutsches Jugendinstitut (DJI), München. 2003. Wandel und Entwicklung familialer Lebensformen: 3. Welle (Familiensurvey). GESIS Datenarchiv, Köln. ZA3920 Datenfile Version 1.0.0, doi:10.4232/1.3920.

Diekmann, Andreas. 2007. *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. 18. Aufl.. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag.

Drobnič, Sonja. 2000. The effects of children on married and lone mothers' employment in the United States and (West) Germany. *European Sociological Review* 16: 137–157.

Drobnič, Sonja, Hans-Peter Blossfeld, und Götz Rohwer. 1999. Dynamics of women's employment patterns over the family life course: A comparison of the United States and Germany. *Journal of Marriage and Family* 61: 133–146.

Ebenstein, Avraham. 2009. When is the Local Average Treatment Close to the Average? Evidence from Fertility and Labor Supply. *The Journal of Human Resources* 44: 955–975.

Erikson, Robert, und John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.

Even, William E. 1987. Career Interruptions Following Childbirth. *Journal of Labor Economics* 5: 255–277.

Felmlee, Diane H. 1993. The dynamic interdependence of women's employment and fertility. *Social Science Research* 22: 333–360.

Forschungsgruppe Wahlen, Mannheim, Max Kaase, Hans-Dieter Klingemann, Manfred Küchler, Franz U. Pappi, und Holli A. Semetko. 1991. Wahlstudie 1990 (Panelstudie). GESIS Datenarchiv, Köln. ZA1919 Datenfile Version 1.0.0, doi:10.4232/1.1919.

Fouarge, Didier, Anna Manzoni, Ruud Muffels, und Ruud Luijkx. 2010. Childbirth and cohort effects on mothers' labour supply: a comparative study using life history data for Germany, the Netherlands and Great Britain. *Work, Employment and Society* 24: 487–507.

Franz, Wolfgang. 1985. An Economic Analysis of Female Work Participation, Education, and Fertility: Theory and Empirical Evidence for the Federal Republic of Germany. *Journal of Labor Economics* 3: 218–234.

Frenette, Marc. 2011. How does the stork delegate work? Childbearing and the gender division of paid and unpaid labour. *Journal of Population Economics* 24: 895–910.

Funk, Walter. 1993. *Determinanten der Erwerbsbeteiligung von Frauen im internationalen Vergleich: Eine Sekundäranalyse des ISSP 1988 für die Bundesrepublik Deutschland, die USA und Australien.* Frankfurt et al.: Peter Lang Verlag.

Gerfin, Michael. 1996. Parametric and semi-parametric estimation of the binary response model of labour market participation. *Journal of Applied Econometrics* 11: 321–339.

GESIS - Leibniz-Institut für Sozialwissenschaften. 2003. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2000 (CAPI-Version). GESIS Datenarchiv, Köln. ZA3451 Datenfile Version 1.0.0, doi:10.4232/1.3451.

Giannelli, Gianna Claudia. 1996. Women's transitions in the labour market: A competing risks analysis on German panel data. *Journal of Population Economics* 9: 287–300.

Gould, William, Jeffrey Pitblado, und Brian Poi. 2010. *Maximum likelihood estimation with Stata.* 4. Aufl.. College Station, TX: Stata Press.

Greene, William H. 2000. *Econometric analysis.* 4. Aufl.. Upper Saddle River, NJ: Prentice-Hall.

Grunow, Daniela, Heather Hofmeister, und Sandra Buchholz. 2006. Late 20th-century persistence and decline of the female homemaker in Germany and the United States. *International Sociology* 21: 101–131.

Haisken-De New, John P., und Joachim R. Frick. 1998. *DTC: Desktop Companion to the German Socio-Economic Panel Study (GSOEP).* Berlin. Version 2.2.

Halaby, Charles N. 2004. Panel models in sociological research: Theory into practice. *Annual Review of Sociology* 30: 507–544.

Heckman, James J. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42: 679–694.

Heckman, James J. 1981. Heterogeneity and State Dependence. In *Studies in Labor Markets*, Hrsg. Sherwin Rosen. Chicago, IL: University of Chicago Press, 91–140.

Henkens, Kène, Yolanda Grift, und Jacques Siegers. 2002. Changes in female labour supply in The Netherlands 1989-1998: The case of married and cohabiting women. *European Journal of Population* 18: 39–57.

Hoch, Irving. 1962. Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data. *Econometrica* 30: 34–53.

Hofbauer, Hans. 1979. Zum Erwerbsverhalten verheirateter Frauen. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 12: 217–240.

Hoffman, Saul D., und E. Michael Foster. 2000. AFDC Benefits and Nonmarital Births to Young Women. *The Journal of Human Resources* 35: 376–391.

Hsiao, Cheng. 1986. *Analysis of panel data*. Cambridge, New York, NY: Cambridge University Press.

Hsiao, Cheng. 2003. *Analysis of panel data*. 2. Aufl.. Cambridge, New York, NY: Cambridge University Press.

Huinink, Johannes. 1989. Ausbildung, Erwerbsbeteiligung von Frauen und Familienbildung im Kohortenvergleich. In *Familienbildung und Erwerbstätigkeit im demographischen Wandel*, Hrsg. Gert Wagner, Notburga Ott, und Hans J. Hoffmann-Nowotny. Berlin, Heidelberg: Springer, 136–158.

Huinink, Johannes, Josef Brüderl, Bernhard Nauck, Sabine Walper, Laura Castiglioni, und Michael Feldhaus. 2011. Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung* 23: 77–101.

Hyslop, Dean R. 1999. State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica* 67: 1255–1294.

Infratest Burke Sozialforschung. 2000. Familie und Partnerbeziehungen in der Bundesrepublik Deutschland (Familiensurvey 2000). Methodenbericht.

Inglehart, Ronald. 1977. *The Silent Revolution: Changing Values and Political Styles among Western Publics*. Princeton, NJ: Princeton University Press.

Jacobsen, Joyce P., James Wishart III Pearce, und Joshua L. Rosenbloom. 1999. The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment. *The Journal of Human Resources* 34: 449–474.

Jagodzinski, Wolfgang, und Steffen-Matthias Kühnel. 1990. Zur Schätzung der relativen Effekte von Issue-Orientierungen, Kandidatenpräferenz und langfristiger Parteibindung auf die Wahlabsicht. In *Wahlen, Parteieliten, politische Einstellungen: Neuere*

Forschungsergebnisse. Frankfurt: Peter Lang Verlag, 5–63.

Joesch, Jutta M. 1997. Paid Leave and the Timing of Women's Employment Before and After Birth. *Journal of Marriage and Family* 59: 1008–1021.

Johnson, Norman L., Samuel Kotz, und N. Balakrishnan. 1994. *Continuous univariate distributions*. 2. Aufl.. New York, NY et al.: Wiley.

Joshi, Heather, und P. R. Andrew Hinde. 1993. Employment after Childbearing in Post-War Britain: Cohort-Study Evidence on Contrasts within and across Generations. *European Sociological Review* 9: 203–227.

Kaase, Max, und Hans-Dieter Klingemann. 1994. Electoral research in the Federal Republic of Germany. *European Journal of Political Research* 25: 343–366.

Kalbfleisch, John D., und Ross L. Prentice. 2002. *The statistical analysis of failure time data*. 2. Aufl.. Hoboken, NJ: Wiley.

Kirner, Ellen, und Erika Schulz. 1991. Die Erwerbsbeteiligung im Lebensverlauf von Frauen in Abhängigkeit von der Kinderzahl: Unterschiede zwischen der Bundesrepublik Deutschland und der ehemaligen Deutschen Demokratischen Republik. In *Frauenalterssicherung: Lebensläufe von Frauen und ihre Benachteiligung im Alter*, Hrsg. Claudia Gather, Ute Gerhard, Karin Prinz, und Mechthild Veil. Berlin: Sigma, 62–82.

Kish, Leslie. 1965. *Survey sampling*. New York, NY et al.: John Wiley & Sons.

Kohler, Ulrich. 2002. *Der demokratische Klassenkampf: Zum Zusammenhang von Sozialstruktur und Parteipräferenz*. Frankfurt am Main, New York, NY: Campus Verlag.

Kohler, Ulrich. 2005. Changing Class Locations and Partisanship in Germany. In *The Social Logic of Politics: Personal Networks as Contexts for Political Behavior*, Hrsg. Alan S. Zuckerman, Kap. 6. Philadelphia, PA: Temple University Press, 117–131.

Kravdal, Oystein. 1992. Forgone Labor Participation and Earning Due to Childbearing Among Norwegian Women. *Demography* 29: 545–563.

Kuh, Edwin. 1959. The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications. *Econometrica* 27: 197–214.

Kühnel, Steffen-Matthias. 1993. *Zwischen Boykott und Kooperation: Teilnahmeabsicht und Teilnahmeverhalten bei der Volkszählung 1987*. Frankfurt am Main et al.: Peter Lang Verlag.

Lancaster, Tony. 1990. *The econometric analysis of transition data*. Cambridge, MA et al.: Cambridge University Press.

Lange, Yvonne. 2007. *Fertilität und Erwerbsbeteiligung von Frauen in Deutschland: Eine empirische Analyse*. Frankfurt et al.: Peter Lang Verlag.

Lauterbach, Wolfgang. 1994. *Berufsverläufe von Frauen: Erwerbstätigkeit, Unterbrechung und Wiedereintritt*. Frankfurt, New York, NY: Campus Verlag.

Lazarsfeld, Paul F., Bernard Berelson, und Hazel Gaudet. 1944. *The people's choice: How the voter makes up his mind in a presidential campaign.* New York, NY: Duell, Sloan and Pearce.

Lee, Myoung-jae. 2002. *Panel data econometrics: Methods-of-moments and limited dependent variables.* San Diego, CA et al.: Academic Press.

Levy, Paul S., und Stanley Lemeshow. 2008. *Sampling of Populations: Methods and Applications.* 4. Aufl.. Hoboken, NJ: Wiley.

Lind, Jo T. 2007. Does Permanent Income Determine the Vote? *The B.E. Journal of Macroeconomics* 7.

Lipset, Seymour M., und Stein Rokkan, Hrsg.. 1967. *Party systems and voter alignments: Cross-national perspectives.* New York, NY: Free Press.

Long, J. Scott. 1997. *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA et al.: Sage Publications.

Lynn, Peter J. 2005. Weighting. In *Encyclopedia of Social Measurement 1*, Hrsg. Kimberly Kempf-Leonard. San Diego, CA: Elsevier, 967–973.

Maddala, G. S. 1983. *Limited-dependent and qualitative variables in econometrics.* Cambridge, New York, NY: Cambridge University Press.

Maddala, G. S. 2001. *Introduction to econometrics.* 3. Aufl.. Chichester, New York, NY: Wiley Press.

Maier, Gunther, und Peter Weiss. 1990. *Modelle diskreter Entscheidungen: Theorie und Anwendung in den Sozial- und Wirtschaftswissenschaften.* Wien, New York, NY: Springer-Verlag.

Malchow, Matthew B., und Adib Kanafani. 2004. A disaggregate analysis of port selection. *Transportation Research, Part E* 40: 317–337.

Matysiak, Anna. 2009. Employment first, then childbearing: Women's strategy in post-socialist Poland. *Population Studies: A Journal of Demography* 63: 253–276.

Matysiak, Anna, und Stephanie Steinmetz. 2008. Finding their way female employment patterns in West Germany, EastGermany, and Poland. *European Sociological Review* 24: 331–345.

Matysiak, Anna, und Daniele Vignoli. 2010. Employment around first birth in two adverse institutional settings: Evidence from Italy and Poland. *Zeitschrift für Familienforschung* 22: 331–349.

McFadden, Daniel. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in econometrics*, Hrsg. Paul Zarembka. New York, NY et al.: Academic Press, 105–142.

McFadden, Daniel. 2001. Economic choices. *The American Economic Review* 91: 351–378.

Michaud, Pierre-Carl, und Konstantinos Tatsiramos. 2011. Fertility and female employment dynamics in Europe: the effect of using alternative econometric modeling assumptions. *Journal of Applied Econometrics* 26: 641–668.

Mincer, Jacob. 1962. On-the-Job Training: Costs, Returns, and Some Implications. *Journal of Political Economy* 70: 50–79.

Mood, Carina. 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26: 67–82.

Morgan, Stephen L., und Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge et al.: Cambridge University Press.

Müller, Walter. 1998. Klassenstruktur und Parteiensystem. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 50: 3–47.

Mundlak, Yair. 1961. Empirical Production Function Free of Management Bias. *Journal of Farm Economics* 43: 44–56.

Mundlak, Yair. 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica* 46: 69–85.

Neuhaus, John M., und Jack D. Kalbfleisch. 1998. Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data. *Biometrics* 54: 638–645.

Neundorf, Anja, Daniel Stegmüller, und Thomas J. Scotto. 2011. The individual-level dynamics of bounded partisanship. *Public Opinion Quarterly* 3: 458–482.

Nobelprize.org. 2000. The Prize in Economics 2000 – Press Release. 05.20.2011. `http://nobelprize.org/nobel_prizes/economics/laureates/2000/press.html`.

Ondrich, Jan, C. Katharina Spiess, Qing Yang, und Gert G. Wagner. 2000. Full time or part time? German parental leave policy and the return to work after childbirth in Germany. *Research in Labor Economics* 18: 41–74.

Pappi, Franz Urban, und Jens Brandenburg. 2010. Sozialstrukturelle Interessenlagen und: Parteipräferenz in Deutschland: Stabilität und Wandel seit 1980. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 459–483.

Pitt, Mark M., und Mark R. Rosenzweig. 1990. Estimating the Intrahousehold Incidence of Illness: Child Health and Gender-Inequality in the Allocation of Time. *International Economic Review* 31: 969–989.

Pradhan, Menno. 1998. Enrolment and delayed enrolment of secondary school age children in Indonesia. *Oxford bulletin of economics and statistics* 60: 413–430.

Rønsen, Marit, und Marianne Sundström. 1996. Maternal Employment in Scandinavia: A Comparison of the After-Birth Employment Activity of Norwegian and Swedish Women. *Journal of Population Economics* 9: 267–285.

Rønsen, Marit, und Marianne Sundström. 2002. Family Policy and After-Birth Employment Among New Mothers: A Comparison of Finland, Norway and Sweden. *European Journal of Population* 18: 121–152.

Rosenzweig, Mark R. 1993. Women, Insurance Capital, and Economic Development in Rural India. *The Journal of Human Resources* 28: 735–758.

Rosenzweig, Mark R., und Kenneth I. Wolpin. 1994. Parental and Public Transfers to Young Women and Their Children. *The American Economic Review* 84: 1195–1212.

Royalty, Anne B., und Neil Solomon. 1999. Health Plan Choice: Price Elasticities in a Managed Competition Setting. *The Journal of Human Resources* 34: 1–41.

Saurel-Cubizolles, Marie-Josèphe, Patrizia Romito, Vicenta Escribà-Aguir, Nathalie Lelong, Rosa Mas Pons, und Pierre-Yves Ancel. 1999. Returning to Work after Childbirth in France, Italy, and Spain. *European Sociological Review* 15: 179–194.

Scanlon, Dennis P., Michael Chernew, Catherine McLaughlin, und Gary Solon. 2002. The impact of health plan report cards on managed care enrollment. *Journal of Health Economics* 21: 19–41.

Schnabel, Reinhard. 1994. *Das intertemporale Arbeitsangebot verheirateter Frauen: Eine empirische Analyse auf der Basis des sozio-ökonomischen Panels.* Frankfurt et al.: Campus Verlag.

Schröder, Jette. 2010. *Der Zusammenhang zwischen der Erwerbstätigkeit von Frauen und ihrer Fertilität.* Würzburg: Ergon Verlag.

Schröder, Jette, und Klaus Pforr. 2009. The relationship between women's employment and fertility: A review of the current state of research. *Zeitschrift für Familienforschung* 21: 218–244.

Schultz, T. Paul. 1987. The Influence of Fertility on labor supply of married women: Simultaneous equation estimates. In *Reserach in Labor Economics*, Hrsg. Ronald G. Ehrenberg, Jg. 2. Greenwich: JAI Press, 273–351.

Schulze, Gerhard. 1992. *Die Erlebnissgesellschaft: Kultursoziologie der Gegenwart.* Frankfurt, New York, NY: Campus Verlag.

Shafer, Emily Fitzgibbons. 2011. Wives' Relative Wages, Husbands' Paid Work Hours, and Wives' Labor-Force Exit. *Journal of Marriage and Family* 73: 250–263.

Shaw, Kathryn. 1994. The Persistence of Female Labor Supply: Empirical Evidence and Implications. *The Journal of Human Resources* 29: 348–378.

Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2010, Version 27, SOEP. 2011. doi:10.5684/soep.v27.

Späth, Helmuth. 1994. *Numerik: Eine Einführung für Mathematiker und Informatiker.* Braunschweig, Wiesbaden: Vieweg.

StataCorp LP. 2009a. *Mata reference manual: release 11.* College Station, TX.

StataCorp LP. 2009b. *Stata longitudinal-data/panel-data reference manual: release 11*. College Station, TX.

StataCorp LP. 2009c. *Stata programming: reference manual: release 11*. College Station, TX.

StataCorp LP. 2009d. *Stata reference manual: release 11*. College Station, TX.

Steele, Fiona. 2011. Multilevel discrete-time event history analysis with applications to the analysis of recurrent employment transitions. *Australia and New Zealand Journal of Statistics* 53: 1–26.

Steinbrecher, Markus. 2009. *Politische Partizipation in Deutschland*, Jg. 11 von *Studien zur Wahl- und Einstellungsforschung*. Baden-Baden: Nomos.

Stoer, Josef, und Roland Bulirsch. 1983. *Einführung in die numerische Mathematik: Teil 1*. 4. Aufl.. Berlin et al.: Springer.

Thevenon, Olivier. 2009. Increased Women's Labour Force Participation in Europe: Progress in the Work-Life Balance or Polarization of Behaviours? *Population* 64: 235–272.

Thurner, Paul W. 1998. *Wählen als rationale Entscheidung: Die Modellierung von Politikreaktion im Mehrparteiensystem*. München: Oldenbourg Verlag.

Tölke, Angelika. 1989. *Lebensverläufe von Frauen. Familiäre Ereignisse, Ausbildungs- und Erwerbsverhalten*. München: DJI Verlag.

Train, Kenneth E. 1986. *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*. Cambridge, MA: MIT Press.

Train, Kenneth E. 2009. *Discrete choice methods with simulation*. 2. Aufl.. Cambridg et al.: Cambridge University Press.

Troske, Kenneth R., und Alexandru Voicu. 2010. Joint estimation of sequential labor force participation and fertility decisions using Markov chain Monte Carlo techniques. *Labour Economics* 17: 150–169.

Untiedt, Gerhard. 1990. *Das Erwerbsverhalten verheirateter Frauen in der Bundesrepublik Deutschland: Eine mikroökonometrische Untersuchung*. Heidelberg: Physica-Verlag.

Van der Lippe, Tanja. 2001. The effect of individual and institutional constraints on hours of paid work of women. In *Women's employment in a comparative perspective*, Hrsg. Tanja Van der Lippe, und Liset van Dijk. New York, NY: De Gruyter, 221–243.

Vlasblom, Jan Dirk, und Joop J. Schippers. 2004. Increases in female labour force participation in Europe: Similarities and differences. *European Journal of Population* 20: 375–392.

Wagner, Gert G., Joachim R. Frick, und Jürgen Schupp. 2007. The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements. *Schmollers Jahrbuch* 127: 139–169.

Waldfogel, Jane, Yoshio Higuchi, und Masahiro Abe. 1999. Family leave policies and women’s retention after childbirth: Evidence from the United States, Britain, and Japan. *Journal of Population Economics* 12: 523–545.

Weber, Andrea Maria. 2004. *Wann kehren junge Mütter auf den Arbeitsmarkt zurück? Eine Verweildaueranalyse für Deutschland.* Discussion Paper 04-08, ZEW.

Wenk, Deeann, und Patricia Garrett. 1992. Having a Baby: Some Predictions of Maternal Employment around Childbirth. *Gender and Society* 6: 49–65.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* 2. Aufl.. Cambridge, MA, London: MIT Press.

Fixed-effects-Modelle sind zu einem wichtigen Werkzeug der Kausalanalyse geworden, da sie die Kontrolle von unbeobachteter Heterogenität ermöglichen. Bis heute wurden fixed-effects-Modelle für kontinuierliche und dichotome abhängige Variablen entwickelt und für viele Statistikpakete implementiert. Für multinominale abhängige Variable wurde von Chamberlain (1980) ein solches Modell entwickelt, es liegt aber keine allgemein anwendbare Umsetzung vor. Die vorliegende Arbeit schließt diese Forschungslücke mit der ersten Umsetzung des Modells von Chamberlain in einem weitverbreiteten Statistikpaket (Stata). Die Anwendbarkeit wird durch Erweiterungen der Arbeiten von Schröder (2010) im Bereich der Familiensoziologie und Kohler (2005) im Bereich der politischen Soziologie gezeigt.

Fixed effects models have become a prime tool for causal analysis, as they allow to control for unobserved heterogeneity. As of today, fixed effects models have been derived and implemented for many statistical software packages for continuous, dichotomous and count-data dependent variables. For multinomial categorical dependent variables such a model has been derived in a seminal paper by Chamberlain (1980), but no implementation is available. The dissertation on hand closes this research gap by delivering the first implementation of Chamberlain's model in a widely available statistical package (Stata). Its applicability is shown by extending Schröder's (2010) work in the sociology of the family and Kohler's (2005) work in political sociology.

gesis Leibniz-Institut für Sozialwissenschaften

ISBN 978-3-86819-020-5
ISSN 1869-2869

19,- EUR