

Inaugural-Dissertation zur Erlangung des Grades eines Doktor  
oeconomiae publicae (Dr. oec. publ.) an der  
Ludwig-Maximilians-Universität München

**Aufsätze zur ökonometrischen Analyse  
diskreter Strukturen  
in der Arbeitsmarkt- und  
Entwicklungsökonomie**

vorgelegt von  
Hermann Gartner

2002

Referent: Prof. Stephan Klasen, Ph.D.

Korreferent: Prof. Dr. Gebhard Flaig

Promotionsabschlussberatung: 5. Februar 2003

# Vorwort

Für den erfolgreichen Abschluss der vorliegenden Arbeit möchte ich mich bei vielen Menschen herzlich bedanken. An erster Stelle bedanke ich mich bei Prof. Stephan Klasen für die wertvolle fachliche Betreuung. Er hat mir viel Geduld und Vertrauen entgegengebracht und gab mir in vielen Diskussionen wichtige Anregungen. Weiterhin bedanke ich mich bei Prof. Gebhard Flaig, der sich bereit erklärte, das Zweitgutachten zu erstellen.

Für wertvolle Anregungen und Kommentare bedanke ich mich auch bei Prof. Regina Riphahn und Prof. John Ermisch. Besonders danken möchte ich auch meinen Kollegen am Seminar für empirische Wirtschaftsforschung der LMU Ph.D. Joachim Wolff und Dr. René Böheim, die verschiedene Teile des Manuskripts lasen und ebenso hilfreiche wie kritische Kommentare abgaben. Ich bedanke mich weiterhin bei allen, die an den Seminaren des volkswirtschaftlichen Institutes der LMU teilnahmen, worin ich meine Arbeitsergebnisse präsentierte, und die diese mit mir kritisch diskutierten.

Für die Hilfe bei statistischen Problemen bedanke ich mich bei den Kollegen und Kolleginnen vom Institut für Statistik an der LMU: bei Dr. Andreas Fieger und Dipl. Math. Leyre Estibaliz Osuna Echavarria. Teile der Arbeit sind aus Kooperationen mit Kollegen des Instituts für Statistik hervorgegangen. Besonderer Dank gebührt daher Dipl. Stat. Sandro Scheid, der bei Problemen der multiplen Imputation hilfreiche Unterstützung leistete, sowie Dipl. Math. oec. Torsten Scholtz, der mir semiparametrische Verfahren nahe brachte.

Großen Dank verdienen schließlich meine Eltern für die materielle und die moralische Unterstützung während meiner Qualifikationsphase.

München, September 2002  
Hermann Gartner



# Inhaltsverzeichnis

<b>Tabellenverzeichnis</b>	<b>6</b>
<b>Abbildungsverzeichnis</b>	<b>8</b>
<b>1 Einleitung</b>	<b>9</b>
<b>1 Der Auszug junger Menschen aus dem Elternhaus in Westdeutschland von 1984 bis 1999</b>	<b>12</b>
<b>2 Einführung</b>	<b>14</b>
<b>3 Forschungsstand</b>	<b>16</b>
<b>4 Theoretischer Rahmen</b>	<b>21</b>
<b>5 Das ökonometrische Modell</b>	<b>25</b>
5.1 Lineare Modellierung . . . . .	25
5.2 Semiparametrische Modellierung . . . . .	27
<b>6 Datenquelle und deskriptive Ergebnisse</b>	<b>31</b>
<b>7 Schätzergebnisse</b>	<b>39</b>
7.1 Lineares Modell . . . . .	40
7.2 Semiparametrisches Modell . . . . .	52
<b>8 Schlussfolgerung</b>	<b>58</b>
<b>Literatur</b>	<b>59</b>

## *Inhaltsverzeichnis*

<b>II Wann verlassen junge Menschen in Italien das Elternhaus?</b>	<b>64</b>
<b>9 Einführung</b>	<b>66</b>
<b>10 Bisherige Untersuchungen</b>	<b>69</b>
<b>11 Modell</b>	<b>72</b>
<b>12 Daten</b>	<b>75</b>
<b>13 Ergebnisse</b>	<b>79</b>
<b>14 Schlussfolgerung</b>	<b>100</b>
<b>Literatur</b>	<b>101</b>
<b>III Multiple Imputation von fehlenden Werten mit Daten über Unterernährung und Kindersterblichkeit</b>	<b>104</b>
<b>15 Einführung</b>	<b>106</b>
<b>16 Unterernährung und Kindersterblichkeit</b>	<b>107</b>
16.1 Fragestellung . . . . .	107
16.2 Datensatz . . . . .	107
<b>17 Die Complete-case-Analyse in Klasen (2000)</b>	<b>111</b>
<b>18 Die Ersetzung fehlender Werte</b>	<b>117</b>
18.1 Klassische Verfahren . . . . .	117
18.2 Multiple Imputation mit Markov-Chain-Monte-Carlo Verfahren . . . . .	118
18.2.1 Annahmen . . . . .	118
18.2.2 Multiple Imputation . . . . .	119
18.2.3 Markov-Chain-Monte-Carlo (MCMC) Verfahren . . . . .	120
18.2.4 MCMC-Verfahren in einem Modell mit gemischten Variablen . . .	122
<b>19 Ergebnisse</b>	<b>124</b>
<b>20 Schlussfolgerung</b>	<b>130</b>
<b>Literatur</b>	<b>131</b>

# Tabellenverzeichnis

6.1	Zustand am Ende der Beobachtungsdauer . . . . .	32
6.2	Beobachtete Dauer im Elternhaus . . . . .	33
6.3	Mehrfachnennungen bei Erwerbsstatus . . . . .	35
6.4	Mittelwerte und Standardabweichungen der Variablen . . . . .	38
7.1	Hazardratenmodell 1, Männer . . . . .	43
7.2	Hazardratenmodell 1, Frauen . . . . .	44
7.3	Hazardratenmodell 2, Männer . . . . .	45
7.4	Hazardratenmodell 2, Frauen . . . . .	46
7.5	Hazardratenmodell 3, Männer . . . . .	47
7.6	Hazardratenmodell 3, Frauen . . . . .	48
7.7	Prognose der Auszugswahrscheinlichkeit . . . . .	49
7.8	Veränderung der Auszugswahrscheinlichkeit . . . . .	50
7.9	Modell mit unbeobachteter Heterogenität: Frauen . . . . .	50
7.10	Modell mit unbeobachteter Heterogenität: Männer . . . . .	51
7.11	Semiparametrisches Modell - Koeffizienten: Männer . . . . .	56
7.12	Semiparametrisches Modell - Koeffizienten: Frauen . . . . .	57
9.1	Durchschnittliches Auszugsalter 1996 . . . . .	67
9.2	Jugendarbeitslosigkeit . . . . .	67
9.3	Arbeitsmarktprogramme für Jugendliche . . . . .	68
12.1	Variablenbeschreibung: Daten 1991/1993 . . . . .	77
12.2	Variablenbeschreibung: Daten 1995 . . . . .	78
13.1	Auszugswahrscheinlichkeit 1991/93: Logit-Modell 1 . . . . .	82
13.2	Marginale Effekte . . . . .	83
13.3	Auszugswahrscheinlichkeit 1991/93: Logit-Modell 1: Männer . . . . .	84

## Tabellenverzeichnis

13.4 Marginale Effekte: Männer . . . . .	85
13.5 Auszugswahrscheinlichkeit 1991/93: Logit-Modell 1: Frauen . . . . .	86
13.6 Marginale Effekte: Frauen . . . . .	87
13.7 Auszugswahrscheinlichkeit 1991/93: Logit-Modell 2, Männer . . . . .	88
13.8 Marginale Effekte: Männer . . . . .	89
13.9 Auszugswahrscheinlichkeit 1991/93: Logit-Modell 2: Frauen . . . . .	90
13.10 Marginale Effekte: Frauen . . . . .	91
13.11 Auszugswahrscheinlichkeit 1995: Logit-Modell 1, Männer . . . . .	92
13.12 Marginale Effekte: Männer . . . . .	93
13.13 Auszugswahrscheinlichkeit 1995: Logit-Modell 1, Frauen . . . . .	94
13.14 Marginale Effekte: Frauen . . . . .	95
13.15 Auszugswahrscheinlichkeit 1995: Logit-Modell 2, Männer . . . . .	96
13.16 Marginale Effekte: Männer . . . . .	97
13.17 Auszugswahrscheinlichkeit 1995: Logit-Modell 2, Frauen . . . . .	98
13.18 Marginale Effekte: Frauen . . . . .	99
17.1 Beschreibung des Datensatzes . . . . .	112
17.2 Unterernährungsmodelle mit unvollständigen Datensatz . . . . .	113
17.3 Sterblichkeitsmodelle mit unvollständigen Datensatz . . . . .	115
19.1 Regression: Geburtsgewicht . . . . .	126
19.2 Regression: Wachstumsverzögerung . . . . .	127
19.3 Regression: Untergewicht . . . . .	128
19.4 Regression: Kindersterblichkeit . . . . .	129
20.1 Liste der im Datensatz vorhandenen Länder . . . . .	133

# Abbildungsverzeichnis

5.1	Illustration von B-Splines 2. Grades . . . . .	29
6.1	Überlebensrate . . . . .	34
6.2	Hazardrate . . . . .	35
6.3	Überlebensraten nach Geschlecht . . . . .	36
7.1	Geschätzter Effekt - Single, Männer . . . . .	53
7.2	Geschätzter Effekt - mit Partnerin, Männer . . . . .	53
7.3	Geschätzte Effekte - Single, Frauen . . . . .	54
7.4	Geschätzte Effekte - mit Partner, Frauen . . . . .	55

# 1 Einleitung

Die hier veröffentlichten Arbeiten sind im Rahmen des Teilprojektes B6 im Sonderforschungsbereich 386 „Statistische Analyse diskreter Strukturen - Anwendungen in der Biometrie und Ökonometrie“ entstanden. Das Teilprojekt befasst sich mit der ökonometrischen Analyse diskreter Strukturen in arbeits- und bevölkerungsökonomischen Modellen.

Viele Erscheinungen in der Ökonomie haben diskrete Strukturen, wie zum Beispiel ‚beschäftigt‘, ‚arbeitslos‘ oder ‚kaufen‘, ‚nicht kaufen‘. Um diese empirisch zu analysieren sind geeignete statistische Methoden erforderlich. Im Rahmen des Sonderforschungsbereich 386 werden am statistischen Institut der Ludwig-Maximilians-Universität derartige Verfahren entwickelt. Besonders berücksichtigt werden dort Probleme, mit denen die angewandte Wirtschaftsforschung häufig konfrontiert ist, wie fehlende Werte, Messfehler oder unbeobachtete Heterogenität, aber auch Fragen der flexiblen Analyse nicht-linearer Zusammenhänge. Ein Ziel der Arbeit im Teilprojekt B6, das von Prof. Stephan Klasen geleitet wird, ist diese neu entwickelten statistischen Verfahren auf die Anwendbarkeit auf konkrete Fragen der Ökonomie zu prüfen.

Inhaltlich zentrale Fragestellung in diesem Forschungsprojekt ist die Frage, was die Bestimmungsfaktoren für eine hohe Lebenserwartung, ausreichende Ernährung- und Beschäftigungsmöglichkeiten sind. Sen (1999) bezeichnet diese Größen als die wichtigsten Determinanten der Wohlfahrt in Entwicklungs- und Transformationsländern.

Aufgrund der wohlfahrtstheoretischen Bedeutung der Beschäftigung ist auch die Frage von Interesse, wie in einer Gesellschaft mit Individuen umgegangen wird, die sich mit ungünstigen Beschäftigungsmöglichkeiten konfrontiert sehen. Eine Fragestellung in diesem Forschungsprojekt ist, welche Rolle informelle, familiäre Strukturen haben, um deren Mitglieder vor Risiken wie Arbeitslosigkeit abzusichern. Vor allem in Entwicklungs- und Transformationsländern sind familiäre Versicherungssysteme von großer Wichtigkeit. In zwei Kapiteln dieser Arbeit geht es um die Frage, ob auch in den Industrieländern Deutschland (West) und Italien junge Menschen auf Arbeitslosigkeit reagieren, indem sie länger bei den Eltern bleiben.

In Teil I wird untersucht, ob Arbeitslosigkeit einen Einfluss auf das Auszugsverhalten

## 1 Einleitung

junger Menschen in der Bundesrepublik (West) hat. Die Untersuchung wird auf Basis von Daten des sozioökonomischen Panels für die Geburtsjahrgänge von 1968 bis 1981 durchgeführt. Es wird ein Hazardratenmodell geschätzt, in welchem die Zielzustände ‚Single-Haushalt‘ und ‚Wohnung mit Partner‘ unterschieden werden. Die Auswirkung des Einkommens auf die Auszugswahrscheinlichkeit wird semi-parametrisch modelliert. Es zeigt sich zwar, dass junge Männer, die arbeitslos werden, mit einer geringeren Wahrscheinlichkeit einen Haushalt mit einer Partnerin gründen. Darüber hinaus hat Arbeitslosigkeit jedoch keinen Einfluss auf die Auszugswahrscheinlichkeit. Die Institution Familie fungiert also für junge Menschen nur in geringem Maße als Versicherung gegen Arbeitslosigkeit.

Ziel des II. Teil dieser Untersuchung ist, den Einfluss ökonomischer Faktoren auf die Wahrscheinlichkeit zu bestimmen, mit der junge Menschen in Italien das Elternhaus verlassen. Hierzu werden mit den Wellen von 1991 bis 1998 des Householdsurveys der Banca d' Italia mehrere Logit-Modelle geschätzt. Ein Ergebnis ist, dass mit der Beendigung der Schulausbildung die Auszugswahrscheinlichkeit zunimmt. Arbeitslose Männer haben gegenüber Beschäftigten eine geringere Auszugswahrscheinlichkeit. Ökonomische Faktoren haben in Italien somit einen hohen Einfluss auf das Auszugsverhalten junger Menschen. Die Bedeutung ökonomischer Größen deutet auf die Versicherungsfunktion hin, die familiäre Netzwerke in Italien einnehmen.

Im Teil III der vorliegenden Arbeit werden die Bestimmungsfaktoren von Unterernährung und Kindersterblichkeit analysiert. Untersuchungen zu dieser Frage stehen oft vor dem Problem, dass die Datengrundlage in Entwicklungsländern schlecht ist. Im hier verwendeten Makro-Datensatz fehlen zahlreiche Beobachtungen. Bei einer in solchen Situationen üblichen Complete-case Analyse gehen viele der vorhandenen Informationen im Datensatz verloren. Mit einem Verfahren der multiplen Imputation der fehlenden Werte wird versucht, dieses Problem zu lösen. Hierzu werden mit einem Markov-Chain-Monte-Carlo-Verfahren mehrere vervollständigte Datensätze generiert, welche dann weiter bearbeitet werden. Durch die Auswertung mehrerer vervollständigter Datensätze wird eine höhere Effizienz der Schätzer erreicht.

Ein Vergleich von Regressionsanalysen, die mit dem vervollständigten Daten durchgeführt wurden, mit einer Complete-case-Analyse hat gezeigt, dass sich bestimmte Koeffizienten in ihrer Größenordnung geändert haben. Bei manchen Koeffizienten sind unplausible Vorzeichen aus der Complete-case Analyse verschwunden. Es ist also vorteilhaft, bei Problemen mit fehlenden Werten moderne Imputationsverfahren zu verwenden. Durch die Ersetzung der fehlenden Werte konnten eine Reihe von Variablen zugänglich

## *1 Einleitung*

gemacht werden, die bei einer Complete-case Analyse nicht brauchbar sind, da bei deren Verwendung auf noch mehr Beobachtungen hätte verzichtet werden müssen.

## **Teil I**

# **Der Auszug junger Menschen aus dem Elternhaus in Westdeutschland von 1984 bis 1999**

## **Zusammenfassung**

In dieser Arbeit wird untersucht, ob Arbeitslosigkeit einen Einfluss auf das Auszugsverhalten junger Menschen in der Bundesrepublik (West) hat. Die Untersuchung wird auf Basis von Daten des sozioökonomischen Panels für die Geburtsjahrgänge von 1968 bis 1981 durchgeführt. Es wird ein Hazardratenmodell geschätzt, in welchem die Zielzustände ‚Single-Haushalt‘ und ‚Wohnung mit Partner‘ unterschieden werden. Es zeigt sich, dass Männer, die arbeitslos werden, mit einer geringeren Wahrscheinlichkeit einen Haushalt mit einer Partnerin gründen. Darüber hinaus hat Arbeitslosigkeit keinen Einfluss auf die Auszugswahrscheinlichkeit. Die Institution Familie fungiert also für junge Menschen nur in geringem Maße als Versicherung gegen Arbeitslosigkeit. Ein höheres eigenes Einkommen führt zu einer höheren Auszugswahrscheinlichkeit. Ein höheres Elterneinkommen führt entgegen der theoretischen Erwartung eher zu einer höheren Auszugswahrscheinlichkeit.

JEL-Klassifikation: D19, C41

## 2 Einführung

Die Art und Weise wie Personen sich in Haushalten bzw. Familien organisieren, ist von hoher Relevanz für sozialpolitische, demographische, wohnungsmarktökonomische und arbeitsmarktökonomische Fragestellungen. Ein Grund für die Relevanz liegt in der Versicherungsfunktion gegenüber einer Reihe von Risiken, welche die Familie erfüllt.

In je stärkerem Maße die Institution Familie Versicherungsfunktionen übernimmt, desto größer ist die Rolle von Arbeitsmarktereignissen bei der Entscheidung darüber, mit wem man sich in einem Haushalt organisiert. Desto stärker sind damit Abhängigkeiten im Rahmen unmittelbarer sozialer Beziehungen. Heranwachsende Menschen stehen dann in einer ökonomischen Abhängigkeitsbeziehung zu ihren Eltern. Es ist daher von gesellschaftspolitischer Relevanz, welche Einflussfaktoren das Haushaltsformationsverhalten bestimmen.

Die Frage, inwiefern Familien Versicherungsfunktionen übernehmen, ist auch unter Effizienzgesichtspunkten von Interesse. Die Institution Familie kann als Versicherungseinrichtung weniger effizient sein als marktwirtschaftliche oder öffentliche Einrichtungen, da innerhalb von Familien nur wenig diversifiziert werden kann. Ein Vorteil der Institution Familie ist hingegen, dass hier geringere Informationsprobleme bestehen.

Von arbeitsmarktpolitischer Bedeutung ist die Institution Familie, weil durch diese Arbeitsmarkttrigiditäten entstehen können. In einer Untersuchung Südafrikas (Klasen und Woolard, 1999), einem Land, in welchem nur in geringem Maße öffentliche Transfers an Arbeitslose bezahlt werden, wird argumentiert, dass Arbeitslose häufig ihren Haushalt auflösen und zu Familienmitgliedern ziehen, welche noch über weitere Einkommensquellen verfügen. In Ländern ohne bzw. mit geringen Sozialtransfers, die bei Arbeitslosigkeit geleistet werden, wird also auf Ressourcen der Familie zurückgegriffen. Transfers innerhalb der Familie können somit ebenso wie öffentliche Transferzahlungen einen hohen Anspruchslohn verursachen oder zu geringerer Mobilität führen.

Die Familie ist auch eine Institution, innerhalb deren die Umverteilung von Ressourcen zwischen den Generationen stattfindet. Erwachsene, die Kinder großziehen investieren einen Teil ihres Einkommens in das künftige Einkommen ihrer Kinder. Unter diesem Aspekt der intergenerativen Verteilung ist es von Interesse, wann junge Menschen das

## 2 Einführung

Elternhaus verlassen. Fogli (2000) präsentiert ein Modell, das diesen Zusammenhang betont.

Auch unter wohnungsmarktpolitischen Gesichtspunkten spielt das Auszugsverhalten eine Rolle. Die Entscheidung zum Auszug ist zugleich eine Entscheidung, Wohnraum nachzufragen (vgl. Haurin et al., 1994). Wenn junge Menschen früher das Elternhaus verlassen, steigt damit der Bedarf an Wohnraum.

In dieser Arbeit soll das Auszugsverhalten Heranwachsender in der Bundesrepublik Deutschland (West) in der Zeit von 1984 bis 1999 auf Basis von Daten des Sozio-ökonomischen Panels (GSOEP) untersucht werden. Wie in den Untersuchungen zum Auszugsverhalten üblich, werden im Folgenden Heranwachsende, die bei den Eltern leben als „Kinder“ bezeichnet, auch wenn die hier untersuchten Personen dem Kindesalter bereits entwachsen sind.

Die Arbeit ist wie folgt aufgebaut. Zunächst werden die wichtigsten Ergebnisse bisheriger Studien vorgestellt. In den darauf folgenden Abschnitten wird auf das theoretische Modell und die Schätzmethode eingegangen. Sodann werden die verwendeten Daten diskutiert und einige deskriptive Ergebnisse dargestellt. Im nächsten Abschnitt werden dann Hazardratenmodelle geschätzt. Die Arbeit schließt mit der Zusammenfassung.

### 3 Forschungsstand

Innerhalb Europas gibt es erhebliche Unterschiede im Haushaltsformationsverhalten junger Menschen. Einen räumlichen und historischen Vergleich Europäischer Länder der Gegenwart mit dem vorindustriellen Europa bietet hierzu Wall (1989). Aktuelle Trends im Auszugsverhalten werden in Iacovou (1998) und Iacovou und Berthoud (2001) dargestellt. Sie sprechen von zwei Modellen des Verhaltens in Europa: Einem süd-europäischen Modell und einem nord-europäischen Modell. Im Süden bleiben junge Menschen länger im Elternhaus und verlassen es meist mit der Gründung einer eigenen Familie. Im Norden verlassen sie das Elternhaus früher und leben zunächst allein oder in einer Partnerschaft ohne Kinder. Die Gründe für diese Unterschiede werden aber nicht systematisch untersucht. Iacovou vermutet aber, dass diese Unterschiede zum Teil mit Unterschieden in den sozialstaatlichen Institutionen erklärbar sind.

Auch im außereuropäischen Vergleich zeigen sich Unterschiede im Auszugsverhalten. Yi, Coale, Choe, Zhiwu und Li (1994) vergleichen ostasiatische Länder mit westlichen Ländern. In ostasiatischen Ländern (China, Japan, Süd-Korea) ist das Auszugsalter 2 bis 3 Jahre höher als im westlichen Durchschnitt (USA, Frankreich, Schweden). Auch hier werden die Ursachen dieser Unterschiede jedoch nicht weiter untersucht.

In den vergangenen Jahren sind international eine Reihe von Studien entstanden, welche die Determinanten der Auszugswahrscheinlichkeit zum Gegenstand haben. Ein großer Teil der Studien ist soziologischen Ursprungs. Einen Überblick über die internationale soziologische Literatur hierzu gibt White (1994). Häufig finden in diesen Arbeiten auch bestimmte ökonomische Erklärungsfaktoren eine Berücksichtigung. Aus ökonomischer Sicht kommen als Determinanten vor allem das Einkommen der Kinder und das der Eltern als Erklärungsfaktor in Frage. Damit in Verbindung können auch der Erwerbsstatus und die Ausbildung eine Rolle spielen. Ein weiterer Erklärungsfaktor sind die Mietpreise (der Preis für das Gut, das mit der Auszugsentscheidung nachgefragt wird).

Di, Yang und Liu (2002) untersuchen insbesondere ethnische Unterschiede und Geschlechtsunterschiede im Auszugsverhalten junger Menschen in den USA. Sie verwenden Daten des Current Population Survey und der Panel Study of Income Dynamics für die

### 3 Forschungsstand

Schätzung logistischer Regressionsmodelle. Sie stellen fest, dass das persönliche Einkommen der wichtigste Faktor ist, um das Haushaltsarrangement zu erklären, in welchem 25 bis 34 jährige Personen leben. Wer ein geringeres Einkommen hat, lebt mit einer größeren Wahrscheinlichkeit bei den Eltern. Ein weiterer Schwerpunkt der Untersuchung ist die Erklärung längerfristiger Änderungen im Auszugsverhalten. In den 90er Jahren hat das Auszugsalter in den USA (ähnlich wie in anderen Industrienationen) zugenommen. Diesen Trend erklären Di et al. (2002) mit dem steigenden Anteil junger Menschen, die nur über ein geringes Einkommen verfügen.

Buck und Scott (1993) arbeiten ebenfalls mit der Panel Study of Income Dynamics und schätzen Hazardratenmodelle für die Auszugswahrscheinlichkeit in den 70er und 80er Jahren. Auch sie finden einen positiven Einfluss des eigenen Einkommens auf die Auszugswahrscheinlichkeit. Für den Einfluss des Elterneinkommens finden sie uneinheitliche Resultate. Sie stellen weiterhin fest, dass eine hohe nationale Arbeitslosigkeit die Auszugswahrscheinlichkeit verringert, eine hohe regionale Arbeitslosigkeit hat hingegen keinen Einfluss.

Börsch-Supan (1986) untersucht die Auszugswahrscheinlichkeit in den USA mit einem Querschnittsdatensatz von 1976 und schätzt hierzu ein multinomiales Logitmodell. Als Ziel-Alternativen unterscheidet er nicht verschiedene Haushaltsarrangements, sondern eigene Wohnung und Mietwohnung sowie verschiedene Wohnungsgrößen. Als Einflussfaktoren identifiziert er das eigene Einkommen und die Mietpreise.

Haurin, Haurin, Hendershott und Bourassa (1997) untersuchen die Auszugswahrscheinlichkeit in Australien. Auch sie finden einen signifikanten Einfluss der eigenen Verdienstmöglichkeiten auf die Wahrscheinlichkeit, das Elternhaus zu verlassen. Sie schätzen Probit-Modelle mit Hilfe des Australian Longitudinal Surveys. In dieser Arbeit wird auch zwischen der Zielwohnform Single und Zusammenleben mit anderen unterschieden. Dies schätzen sie jedoch mit einem separaten Probit-Modell. Ein höheres Einkommen führt demnach zu einer höheren Wahrscheinlichkeit in eine Singlewohnung zu ziehen, statt mit anderen zusammen zu ziehen. Dies deutet darauf hin, dass es eine Präferenz für Privatsphäre gibt.

Eine methodisch ähnliche Untersuchung (Haurin, Hendershott und Kim, 1993, 1994) mit Daten des National Longitudinal Survey of Youth aus USA stellt ebenfalls fest, dass ein höheres eigenes Einkommen und geringere Mietpreise die Wahrscheinlichkeit, dass junge Menschen einen eigenen Haushalt gründen, erhöht.

Mit der Frage, welche Rolle der Sozialstaat für die Auszugsentscheidung spielt, befasst sich Hu (2000). Er untersucht insbesondere, welche Auswirkung die Veränderung

### 3 Forschungsstand

von staatlichen Transferleistungen an die Kinder und an die Eltern haben, die bezahlt werden, wenn das Kind noch im Elternhaus lebt. Er arbeitet mit einem Probit-Modell und stellt für die USA fest, dass die Wahrscheinlichkeit das Elternhaus zu verlassen um so größer ist, je kleiner der Verlust an Sozialtransfers ist, den die Eltern durch den Auszug erleiden.

Avery, Goldschneider und Speare (1992) legen ihren Schwerpunkt auf die Frage, welchen Einfluss die Ressourcen der Eltern im Auszugsprozess haben. Sie verwenden den Survey of Income and Programm Participation aus USA von 1984 um diese Frage zu klären. Sie stellen fest, dass der Einfluss des elterlichen Einkommens davon abhängt, ob die Kinder das Elternhaus verlassen um eine Familie zu gründen oder um alleine zu wohnen. Unterschiede zwischen den Geschlechtern wurden nicht gefunden.

Card und Lemieux (1997) untersuchten den Einfluss der Arbeitsmarktbedingungen, denen sich junge Menschen gegenüber sehen, auf das Auszugsverhalten. In ihrem Vergleich zwischen USA und Canada stellen sie fest, dass die schlechte Arbeitsmarktsituation für Jugendliche in Canada dazu geführt hat, dass Anfang der 90er Jahre der Anteil der Jugendlichen, die bei den Eltern leben, im Vergleich zur USA zugenommen hat. Sie verwenden für ihre Regressionsanalyse Mikrodaten, die nach Region, Geschlecht und Alter gruppiert sind.

Sehr unterschiedliche Muster des Auszugsverhalten finden auch Billari, Philipov und Baizan (2001). Sie untersuchen 16 europäische Länder mit Hilfe des Family and Fertility Surveys. Ihre Untersuchung schließt auch osteuropäische Länder mit ein. Auch in den Ländern Polen und Ungarn ist das Auszugsalter überdurchschnittlich hoch. Sie weisen darauf hin, dass das Auszugsverhalten in den nordeuropäischen Staaten am homogensten ist. In Südeuropa hingegen ist die Bandbreite des Auszugsalters höher. Sie deuten dies als Hinweis darauf, dass in Südeuropa Unterschiede in den Präferenzen und in den Budgetrestriktionen eine größere Bedeutung haben als in Nordeuropa. Sie vermuten, dass es verschiedene Arrangements der Unterstützung der Generationen untereinander gibt, die diese Unterschiede erklären. In Ländern mit stärkeren Familien-Systemen unterstützen die Eltern länger ihre heranwachsenden Kinder. Söhne und Töchter unterstützen aber auch stärker ihre Eltern im Alter. Länder mit schwächeren Familien-Systemen sind stärker auf Unterstützungszahlungen des Staates angewiesen.

Aassve, Billari, Mazzuco und Ongaro (2001) untersuchen, ob sich die Muster des Auszugsverhaltens in unterschiedlichen sozialstaatlichen Systemen unterscheiden. Sie arbeiten mit Daten des Europäischen Haushaltspanel (ECHP). Aus zehn Europäischen Ländern verwenden sie Daten der in der ersten Welle (1994) 18 bis 34jährigen. Sie verwenden

### 3 Forschungsstand

den eine Verallgemeinerung des Heckman-Selektions-Modells, um die Auszugsentscheidung zu schätzen und stellen fest, dass in Süd-Europäischen Ländern der Einfluss des Erwerbsstatus sehr groß ist. In Skandinavischen Ländern spielen Beschäftigungsstatus und Einkommen eine vernachlässigbare Rolle. In den weiteren kontinentaleuropäischen Ländern ist der Einfluss gemischt.

Ermisch und DiSalvo (1997) haben das Auszugsverhalten junger Menschen für England mit Hazardratenmodellen untersucht, wobei sie sich insbesondere auf ökonomische Faktoren konzentrierten. Sie fanden einen signifikanten Einfluss der Wohnungs- und Arbeitsmarktbedingungen sowie des Einkommens der Eltern und des eigenen Einkommens entsprechend der erwarteten Vorzeichen.

Eine weitere Untersuchung für England liefern Murphey und Wang (1998). Sie schätzen ein Hazardratenmodell mit zufälligen Effekten mit Daten aus dem British Household Panel Study. Ökonomische Einflussfaktoren werden hier nur wenig berücksichtigt. Sie stellen fest, dass eine höhere Qualifikation der Mutter die Auszugswahrscheinlichkeit erhöht. Wenn die Mutter berufstätig ist, verringert sich die Auszugswahrscheinlichkeit.

In USA, Canada und England zeigt sich also, dass junge Menschen auf Arbeitslosigkeit reagieren, indem sie länger im Elternhaus bleiben. Ein höheres eigenes Einkommen erhöht die Wahrscheinlichkeit, den Haushalt der Eltern zu verlassen. Auch für öffentliche Transferzahlungen findet sich ein Einfluss auf die Auszugsentscheidung. Für die Rolle des Elterneinkommens sind die Ergebnisse nicht einheitlich. In den skandinavischen Ländern hingegen haben Faktoren wie Elterneinkommen oder Arbeitsmarktsituation keinen bedeutenden Einfluss auf die Auszugsentscheidung.

Die bisherigen Untersuchungen zum Auszugsverhalten in der Bundesrepublik kommen zum Teil zu unterschiedlichen Ergebnissen. Datengrundlage waren meist die ALLBUS-Daten oder das Sozio-ökonomische Panel. Zu den Ergebnissen, bei welchen Übereinstimmung in den Studien besteht, zählt, dass Männer später das Elternhaus verlassen als Frauen (Hullen, 1995; Schwarz, 1989; Wagner und Huinink, 1991; Weick, 1993).

Auf Arbeitslosigkeit als erklärender Faktor ist bisher nicht eingegangen worden. Andere Ereignisse des Erwerbsverlaufes wie der Beginn der Ausbildung wurden von manchen Studien berücksichtigt. So stellen Ott (1986) und Mayer und Wagner (1986) fest, dass mit Ausbildungsbeginn oder mit Aufnahme einer Erwerbstätigkeit die Auszugswahrscheinlichkeit erhöht wird.

In mehreren internationalen Untersuchungen (zum Beispiel Ermisch und DiSalvo (1997) für England und Buck und Scott (1993); Richards, Withe und Tsui (1987) für die USA) wird zwischen verschiedenen Haushaltsformen im Zielzustand unterscheiden. In den bis-

### 3 Forschungsstand

herigen Untersuchungen für die Bundesrepublik wurde eine derartige Unterscheidung nicht getroffen. Lediglich Hullen (1995) berücksichtigt verschiedene Kombinationen von Ereignissen, wie Auszug mit und ohne Heirat. Es zeigt sich, dass das Auszugsalter bei jenen höher ist, bei denen der Auszug mit einer Heirat oder der Geburt eines Kindes verbunden war. Auch in Hullen's Untersuchung werden jedoch ökonomische Erklärungsfaktoren außer Acht gelassen.

Eine Reihe von Untersuchungen hatte ihren Schwerpunkt auf der Veränderung des Auszugsverhaltens über verschiedene Kohorten hinweg. So stellen Wagner und Huinink (1991) einen Rückgang des Auszugsalters für die Jahrgänge bis 1955 fest. Bei dem späteren Jahrgängen ist das Auszugsalter wieder angestiegen.

Wagner und Huinink (1991) machen für das zunehmende Auszugsalter eine Reihe von Faktoren verantwortlich. Die zunehmende Bedeutung des Berufseintrittes für die Auszugswahrscheinlichkeit kann dies nur zum Teil erklären. Insbesondere in den Kohorten 1963-1965 zeigte sich, dass die Auszugsraten unabhängig vom Eintritt in das Erwerbsleben gefallen sind. Wagner und Huinink (1991) nennen als Faktoren, die eine Rolle spielen, die Bildungsexpansion und Engpässe am Wohnungsmarkt. Zum Teil spielt auch das spätere Heiratsalter eine Rolle. Die Untersuchungen kommen also zu dem Ergebnis, dass ein höheres eigenes Einkommen die Auszugswahrscheinlichkeit erhöht. Der Einfluss des Elterneinkommens ist uneinheitlich. Meist führt aber ein höheres Einkommen der Eltern zu einer geringeren Wahrscheinlichkeit das Elternhaus zu verlassen.

In dieser Arbeit soll nun mit Daten des Sozioökonomischen Panels untersucht werden, welche Bedeutung diese ökonomischen Faktoren in West-Deutschland haben. Es werden die Wellen von 1984 bis 1999 verwendet. Die Untersuchung umfasst damit mehr Kohorten als die Arbeiten mit den ECHP, das erst seit 1994 erhoben wurde. Anders als in den meisten bisherigen Arbeiten zur Bundesrepublik Deutschland wird auch zwischen verschiedenen möglichen Haushaltsarrangements unterschieden, welche die jungen Menschen eingehen, wenn sie das Elternhaus verlassen. Stetige Kovariablen werden, anders als in den bisherigen Studien, semi-parametrisch modelliert.

## 4 Theoretischer Rahmen

Die Haushaltsstruktur als Lösung von Nutzenmaximierungsproblemen wurde zuerst von Becker (1981) analysiert. Die Haushaltsformations-Entscheidung lässt sich damit zurückführen auf die Präferenzen der beteiligten Individuen und deren zur Verfügung stehenden Ressourcen. Das Auszugsverhalten wird dort allerdings nicht explizit modelliert. Das erste Modell, welches das Auszugsverhalten analysiert findet sich in McElroy (1985). Für eine ausführlicher Diskussion der verschiedenen Modellierungen der Auszugsentscheidung in der Literatur wird auf das Kapitel 11 der vorliegenden Arbeit verwiesen. Die bisherigen Modelle des Auszugsverhalten junger Menschen unterscheiden lediglich die Zustände 'lebt bei den Eltern' und 'zieht aus' (siehe auch Ermisch (1999); Ghidoni (2001)). Verschiedene Zielzustände (Single-Wohnung, Lebensgemeinschaft) werden nicht unterschieden. Das Modell des Auszugsverhaltens von Ermisch und DiSalvo (1997) bzw. Ermisch (1999) wird im folgenden verallgemeinert, so dass verschiedene Haushaltsarrangements, in welche die Kinder einziehen, unterschieden werden können.

Der Index  $c$  kennzeichnet jeweils Variablen, die sich auf die Kinder beziehen,  $p$  bezieht sich auf Eltern. Die Variable  $C$  steht für den Konsum.

Die Nutzenfunktion des Kindes lautet

$$U_c = U_c(C_c, j). \quad (4.1)$$

$j$  kennzeichnet die verschiedenen Haushaltsarrangements, die hier unterschieden werden.

$$j = \begin{cases} in & \text{lebt bei den Eltern} \\ out_{part} & \text{mit Partner/in} \\ out_s & \text{Single} \end{cases} \quad (4.2)$$

Der Nutzen wird maximiert unter der Nebenbedingung:

$$y_c + T(j) = C_c. \quad (4.3)$$

$T(j)$  sind Transfers in Form von Geld oder Gütern der Eltern an die Kinder,  $y$  ist das Einkommen. Die Höhe der Transfers hängt davon ab, ob die Kinder im elterlichen Haushalt

#### 4 Theoretischer Rahmen

leben. Es wird angenommen, die Eltern sind gegenüber ihren Kindern altruistisch. Der Nutzen der Eltern ist also eine Funktion des Nutzens der Kinder. Die Nutzenfunktion der Eltern lautet

$$U_p = U_p(C_p, j, U^c(C_c, j)), \quad (4.4)$$

die Budgetrestriktion

$$y_p = C_p + T(j). \quad (4.5)$$

In der Literatur wird oft unterstellt, dass der Nutzen der Eltern steigt, wenn das Kind bei ihnen wohnt. Der Nutzen des Kindes hingegen steigt, wenn es das Elternhaus verlässt, da es die Privatsphäre einer eigenen Wohnung vorzieht. Blanco und Kluve (2002) untersuchen die Frage, ob es ein „Gut“ oder ein „Schlecht“ für die Eltern ist, wenn die Kinder im Hause sind für Deutschland. Sie fanden jedoch eher für die gegenteilige These empirische Unterstützung. Für Eltern sinkt demnach der Nutzen, wenn die Kinder im Hause leben. Söhne und Töchter fühlen sich nach Blanco und Kluve (2002) hingegen ganz wohl, wenn sie bei den Eltern leben. Die grundlegende Struktur des hier verwendeten Modells ist jedoch unabhängig davon, ob diese These nun zutrifft. Ein anderer Grund dafür, dass das Zusammenleben der Eltern mit den Kindern den Nutzen der Eltern steigern könnte, ist, dass Transfers der Eltern an die Kinder in diesem Fall günstiger sind. Transfers können dann stattfinden in Form von gemeinsam benutzten Gütern, wie zum Beispiel der Wohnraum, Geräte wie eine Waschmaschine usw. Sie sind innerhalb des Haushaltes ein öffentliches Gut (vgl. Rosenzweig und Wolpin, 1993).

In der ersten Stufe des Entscheidungsprozesses maximieren die Eltern ihren Nutzen unter Kenntnis der Präferenzstruktur des Kindes. Dabei entscheiden sie auch über die Transfers an das Kind. In der zweiten Stufe maximiert das Kind seinen Nutzen, gegeben die Entscheidung der Eltern.

Die indirekte Nutzenfunktion des Kindes, die sich aus diesen Entscheidungen ergibt, lautet

$$V_j = V_j(y_c, y_p, \mathbf{x}, u_j). \quad (4.6)$$

$u_j$  ist eine stochastische Nutzenkomponente. Die Kinder werden sich entscheiden das Elternhaus zu verlassen, wenn deren Nutzen außerhalb des Elternhauses höher ist als im Elternhaus, also wenn (für  $j \neq in$ )

$$V_{in} < V_j. \quad (4.7)$$

Die Nutzenfunktion wird zur Vereinfachung linear spezifiziert:

$$V_j = \beta_{0j} + \beta_{1j}y_c + \beta_{2j}y_p + \beta_{3j}\mathbf{x} + u_j. \quad (4.8)$$

#### 4 Theoretischer Rahmen

Die Bedingung 4.7 lässt sich dann auch schreiben als

$$\beta_{0j} - \beta_{0in} + (\beta_{1j} - \beta_{1in})y_c + (\beta_{2j} - \beta_{2in})y_p + (\beta_{3j} - \beta_{3in})\mathbf{x} + u_{in} - u_j > 0, \quad (4.9)$$

Wir definieren  $\alpha_{kj} = \beta_{kj} - \beta_{k,in}$  und  $\epsilon_j = u_{in} - u_j$ , ( $j \neq in$ ). Damit ergibt sich die Bedingung für einen Auszug wie folgt:

$$\alpha_{0j} + \alpha_{1j}y_c + \alpha_{2j}y_p + \alpha_{3j}\mathbf{x} > -\epsilon_j. \quad (4.10)$$

Wenn  $\epsilon_j$  logistisch verteilt ist, lässt sich die Wahrscheinlichkeit in den Zustand  $j$  zu wechseln schreiben als:

$$p_j(y_c, y_p, \mathbf{x}) = \frac{e^{\alpha_{0j} + \alpha_{1j}y_c + \alpha_{2j}y_p + \alpha_{3j}\mathbf{x}}}{1 + \sum_l e^{\alpha_{0l} + \alpha_{1l}y_c + \alpha_{2l}y_p + \alpha_{3l}\mathbf{x}}}. \quad (4.11)$$

Der Index  $l$  läuft über die möglichen Zielzustände  $j \neq in$ . Die Koeffizienten dieser Funktion sind durch ein geeignetes Verfahren zu schätzen. Die Gleichung lässt sich im schätztheoretischen Rahmen allgemeiner linearer Modelle (generalized linear models, GLM), wie sie in Nelder und Wedderburn (1972) beschrieben werden, formulieren. Es gilt

$$p_j(y_c, y_p, \mathbf{x}) = h_j(\eta_j), \quad (4.12)$$

wobei  $h_j(\eta_j)$  die Response-Funktion ist und  $\eta_j$  der lineare Prädiktor:

$$\eta_j = \alpha_{0j} + \alpha_{1j}y_c + \alpha_{2j}y_p + \alpha_{3j}\mathbf{x}. \quad (4.13)$$

Ermisch (1999); Ermisch und DiSalvo (1997) analysieren dieses Modell mit einer CES-Nutzenfunktion und kommen damit zu der Erwartung, dass  $\alpha_{1j} > 0$  und  $\alpha_{2j} < 0$ . Mit steigendem Einkommen des Kindes sollte also die Wahrscheinlichkeit eines Auszugs zunehmen. Mit zunehmendem Einkommen der Eltern nimmt sie ab.

Solange eine Person bei den Eltern lebt, muss die Entscheidung, in einer bestimmten Periode aus dem Elternhaus auszuziehen oder dort zu verweilen, in jeder Periode neu getroffen werden. Die Perioden werden mit  $t$  gekennzeichnet.  $t$  zählt hierbei die Jahre nach dem Erreichen des 16-ten Lebensjahres. Wird die Auszugswahrscheinlichkeit in Gleichung 4.11 um diesen Zeitindex ergänzt, lässt sie sich als Hazardrate interpretieren. Sie gibt die Wahrscheinlichkeit an zum Zeitpunkt  $t$  in den Zustand  $j$  zu wechseln, gegeben die Person befindet sich in  $t$  noch im Elternhaus. Der Prädiktor  $\eta_j$  wird um einen Term  $\theta_{jt}$  ergänzt, der die Abhängigkeit der Auszugsneigung vom Alter modellieren soll, und lautet damit

$$\eta_{jt} = \theta_{jt} + \alpha_{0j} + \alpha_{1j}y_{ct} + \alpha_{2j}y_{pt} + \alpha_{3j}\mathbf{x}_t. \quad (4.14)$$

#### 4 Theoretischer Rahmen

Die Auszugswahrscheinlichkeit ist dann

$$p_{jt}(y_{ct}, y_{pt}, \mathbf{x}_t) = \frac{e^{\theta_{jt} + \alpha_{0j} + \alpha_{1j}y_{ct} + \alpha_{2j}y_{pt} + \alpha_{3j}\mathbf{x}_t}}{1 + \sum_l e^{\theta_{jt} + \alpha_{0l} + \alpha_{1l}y_{ct} + \alpha_{2l}y_{pt} + \alpha_{3l}\mathbf{x}_t}}. \quad (4.15)$$

# 5 Das ökonometrische Modell

## 5.1 Lineare Modellierung

Es wird für das ökonometrische Modell davon ausgegangen, dass junge Menschen eine bestimmte Dauer im Elternhaus verweilen und dann einen eigenen Haushalt gründen. Der Zielzustand kann lauten: ‚Single-Wohnung‘ oder ‚Wohnung mit Partner‘. Es ist also ein Modell erforderlich, welches den Wechsel vom Zustand ‚Haushalt der Eltern‘ in den Zustand ‚Zielhaushalt‘ abbildet. Hierfür ist ein Hazardratenmodell geeignet. Da die Beobachtungen jährlich vorliegen, wird von einem diskreten Hazardratenmodell ausgegangen<sup>1</sup>.

Die Wahrscheinlichkeit, dass das Individuum  $i$  in der Periode  $t$  in den Zielzustand  $j$  wechselt, gegeben es wohnt am Anfang dieser Periode noch bei den Eltern, wird bestimmt durch:

$$h_{ijt} = P(T_i = t, J_{iT_i} = j | T_i \geq t; \mathbf{x}_{it}) \quad (5.1)$$

$T_i$  und  $J_{iT_i}$  sind Zufallsvariablen, die angeben, nach wie viel Jahren  $T_i$  das Individuum  $i$  in den Zustand  $J_{iT_i} \in \{out_{part}, out_s\}$  wechselt. Die Hazardrate wird hier entsprechend der Gleichung 4.15 als eine logistische Funktion spezifiziert.

$$h_{ijt} = \frac{e^{\theta_{jt} + \alpha_j \mathbf{x}_{it}}}{1 + \sum_l e^{\theta_{lt} + \alpha_l \mathbf{x}_{it}}} \quad (5.2)$$

Die Kovariablen wurden mit dem Index  $i$  für die Individuen ergänzt. Die Koeffizienten  $\alpha_{1j}$  sind zur Vereinfachung der Schreibweise im Vektor  $\alpha_j$  zusammengefasst. Die entsprechenden Kovariablen sind im Vektor  $\mathbf{x}_{it}$  zusammengefasst.

Um die Likelihood-Funktion herzuleiten, die zur Schätzung der gesuchten Parameter erforderlich ist, sind mehrere Schritte notwendig. Wir definieren die Auszugswahrscheinlichkeit

$$h_{it} = \sum_{j=1}^m h_{ijt}. \quad (5.3)$$

---

<sup>1</sup>zum Folgenden vgl. Jenkins (1995), Allison (1982) oder Wolff (1998)

## 5 Das ökonometrische Modell

$m$  ist die Zahl der alternativen Zustände, in welche die Person wechseln kann. Damit ist  $(1 - h_{it})$  die Wahrscheinlichkeit, in der Periode  $t$  im Ausgangszustand zu verweilen, gegeben die Person ist zu Beginn der Periode in diesem Zustand. Die unbedingte Verweilwahrscheinlichkeit ist das Produkt dieser bedingten Wahrscheinlichkeiten  $(1 - h_{it})$  über alle Perioden bis zum Zeitpunkt  $t_i$ . Dies ist die Überlebensfunktion

$$S_{it_i} = \prod_{t=1}^{t_i} (1 - h_{it}), \quad (5.4)$$

in unserem Fall die Wahrscheinlichkeit, dass eine Person  $i$  am Ende des Beobachtungszeitraumes  $t_i$  noch bei den Eltern lebt. Dies ist also die Wahrscheinlichkeit, dass die Beobachtung der Person  $i$  rechtszensiert ist. Bei Beobachtungen die nicht rechtszensiert sind, ist am Ende der Beobachtungsdauer ein Ereignis  $j \in \{out_{part}, out_s\}$  zu beobachten. Die (unbedingte) Wahrscheinlichkeit, dies zu beobachten ist

$$P(t_i = t, j) = h_{ijt} \prod_{t=1}^{t_i-1} (1 - h_{it}). \quad (5.5)$$

Um die Likelihood-Funktion formulieren zu können, wird eine Indikatorvariable definiert:

$$\delta_i = \begin{cases} 1 & \text{die Beobachtung von Individuum } i \text{ ist unzensiert} \\ 0 & \text{sonst.} \end{cases} \quad (5.6)$$

Die Likelihood-Funktion lautet damit

$$L = \prod_{i=1}^n \prod_{j=1}^m \left[ \left( h_{ijt_i} \prod_{t=1}^{t_i-1} (1 - h_{it}) \right)^{\delta_i} \prod_{t=1}^{t_i} (1 - h_{it})^{1-\delta_i} \right], \quad (5.7)$$

wobei  $n$  die Zahl der beobachteten Individuen ist. Die Gleichung lässt sich umformen zu

$$L = \prod_{i=1}^n \left[ \prod_{j=1}^m \left( \frac{h_{ijt_i}}{1 - h_{it}} \right)^{\delta_i} \prod_{t=1}^{t_i} (1 - h_{it}) \right]. \quad (5.8)$$

Wir definieren eine neue Variable  $y_{ijt}$  wie folgt:

$$y_{ijt} = \begin{cases} 1 & \text{bei Wechsel von Individuum } i \text{ in Zustand } j \in \{out_{part}, out_s\} \text{ in } t = t_i \\ 0 & \text{sonst,} \end{cases} \quad (5.9)$$

ziehen das Produktzeichen nach vorne und bilden die Loglikelihood-Funktion

$$\log L = \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} y_{ijt} \log [h_{ijt} / (1 - h_{it})] + \sum_{i=1}^n \sum_{t=1}^{t_i} \log (1 - h_{it}). \quad (5.10)$$

Der Ausdruck  $(1 - h_{it})$  kann aus dem ersten Term herausgezogen werden:

$$\log L = \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} y_{ijt} \log h_{ijt} - \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} y_{ijt} \log (1 - h_{it}) + \sum_{i=1}^n \sum_{t=1}^{t_i} \log (1 - h_{it}). \quad (5.11)$$

Der Logarithmus der Verweilwahrscheinlichkeit in der Situation, in welcher ein Auszug zu beobachten ist, wird also im dritten Term addiert und im zweiten Term abgezogen. Daher lässt sich der Ausdruck auch schreiben als

$$\log L = \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} y_{ijt} \log h_{ijt} + \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} (1 - y_{ijt}) \log (1 - h_{it}). \quad (5.12)$$

Setzen wir schließlich die Hazardrate wie in Gleichung (5.2) spezifiziert ein, folgt daraus die Loglikelihood-Funktion

$$\log L = \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} y_{ijt} \log \left( \frac{e^{\theta_{jt} + \alpha_j \mathbf{x}_{it}}}{1 + \sum_l e^{\theta_{jt} + \alpha_l \mathbf{x}_{it}}} \right) + \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^{t_i} (1 - y_{ijt}) \log \left( \frac{1}{1 + \sum_l e^{\theta_{jt} + \alpha_l \mathbf{x}_{it}}} \right). \quad (5.13)$$

Die Maximum-Likelihood-Funktion, die sich für das Modell ergibt, entspricht der eines multinomialen Logit-Modelles. Da dieses Verfahren in zahlreichen statistischen Softwarepaketen implementiert ist, kann dieses Hazardraten-Modell einfach geschätzt werden.

## 5.2 Semiparametrische Modellierung

In den vergangenen Jahren wurden semiparametrische Verfahren zunehmend weiterentwickelt (vgl. Härdle und Linton, 1994). Diese Verfahren ermöglichen es, Zusammenhänge flexibler zu modellieren, als es parametrische Verfahren können.

In den bisherigen Arbeiten zum Haushaltsformationsverhalten wurde der Einfluss metrischer Variablen linear modelliert. Eine Ausnahme ist die Arbeit von Murphey und Wang (1998), die verschiedene demographische Variablen mit linearen Splines modellieren.

Insbesondere bei Variablen wie dem eigenen Einkommen und dem Einkommen der Eltern ist ein nicht-linearer Zusammenhang zu vermuten. Denkbar ist etwa, dass bei geringerem Elterneinkommen ein Anstieg des Einkommens der Eltern die Auszugswahrscheinlichkeit verringert, da es die Kinder attraktiver finden im Elternhaus zu bleiben. Ab einem höheren Einkommen kann sich die Auszugswahrscheinlichkeit aber erhöhen,

weil die Eltern so hohe Transfers an die Kinder zahlen, dass diese sich leichter eine eigene Wohnung finanzieren können.

Auch bei einer weiteren Kovariable, die im Modell aufgenommen wird, den Jahren seit Arbeitsbeginn, ist ein nicht-linearer Effekt zu erwarten. In den ersten Jahren dürfte der Effekt größer sein. Wenn der Arbeitsbeginn bereits mehrere Jahre zurück liegt, ist von einem zusätzlichen Jahr kein großer Einfluss mehr zu erwarten. Daher wird in dieser Arbeit neben einer linearen Modellierung auch ein semiparametrisches Modell geschätzt, das weniger restriktive Annahmen über den Einfluss der Kovariablen zugrunde legt.

Semiparametrische Glättungsverfahren mit B-Splines werden in Eilers und Marx (1996) diskutiert. Die Anwendung auf multikategoriale Daten wird in Tutz und Scholz (2000) dargestellt. Ein Vorteil dieses Ansatzes ist, dass er ebenso wie das oben dargestellte lineare Modell im Rahmen der parametrischen verallgemeinerten linearen Modelle (GLM, Nelder und Wedderburn (1972)) geschätzt werden kann.

Der Prädiktor  $\eta_{jt}$  wird nun nur von den Variablen im Vektor  $\mathbf{x}_t$  linear beeinflusst. Der Zusammenhang der Variablen  $y_{kt}$  mit dem Prädiktor wird mit einer allgemeinen Funktion beschrieben:

$$\eta_{jt} = \alpha_{0jt} + \sum_{k=1}^q \gamma_{kj}(y_{kt}) + \alpha_{1j} \mathbf{x}_t. \quad (5.14)$$

$q$  ist die Zahl der Variablen, die flexibel modelliert werden. Die Funktion  $\gamma_{kj}(y)$  ist eine Linearkombination aus Basisfunktionen:

$$\gamma_{kj}(y) = \sum_{s=1}^{M_k} \gamma_{kjs} G_{ks}(y) \quad (5.15)$$

Eine Basisfunktion  $G_{ks}(y)$  ist jeweils verbunden mit einem Knoten  $y_{k(s)}$ . Die Knoten liegen im Wertebereich der entsprechenden Variablen im gleichen Abstand voneinander entfernt.

Die Basisfunktionen (die B-Splines) bestehen aus mehreren zusammengeführten Abschnitten von Polynomen. Der einfachste Fall wäre ein B-Spline vom Grade 1. Diese Basisfunktion hat eine einfache Dreiecksform  $G_{ks}(y) = |y - y_{k(s)}|$  im Intervall  $[y_{k(s)} - 1, y_{k(s)} + 1]$ . Ein B-Spline vom Grade 2 besteht aus drei zusammengesetzten Abschnitten eines Polynoms 2. Grades und hat eine Glockenform. Dieser B-Spline erstreckt sich über 4 Knoten (siehe Abbildung 5.1). Für die Schätzung in dieser Arbeit wurden B-Splines 2. Grades verwandt.

Die Frage, welche Zahl von Knoten die beste ist, ist Gegenstand intensiver Diskussion. Bei zu vielen Knoten wird die Kurve zu gut an die Beobachtungen angepasst, so dass keine Struktur zu erkennen ist. Bei zu wenig Knoten verläuft die Funktion zu glatt.

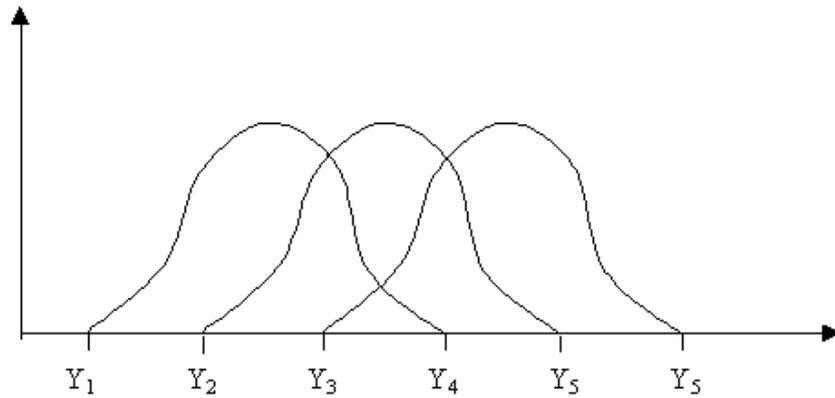


Abbildung 5.1: Illustration von B-Splines 2. Grades

Anstatt weniger Knoten zu wählen um eine hinreichend glatte Funktion zu erhalten, ist es jedoch auch möglich, viele Knoten zu wählen und die Differenz zwischen den Gewichten benachbarter Basisfunktionen gering zu halten. Zu große Sprünge in der Funktion sollen also bestraft werden.

Formal ist dies zu erreichen, indem der Likelihoodfunktion noch ein Bestrafungsterm (Penalisierungsterm) für zu große Veränderungen benachbarter  $\gamma_{kjs}$  angefügt wird. Zu maximieren ist dann die penalisierte Likelihoodfunktion

$$PL = \log L - \frac{1}{2} \cdot Pen \quad (5.16)$$

Der Penalisierungsterm  $Pen$  lautet:

$$Pen = \sum_{k=1}^q \lambda_k \sum_{j=1}^m \sum_{s=d_{kj}+1}^{P_k} (\Delta^{d_{kj}} \gamma_{kjs})^2 \quad (5.17)$$

$s$  ist der Index für die Knoten.  $\Delta$  ist ein Differenzoperator.  $d_{kj}$  ist der gewählte Grad der Differenz für die Variable  $k$  und dem Zielzustand  $j$ . In dieser Arbeit wurde wie bei diesem Verfahren üblich  $d_{kj} = 2$  gewählt. Das heißt, die Differenz der Differenz benachbarter  $\gamma_{kjs}$  wird bestraft. Mit anderen Worten: Die Steigung der Funktion darf sich nicht zu stark ändern. Mit dem Parameter  $\lambda_k$  lässt sich die gewünschte Glätte der Funktion einstellen. Um den optimalen Wert für  $\lambda_k$  zu finden, sind verschiedene Verfahren möglich.

Ein plausibles Kriterium, um die beste Anpassung von  $\lambda_k$  zu finden, ist das Akaike-Informationen-Kriterium (AIC), welches die Loglikelihood der Daten um die Zahl der

zu schätzenden Parameter korrigiert. Es lassen sich jedoch keine Aussagen über die Konkavität dieser Zielfunktion machen. Die Zielfunktion kann in Abhängigkeit von den Daten sehr unregelmäßig verlaufen und mehrere lokale Optima haben. Algorithmen zur Suche eines Optimums wie z.B. der Newton-Raphson-Algorithmus versagen hier.

In dieser Arbeit wurde daher ein genetischer Algorithmus verwendet, um die optimalen  $\lambda_k$  zu finden. Genetische Algorithmen<sup>2</sup> sind stochastische heuristische Suchverfahren, die bei hochkomplexen Zielfunktionen ein Optimum finden sollen, und sind von Holland (1975) entwickelt worden. Die  $\lambda_k$  für die  $q$  verschiedenen Variablen bilden zusammen einen String, das Genom, das alle genetischen Informationen einer Einheit umfasst. Ein genetischer Algorithmus beginnt mit einer Population von Strings mit verschiedenen, zufällig gebildeten  $\lambda_k$ , die zu unterschiedlich guten Anpassungen der geschätzten Kurve an die Daten führen. Die Strings, die für eine bessere Anpassung sorgen (für einen besseren Fit), reproduzieren sich in der nächsten Generation der Population mit einer höheren Wahrscheinlichkeit als andere. Zufällige Mutation und Rekombination sorgen dafür, dass sich ein Teil der  $\lambda_k$  in der nächsten Generation ändert. Nach einigen Generationen haben sich stabile Werte der  $\lambda_k$  durchgesetzt.

---

<sup>2</sup>Eine Darstellung genetischer Algorithmen findet sich z.B. in Goldberg (1989)

## 6 Datenquelle und deskriptive Ergebnisse

Bei den Umfragen des SOEP wird in jeder Welle nach der Stellung der befragten Person zum Haushaltsvorstand gefragt. Diese Informationen für alle Personen eines Haushaltes konnten verwendet werden, um die Stellung einer Person im Haushalt und die Haushaltsform zu bestimmen. Die Veränderungen der Stellung im Haushalt und der Haushaltsform können dann über den Beobachtungszeitraum verfolgt werden. Es wurde eine Variable ‚Haushaltsform‘ kreiert, die diese Informationen enthält. Wer angab, dass er oder sie Kind oder Enkel des Haushaltsvorstandes sei, erhielt die Ausprägung ‚bei den Eltern lebend‘. Enkel des Haushaltsvorstandes sind knapp 2% der Beobachtungen im Datensatzes.

Am Ende der Beobachtungsdauer wird zunächst unterschieden zwischen: bei den Eltern lebend, Single, mit einem Partner/einer Partnerin lebend, Familie (mit Partner und wenigstens einem Kind), alleinerziehend oder mit sonstigen anderen Personen in einem Haushalt lebend. Letzteres sind zum Beispiel Wohngemeinschaften. Tabelle 6.1 zeigt die Häufigkeiten dieser Zustände, die im verwendeten Datensatz auftauchen. Wie sich zeigt, zieht gut die Hälfte der Personen in eine Single-Wohnung. Knapp 47% derer, die das Elternhaus verlassen, gründen mit dem Partner (zum Teil bereits mit den eigenen Kindern) einen Haushalt. In Tabelle 6.1 werden auch die Altersgruppen der bis 21jährigen und der über 22jährigen unterschieden. Die jüngere Gruppe zieht in über der Hälfte der Fälle in Single-Wohnungen. Die ältere Gruppe zieht häufiger als die jüngere Gruppe mit einem Partner zusammen.

Für diese Untersuchung wurden im Datensatz jene Personen aufgenommen, die im Beobachtungszeitraum (1984 bis 1999) 16 Jahre alt wurden. Diese Personen wurden so lange weiterverfolgt, bis sie erstmals nicht mehr bei ihren Eltern wohnten.

Damit fallen jene heraus, die zwar noch bei den Eltern wohnen, aber zu Beginn des Beobachtungszeitraumes bereits älter sind. Würden diese mit aufgenommen, käme es zu einer Überschätzung des Auszugsalters, weil Gleichaltrige, die bereits vor dem Beobachtungszeitraum ausgezogen sind und über deren Auszugszeitpunkt mithin keine Information vorliegt, in der betrachteten Stichprobe nicht auftauchen. Es wären also zuviele Personen, die später ausgezogen sind, in der Stichprobe.

## 6 Datenquelle und deskriptive Ergebnisse

Zielhaushaltsform			Alter: bis 21		Alter: ab 22		
	n	Prozent	n	Prozent	n	Prozent	
Eltern	872	48,26	594		278		
Single	456	25,24	48,77	308	51,25	148	44,31
mit Partner/-in	355	19,65	37,97	216	35,94	139	41,62
Familie	80	4,43	8,56	52	8,65	28	8,38
alleinerziehend	10	0,55	1,07	8	1,33	2	0,60
mit anderen Personen	34	1,88	3,64	17	2,83	17	5,09
Total	1807	100	100	1195	100	612	100

Tabelle 6.1: Zustand am Ende der Beobachtungsdauer

Goldscheider et al. (1993) weisen darauf hin, dass der Auszug aus dem Elternhaus kein punktuellere Ereignis, sondern ein längerer Prozess, welcher über verschiedene Phasen laufen kann. Junge Menschen können zum Beispiel in einer ersten Phase getrennt von den Eltern zur Untermiete oder in einem Studentenwohnheim leben, haben aber selber noch keinen eigenen Haushalt gegründet. Ein häufiges Problem in den bisherigen Untersuchungen ist auch die Behandlung von Männern, die ihren Wehrdienst absolvieren. In Erhebungen des Allbus oder des GSOEP wird dieses Datum von 10% der Männer bei der Frage nach dem ‚Auszug aus dem Elternhaus‘ angegeben (siehe Ziegler und Schladt, 1993; Wagner und Huinink, 1991). Ein großer Teil davon zieht nach dem Aufenthalt in der Kaserne wieder zu den Eltern zurück. Sie haben also noch keinen eigenen Haushalt gegründet. Der Ablösungsprozess kann sich also über mehrere Jahre hinweg ziehen. Für eine empirische Untersuchung mit Hazardratenmodellen ist es jedoch erforderlich, einen bestimmten Zeitpunkt festzulegen.

In dieser Arbeit wird auf die Frage im Erhebungsbogen nach der Stellung zum Haushaltsvorstand zurückgegriffen. Als Auszug aus dem Elternhaus wird nur ein Ereignis gewertet, in welchem die Person tatsächlich in einen neuen Haushalt einzieht. Personen, die vom Elternhaus abwesend sind, weil sie in Kasernen o.ä. leben, führen keinen eigenen Haushalt. Ihr Haushaltsvorstand ist nach wie vor der Vater oder die Mutter. Ein Auszug im hier verwendeten Sinn findet erst statt, wenn eine Person selber Haushaltsvorstand, Partner des Haushaltsvorstandes oder ähnliches wird.

Vor der Diskussion der Einflussfaktoren auf die Auszugswahrscheinlichkeit wird zunächst die Verweildauer im Elternhaus empirisch dargestellt. Die empirische Überlebensrate (Abbildung 6.1) gibt an, wie groß der Anteil derer ist, die im Alter von  $n$  Jahren noch bei ihren Eltern leben. Die Funktion fällt am stärksten in der Mitte des dritten Lebens-

## 6 Datenquelle und deskriptive Ergebnisse

Alter	Total	Auszug	Zensiert	Hazardrate
17	1807	30	129	0,0174
18	1648	84	131	0,0545
19	1433	106	106	0,0799
20	1221	138	96	0,1250
21	987	117	63	0,1304
22	807	126	69	0,1776
23	612	101	54	0,1890
24	457	83	59	0,2150
25	315	54	43	0,2026
26	218	34	35	0,1853
27	149	22	29	0,1781
28	98	24	23	0,3221
29	51	16	17	0,4638
30	18	1	17	0,1111
Summe	9821	936	871	

Tabelle 6.2: Beobachtete Dauer im Elternhaus

jahrzehnts.

Die Auszugswahrscheinlichkeit in einem bestimmten Alter, gegeben, die Person wohnt in diesem Alter noch bei den Eltern, wird durch die empirische Hazardrate dargestellt. Sie ist in der Abbildung 6.2 abgebildet. In den ersten Jahren ist die Auszugswahrscheinlichkeit noch gering, steigt aber von Jahr zu Jahr. Ab dem 25. Lebensjahr nimmt die Auszugswahrscheinlichkeit wieder langsam ab. Eine Spitze in der Hazardrate gibt es im Alter von 29 Jahren. Sie ist jedoch schwer zu interpretieren, da dort nur noch wenige Beobachtungen vorliegen.

Bei den nach Geschlecht getrennten Überlebensraten (Abbildung 6.3) ist in der Literatur bereits oft festgestellte Phänomen zu erkennen, dass Männer länger im Elternhaus verbleiben als Frauen. 50% der Männer sind mit 25 Jahren noch im Elternhaus. Bei den Frauen sind mit 23 Jahren über die Hälfte schon aus dem Elternhaus ausgezogen.

Ein deskriptiver Überblick über die Kovariablen, welche für die Schätzung des Hazardratenmodelles verwandt wurden, findet sich in Tabelle 6.4. Eine Reihe von Faktoren, welche Einfluss auf das Auszugsverhalten junger Menschen haben können, beziehen sich auf Eigenschaften der Eltern. Unterschiede in den Präferenzen, etwa von normativen Vorstellungen, können zum Teil durch das Bildungsniveau der Eltern abgebildet werden.

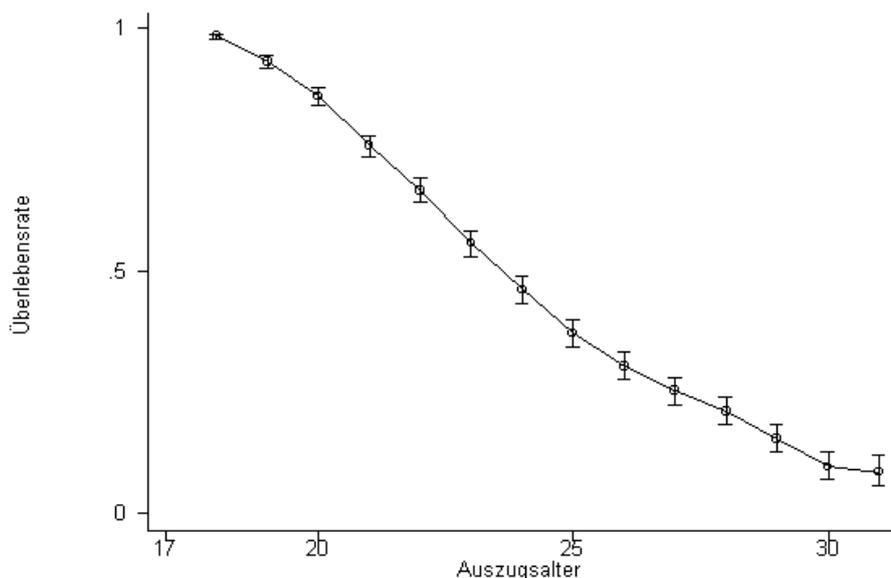


Abbildung 6.1: Überlebensrate

*Anmerkung: Die vertikalen Linien in dieser und den folgenden Abbildungen geben das 95% Konfidenzintervall an.*

Die entsprechenden Dummyvariablen für den Vater bzw. die Mutter sind eins, wenn deren höchster Schulabschluss Realschule oder höher ist. Wie Tabelle 6.4 zeigt, trifft dies bei der Elterngeneration auf etwa 20% des Datensatzes zu. Als Charakteristikum des Elternhaushaltes, welches die Präferenzen der Heranwachsenden beeinflussen könnte, wurde außerdem die Haushaltsgröße aufgenommen.

Das Einkommen der Eltern und das der Kinder sind angegeben als Nettojahreseinkommen in Tausend DM und wurden mit dem Preisindex des privaten Verbrauches der Volkswirtschaftlichen Gesamtrechnung des DIW bezogen auf das Jahr 1996 deflationiert. Es umfasst auch Transfers des Staates wie zum Beispiel Arbeitslosengeld oder Sozialhilfe. Je mehr Personen ein Haushalt umfasst, auf desto mehr Köpfe muss das Elterneinkommen verteilt werden. Um dies zu berücksichtigen, wurde das Einkommen der Eltern mit Hilfe von Äquivalenzskalen entsprechend der Gleichung

$$\text{Äquivalentes Einkommen} = \frac{\text{Einkommen}}{1 + 0.5 * (\text{Erwachsene} - 1) + 0.3 * \text{Kinder}} \quad (6.1)$$

transformiert. Dies entspricht der modifizierten OECD-Skala, die in internationalen Studien üblich ist und zudem ungefähr den im Bundessozialhilfegesetz implizierten Bedarfs-

## 6 Datenquelle und deskriptive Ergebnisse

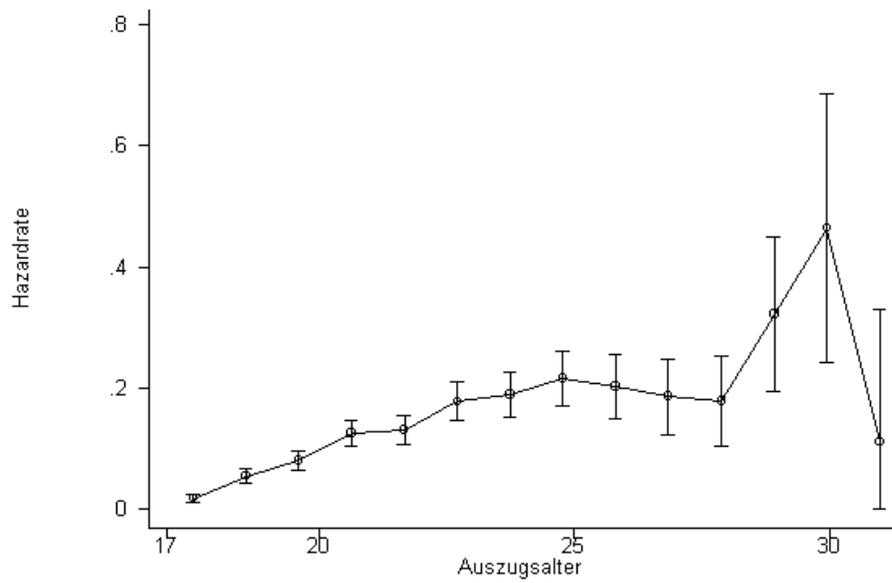


Abbildung 6.2: Hazardrate

Mehrfachnennungen	n	Prozent
0	3837	39,07
1	4452	45,33
2	1345	13,70
3	177	1,80
4	10	0,10
	9821	100,00

Tabelle 6.3: Mehrfachnennungen bei Erwerbsstatus

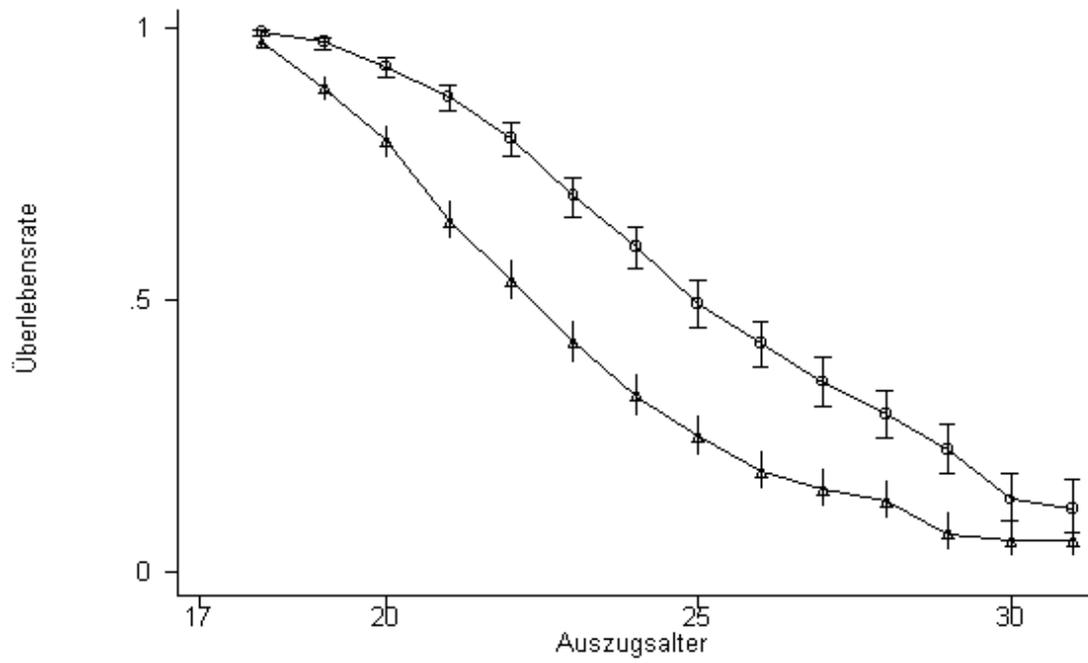


Abbildung 6.3: Überlebensraten nach Geschlecht

Anmerkung:  $\Delta$  steht für Frauen,  $\circ$  steht für Männer

gewichten entspricht (vgl. Faik, 1997).

Die Ressourcen, welche einer Person unmittelbar zufließen, werden mit bestimmt durch deren Erwerbsstatus. Für die Kennzeichnung des Erwerbsstatus wurden in der vorliegenden Arbeit Daten verwandt, welche von der GSOEP-Projektgruppe aus monatlich vorliegenden Daten zu Jahresdaten aggregiert wurden.

Die Befragungen des GSOEP haben in den meisten Fällen im Februar und März stattgefunden. Wenn es eine Änderung in der Stellung zum Haushaltsvorstand gab, hat dies zwischen den Befragungszeitpunkten stattgefunden. Die Variablen, welche über den Erwerbsstatus Auskunft geben, beziehen sich auf das Vorjahr des jeweiligen Befragungszeitpunktes. Fünf verschiedene Zustände des Erwerbsstatus werden unterschieden. In knapp 40% der Fälle hatten die Personen während des ganzen Jahres den gleichen Erwerbsstatus. Mehrfachnennungen gab es bei etwa 60% der Fälle (Siehe Tabelle 6.3). Unter dem Erwerbsstatus ‚sonstiges‘ wurden zusammengefasst: Hausfrauen, Wehrdienst und Personen im Ruhestand (dies sind meist Empfänger von Hinterbliebenenrente).

Eine Person, die sich in Ausbildung befindet, voll erwerbstätig oder teilzeitbeschäftigt ist, verfügt über ein eigenes Einkommen. Daher ist zu erwarten, dass sich deren Auszugswahrscheinlichkeit erhöht.

Wer arbeitslos gemeldet ist, war mit höherer Wahrscheinlichkeit vorher beschäftigt, als Personen in anderem Erwerbsstatus. Um zu verhindern, dass die Variable Arbeitslosigkeit auch diese Arbeitsbiographie reflektiert, wurden die Jahre seit dem Arbeitsbeginn in das Hazardratenmodell aufgenommen. Der Erwerbsstatus Arbeitslosigkeit in einem Jahr kann häufiger mit anderen Zuständen, die mit Beschäftigung einhergehen, verbunden sein, als es in anderen Zuständen der Fall ist, da bei den Arbeitslosen die Fluktuation höher ist, als etwa bei Auszubildenden oder voll Erwerbstätigen. Um diese Korrelationen aufzufangen wurde die Variable ‚Arbeitslos\_Beschäftigt‘ in das Modell aufgenommen. Diese Dummyvariable nimmt den Wert eins an, wenn die Person im selben Jahr sowohl arbeitslos als auch in Ausbildung, voll- oder teilzeitbeschäftigt war.

Weiterhin wurde der Bildungsabschluss in der Untersuchung berücksichtigt. Ein höherer Bildungsabschluss erhöht das Humankapital und damit das mögliche Arbeitseinkommen. Es ist auch denkbar, dass eine höhere Qualifikation das Bedürfnis, unabhängig von den Eltern zu leben, erhöht. Beide Wirkungsketten würden die Auszugswahrscheinlichkeit erhöhen. Unterschieden wird zwischen Personen ohne Abschluss, Personen, deren höchster Bildungstitel Haupt- bzw. Realschulabschluss ist, Abitur oder ein sonstiger Abschluss.

## 6 Datenquelle und deskriptive Ergebnisse

Variable	Mittelwert	Standardabweichung
weiblich	0,427	0,495
Vater: Realschule (od. höher)	0,223	0,416
Mutter: Realschule (od. höher)	0,191	0,393
Alter	19,101	2,945
Haushaltsgröße	4,181	1,347
Erwerbsstatus		
Schule/Studium	0,525	0,499
Ausbildung	0,304	0,460
Arbeitslos	0,060	0,238
Erwerbstätig	0,304	0,460
sonstiges	0,118	0,323
Arbeitslos*Erwerbst.	0,035	0,183
Jahre seit Arbeitsbeginn	1,430	2,275
Bildungsabschl.:		
kein Schulabschluss	0,265	0,442
Haupt-, Realsch.	0,578	0,494
Abitur	0,132	0,339
sonst. Bildungsabschl.	0,024	0,153
Einkommen Eltern	22,760	16,865
Eigenes Einkommen	9,111	13,023
Arbeitslosengeld/-hilfe	0,130	1,027

Tabelle 6.4: Mittelwerte und Standardabweichungen der Variablen

*Anmerkung: Die Mittelwerte und Standardabweichungen beziehen sich auf den Datensatz, in welchem für jede Person pro Jahr eine Beobachtung vorliegt. Einkommensdaten sind angegeben als Nettojahreseinkommen in Tausend DM.*

## 7 Schätzergebnisse

Für das Hazardratenmodell werden die fünf Zielzustände aus Tabelle 1 zu zwei Zielzustände zusammengefasst, da sonst die Fallzahlen für die einzelnen Gruppen zu gering wären. Personen, die mit anderen (WGs) zusammenleben, werden zu den Singles gezählt. Die Zustände ‚alleinerziehend‘ und ‚Familie‘ werden zum Zustand ‚mit Partner‘ gezählt. Die Gruppen werden also nach dem Kriterium der Lebensphase, in der sie sich befinden, unterschieden.

In den meisten Studien zum Auszugsverhalten zeigt sich, dass sich das Verhalten von Männern und Frauen deutlich unterscheidet. Daher werden hier für Frauen und Männer getrennte Hazardratenmodelle geschätzt. Zudem wurde für das hier geschätzte Modell die Hypothese, dass für Männer und Frauen die Koeffizienten gleich sind, mit einem Likelihood-Ratio-Test abgelehnt.

Ein methodisches Problem könnte sich ergeben, wenn junge Menschen das Elternhaus nur kurzzeitig verlassen, um später dort wieder einzuziehen. Die endgültige Gründung eines neuen Haushaltes wäre erst später. Dadurch würde die Auszugswahrscheinlichkeit überschätzt. In der Bundesrepublik kommt dieses Phänomen jedoch selten vor. Werden die Biographien der Personen in der Stichprobe weiter verfolgt, zeigt sich, dass im Beobachtungszeitraum (also bis 1999) nur 0,8% derer, die das Elternhaus bereits verlassen haben, wieder zurückgezogen sind.

Ein möglicher Kritikpunkt an der hier gewählten Modellierung könnte sein, dass die verwendeten Kovariablen des Erwerbsstatus endogen sind, da der Erwerbsstatus möglicherweise von den gleichen Variablen beeinflusst wird wie das Haushaltsformationsverhalten. Eine Schätzung mit der verzögerten Variable, der Erwerbsstatus ein Jahr früher, zeigte jedoch, dass das Modell robust gegenüber dieser Modifikation ist. Hier werden dennoch die Ergebnisse der Schätzung ohne verzögerte Erwerbsstatus-Variable präsentiert, da durch die Verzögerung der Beobachtungszeitraum um ein Jahr kürzer ist und entsprechend viele Beobachtungen wegfallen. Es ist im Übrigen plausibel, den Erwerbsstatus als exogene Größe zu betrachten. In der Literatur wird meist davon ausgegangen, dass der Erwerbsstatus junger Menschen vor allem von der Nachfrageseite bestimmt wird (hierzu Card und Lemieux, 1997) und damit nicht endogen bestimmt ist.

Bei den Erwerbsstatus-Variablen wurde der Status Schule/Studium als Referenzgruppe verwendet und daher im Hazardratenmodell ausgelassen. Bei den Variablen zum Bildungsabschluss bilden die Referenzgruppe die Personen ohne Abschluss.

## 7.1 Lineares Modell

Die Ergebnisse der Schätzung des Hazardratenmodelles sind in den Tabellen 7.1 und folgende ausgewiesen. Es zeigt sich, dass beim Bildungsniveau der Eltern lediglich eine höhere Qualifikation der Mutter auf die Auszugswahrscheinlichkeit der Tochter einen signifikanten, positiven Einfluss hat. Hier können Unterschiede in der Sozialisation eine Rolle spielen.

Einen hochsignifikanten Einfluss auf die Auszugswahrscheinlichkeit hat das Alter der Personen. Referenzgruppe sind Personen im Alter von 16 bis 18 Jahren. Jene, die 25 oder älter sind, wurden zu einer Gruppe zusammengefasst, weil bei einer weiteren Unterteilung in der Gruppe der Älteren zu wenig Beobachtungen vorgelegen wären. Der größte Koeffizient zeigt sich für die Altersgruppe der 23 bis 24-jährigen. Dies bestätigt das Ergebnis der empirischen Hazardrate in Abbildung 6.2. Eine Ausnahme findet sich bei Männern, die mit einer Partnerin zusammen ziehen. Dort findet sich die größte Auszugswahrscheinlichkeit bei den über 24-jährigen.

Die Veränderungen des Auszugsverhaltens über die Kohorten hinweg wurde mit Indikatorvariablen für Kohorten von jeweils 3 Jahren modelliert. Die Referenzgruppe sind die Jahrgänge von 1968 bis 1970. Bei späteren Jahrgängen sinkt bei Männern die Wahrscheinlichkeit, mit einer Partnerin zusammen zu ziehen. Die Wahrscheinlichkeit in eine Single-Wohnung zu ziehen steigt bei den späteren Jahrgängen. Dies kann zum einen mit dem Trend erklärt werden, dass das Heiratsalter in den 80er Jahren zugenommen hat, zum anderen damit, dass zwischen der Lebensphase, in der man bei den Eltern lebt, und der Familiengründungsphase immer häufiger eine Phase eingeschoben wird, in der junge Menschen alleine leben. Bei den Frauen lassen sich keine signifikanten Kohorteneffekte finden.

Unter den Variablen, welche den Erwerbsstatus kennzeichnen, sind die Koeffizienten für Ausbildung und Beschäftigung signifikant (Referenzgruppe ist Schule). Personen, die beschäftigt oder in Ausbildung sind, ziehen mit einer größeren Wahrscheinlichkeit in einen Haushalt mit einem Partner oder einer Partnerin. Auch die Variable Jahre seit ersten Job ist signifikant. Je länger eine Person bereits arbeitet, desto größer ist die Auszugswahrscheinlichkeit.

## 7 Schätzergebnisse

Bei Männern, die arbeitslos sind, ist die Wahrscheinlichkeit mit einer Partnerin zusammenzuziehen geringer als bei Beschäftigten. Ein Likelihood-Ratio-Test zeigt allerdings, dass hier die Koeffizienten Beschäftigt und Arbeitslos nicht signifikant voneinander verschieden sind. In allen anderen Fällen spielt ‚Arbeitslos‘ im Vergleich zu ‚Beschäftigt‘ keine Rolle. Junge Menschen, die arbeitslos sind, bleiben also im Vergleich zu Beschäftigten für gewöhnlich nicht länger im Elternhaus um die elterlichen Ressourcen stärker nutzen zu können.

Die Haushaltsgröße hat bei den Männern einen signifikant positiven Einfluss auf die Wahrscheinlichkeit mit einer Partnerin zusammenzuziehen. Als Ursache dafür sind Unterschiede in den Präferenzen denkbar. Wer in einem Haushalt mit vielen Personen aufgewachsen ist, hat möglicherweise eine geringere Neigung in einem Single-Haushalt zu leben.

Bei den Frauen sind die Erwerbsstatusvariablen Beschäftigt und sonstiges (dies sind vor allem Hausfrauen) signifikant positiv. Der Koeffizient Arbeitslos ist in der Zielkategorie ‚Mit Partner‘ ebenfalls signifikant positiv. Hier unterscheidet sich das Verhalten der Frauen deutlich von dem der Männer. Das Einkommen der Eltern hat bei Frauen einen signifikanten Einfluss auf die Wahrscheinlichkeit in einen Single-Haushalt zu ziehen. Mit höherem Elterneinkommen nimmt die Auszugswahrscheinlichkeit zu. Für Deutschland zeigt sich damit ein anderer Befund, als er für andere Länder gefunden wurde (vgl. Avery et al., 1992; Ghidoni, 2001). Dies deutet darauf hin, dass die Transfers der Eltern an die Kinder nicht sehr stark eingeschränkt werden, wenn das Kind das Elternhaus verlässt. Möglicherweise kommt dieser Effekt zustande durch junge Frauen, die aufgrund des Studiums das Elternhaus verlassen und in eine Single-Wohnung ziehen. Studierend haben überdurchschnittlich häufig Eltern, die über ein höheres Einkommen verfügen. Die Bildungs-Koeffizienten zeigen, dass Abiturientinnen mit einer größeren Wahrscheinlichkeit in eine Singlewohnung ziehen als Personen ohne Abschluss. Anlass zum Auszug aus dem Elternhaus dürfte hier häufig der Studienbeginn sein.

Die Modelle in Tabelle 7.1 und 7.2 wurden ohne die Variable Einkommen des Kindes geschätzt, da hier Endogenitätsprobleme denkbar sind. In der Schätzung in Tabellen 7.3 und 7.4 wurde das Einkommen jedoch mit berücksichtigt. Ein höheres Einkommen erhöht erwartungsgemäß die Auszugswahrscheinlichkeit. Bei Männern erhöht sich die Wahrscheinlichkeit mit einer Partnerin zusammen zu ziehen signifikant. Die Jahre seit dem ersten Job sind weniger signifikant als in der ersten Modellvariante. Da für gewöhnlich das Einkommen mit der Berufserfahrung zunimmt, hat diese Variable einen Teil des Einflusses des eigenen Einkommens aufgefangen. Die Signifikanz und Größe des Koeffi-

zienten Beschäftigung nimmt zwar ab, bleibt aber meist noch signifikant.

Von Interesse ist die Frage, ob der geringe Einfluss von Arbeitslosigkeit auf die Auszugsentscheidung darauf zurückzuführen ist, dass Arbeitslose Arbeitslosengeld und -hilfe beziehen, so dass der Einkommensausfall durch Beschäftigungslosigkeit nicht relevant ist. Um dies zu klären wurden zwei weitere Varianten geschätzt (Tabelle 7.5 und 7.6) in denen als Kovariable zusätzlich der Betrag der erhaltenen Arbeitslosenhilfe und des Arbeitslosengeldes aufgenommen wurde. Die Variable eigenes Einkommen wurde in diesen Modellen um diesen Betrag reduziert. Die Koeffizienten Arbeitslos ändern sich nur minimal und die Koeffizienten werden im Vergleich zu den anderen Modellvarianten nicht signifikanter. Es lässt sich also kein Hinweis darauf finden, dass Arbeitslosigkeit aufgrund von Arbeitslosengeld-Transfers keinen Einfluss auf die Auszugsentscheidung hat.

Die Größenordnung der Koeffizienten in der Schätzung des Hazardratenmodelles lässt sich inhaltlich nur begrenzt interpretieren, da sich an diesen nicht unmittelbar ablesen lässt, wie sich die Auszugswahrscheinlichkeit ändert, wenn sich eine Kovariable ändert. Die Deutung der Schätzergebnisse wird durch die Berechnung der Wahrscheinlichkeitsänderung bei Änderung einer Kovariable erleichtert (siehe hierzu Ronning, 1991, S. 42). Daher wurden die prognostizierten Hazardraten der Modellvariante 2 jeweils für den Fall, dass die Personen in Ausbildung, Arbeitslos bzw. voll Erwerbstätig sind, ermittelt. Hierzu wurden für jede Person im Datensatz die Kovariablen in die Hazardratenfunktion eingesetzt und so die Auszugswahrscheinlichkeit der Person bestimmt, wobei die Variable Ausbildung auf 1, Arbeitslos und voll Erwerbstätig auf 0 gesetzt wurden. Von den sich daraus ergebenden Hazardraten wurde das arithmetische Mittel gebildet. Dieser Wert gibt die durchschnittliche Auszugswahrscheinlichkeit an, wenn alle Personen im Datensatz in Ausbildung wären. Das gleiche wurde für die Ausprägungen Arbeitslos und voll Erwerbstätig durchgeführt (recycled predictions). Die Ergebnisse finden sich in der Tabelle 7.7.

Es zeigt sich, dass bei Männern, die arbeitslos sind, die Wahrscheinlichkeit in eine Single-Wohnung zu ziehen sehr viel höher ist als in einen Haushalt mit einer Partnerin zu ziehen.

Tabelle 7.8 zeigt, wie sich die Verweil- bzw. Auszugswahrscheinlichkeiten ändern, wenn sich der Erwerbsstatus von Ausbildung oder Beschäftigt zu Arbeitslos ändert. Hierzu wurden die entsprechende Differenz der Werte aus der Tabelle 7.7 gebildet. Im Vergleich zu Beschäftigten Frauen und Männern ist die Auszugswahrscheinlichkeit bei arbeitslosen marginal größer. Im Vergleich zu Personen, die sich in Ausbildung befinden, nimmt die Wahrscheinlichkeit bei Männern um 0,7%-Punkte zu, bei Frauen sinkt sie um 4 %-

Tabelle 7.1: Hazardratenmodell 1, Männer

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realschule	-0.112	(0.205)
Mutter: min. Realschule	0.176	(0.200)
Alter 19-20	1.003**	(0.292)
21-22	1.501**	(0.306)
23-24	1.841**	(0.336)
> 24	1.622**	(0.391)
Jahrg. 70-73	0.080	(0.164)
74-76	-0.092	(0.221)
77-79	0.742**	(0.284)
Haushaltsgröße	-0.093	(0.058)
Erwerbsstat.: Ausbildung	0.271	(0.203)
Arbeitslos	0.561	(0.464)
Beschäftigt	0.345 <sup>†</sup>	(0.204)
sonstiges	0.282	(0.207)
besch.*arbeitsl.	-0.145	(0.521)
Jahre seit ersten Job	0.096**	(0.035)
Bildungsabschl.: Haupt-, Realsch.	0.499*	(0.232)
Abitur	0.670*	(0.270)
sonstiges	-0.289	(0.559)
Einkommen Eltern	0.004	(0.004)
Konstante	-5.582**	(0.440)
Mit Partnerin		
Vater: min. Realschule	-0.322	(0.260)
Mutter: min. Realschule	0.057	(0.254)
Alter 19-20	2.211**	(0.619)
21-22	3.217**	(0.620)
23-24	3.453**	(0.641)
> 24	3.541**	(0.670)
Jahrg. 70-73	-0.380*	(0.185)
74-76	-0.648*	(0.286)
77-79	-0.149	(0.491)
Haushaltsgröße	0.131**	(0.049)
Erwerbsstat.: Ausbildung	0.399 <sup>†</sup>	(0.218)
Arbeitslos	-1.328	(1.032)
Beschäftigt	0.503*	(0.231)
sonstiges	-0.201	(0.257)
besch.*arbeitsl.	1.540	(1.061)
Jahre seit ersten Job	0.070 <sup>†</sup>	(0.038)
Bildungsabschl.: Haupt-, Realsch.	0.562*	(0.285)
Abitur	0.362	(0.335)
sonstiges	0.136	(0.511)
Einkommen Eltern	0.005	(0.005)
Konstante	-7.674**	(0.704)
N		5623
Log-likelihood		-1550.022
$\chi^2_{(44)}$		491.866
Signifikanzlevel : <sup>†</sup> : 10%   * : 5%   ** : 1%		

Anmerkung: Ausgelassene Kategorien sind die 16 bis 18-jährigen, die Jahrgänge 1968 und 1969, der Erwerbsstatus Schüler und der Bildungsabschluss ‚keiner‘; Einkommensdaten sind angegeben als Nettojahreseinkommen in Tausend DM.

Tabelle 7.2: Hazardratenmodell 1, Frauen

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realschule	-0.079	(0.197)
Mutter: min. Realschule	0.652**	(0.201)
Alter 19-20	0.826**	(0.193)
21-22	1.063**	(0.230)
23-24	1.213**	(0.288)
> 24	0.808*	(0.382)
Jahrg. 70-73	-0.218	(0.164)
74-76	-0.097	(0.186)
77-79	-0.185	(0.236)
Haushaltsgröße	-0.089	(0.057)
Erwerbsstat.: Ausbildung	0.330 <sup>†</sup>	(0.172)
Arbeitslos	0.385	(0.540)
Beschäftigt	0.493**	(0.178)
sonstiges	0.732**	(0.202)
besch.*arbeitsl.	0.259	(0.592)
Jahre seit ersten Job	0.037	(0.043)
Bildungsabschl.: Haupt-, Realsch.	0.371 <sup>†</sup>	(0.190)
Abitur	0.470*	(0.235)
sonstiges	0.263	(0.626)
Einkommen Eltern	0.007 <sup>†</sup>	(0.004)
Konstante	-4.480**	(0.371)
Mit Partner		
Vater: min. Realschule	-0.244	(0.221)
Mutter: min. Realschule	0.268	(0.232)
Alter 19-20	0.647**	(0.201)
21-22	0.970**	(0.229)
23-24	1.063**	(0.279)
> 24	0.834*	(0.363)
Jahrg. 70-73	-0.098	(0.156)
74-76	-0.155	(0.195)
77-79	-0.092	(0.252)
Haushaltsgröße	0.018	(0.050)
Erwerbsstat.: Ausbildung	0.024	(0.168)
Arbeitslos	0.851*	(0.417)
Beschäftigt	1.120**	(0.180)
sonstiges	1.325**	(0.163)
besch.*arbeitsl.	-0.311	(0.467)
Jahre seit ersten Job	0.074 <sup>†</sup>	(0.040)
Bildungsabschl.: Haupt-, Realsch.	0.367 <sup>†</sup>	(0.189)
Abitur	-0.094	(0.267)
sonstiges	0.382	(0.471)
Einkommen Eltern	-0.003	(0.005)
Konstante	-4.612**	(0.366)
N		4198
Log-likelihood		-1821.703
$\chi^2_{(44)}$		489.546

Signifikanzlevel : † : 10% \* : 5% \*\* : 1%

Tabelle 7.3: Hazardratenmodell 2, Männer

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realschule	-0.100	(0.206)
Mutter: min. Realschule	0.199	(0.201)
Alter 19-20	0.968**	(0.293)
21-22	1.430**	(0.312)
23-24	1.750**	(0.344)
> 24	1.517**	(0.401)
Jahrg. 70-73	0.071	(0.164)
74-76	-0.100	(0.221)
77-79	0.723*	(0.284)
Haushaltsgröße	-0.094	(0.058)
Erwerbsstat.: Ausbildung	0.277	(0.203)
Arbeitslos	0.610	(0.465)
Beschäftigt	0.300	(0.208)
sonstiges	0.307	(0.208)
besch.*arbeitsl.	-0.159	(0.522)
Jahre seit ersten Job	0.088*	(0.036)
Bildungsabschl.: Haupt-, Realsch.	0.465*	(0.233)
Abitur	0.713**	(0.272)
sonstiges	-0.315	(0.559)
Eigenes Einkommen	0.007	(0.005)
Einkommen Eltern	0.004	(0.004)
Konstante	-5.563**	(0.441)
Mit Partnerin		
Vater: min. Realschule	-0.303	(0.260)
Mutter: min. Realschule	0.118	(0.257)
Alter 19-20	2.163**	(0.619)
21-22	3.102**	(0.624)
23-24	3.295**	(0.648)
> 24	3.349**	(0.679)
Jahrg. 70-73	-0.403*	(0.185)
74-76	-0.659*	(0.286)
77-79	-0.181	(0.492)
Haushaltsgröße	0.128**	(0.049)
Erwerbsstat.: Ausbildung	0.426 <sup>†</sup>	(0.219)
Arbeitslos	-1.228	(1.034)
Beschäftigt	0.416 <sup>†</sup>	(0.238)
sonstiges	-0.146	(0.259)
besch.*arbeitsl.	1.514	(1.061)
Jahre seit ersten Job	0.060	(0.039)
Bildungsabschl.: Haupt-, Realsch.	0.513 <sup>†</sup>	(0.286)
Abitur	0.441	(0.338)
sonstiges	0.100	(0.512)
Eigenes Einkommen	0.011 <sup>†</sup>	(0.005)
Einkommen Eltern	0.005	(0.005)
Konstante	-7.656**	(0.704)
N		5623
Log-likelihood		-1547.481
$\chi^2_{(46)}$		496.948
Signifikanzlevel : † : 10% * : 5% ** : 1%		

Tabelle 7.4: Hazardratenmodell 2, Frauen

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realschule	-0.049	(0.198)
Mutter: min. Realschule	0.696**	(0.203)
Alter 19-20	0.755**	(0.196)
21-22	0.883**	(0.241)
23-24	0.941**	(0.307)
> 24	0.578	(0.394)
Jahrg. 70-73	-0.232	(0.164)
74-76	-0.121	(0.187)
77-79	-0.208	(0.236)
Haushaltsgröße	-0.080	(0.057)
Erwerbsstat.: Ausbildung	0.341*	(0.174)
Arbeitslos	0.476	(0.542)
Beschäftigt	0.379*	(0.186)
sonstiges	0.790**	(0.203)
besch.*arbeitsl.	0.234	(0.593)
Jahre seit ersten Job	0.006	(0.045)
Bildungsabschl.: Haupt-, Realsch.	0.307	(0.192)
Abitur	0.550*	(0.237)
sonstiges	0.192	(0.629)
Eigenes Einkommen	0.021**	(0.008)
Einkommen Eltern	0.007†	(0.004)
Konstante	-4.502**	(0.373)
Mit Partner		
Vater: min. Realschule	-0.220	(0.222)
Mutter: min. Realschule	0.294	(0.233)
Alter 19-20	0.601**	(0.203)
21-22	0.857**	(0.239)
23-24	0.888**	(0.297)
> 24	0.682†	(0.374)
Jahrg. 70-73	-0.103	(0.156)
74-76	-0.165	(0.195)
77-79	-0.100	(0.252)
Haushaltsgröße	0.023	(0.050)
Erwerbsstat.: Ausbildung	0.043	(0.170)
Arbeitslos	0.909*	(0.419)
Beschäftigt	1.054**	(0.185)
sonstiges	1.371**	(0.166)
besch.*arbeitsl.	-0.322	(0.467)
Jahre seit ersten Job	0.057	(0.041)
Bildungsabschl.: Haupt-, Realsch.	0.326†	(0.191)
Abitur	-0.037	(0.269)
sonstiges	0.352	(0.471)
Eigenes Einkommen	0.013†	(0.007)
Einkommen Eltern	-0.003	(0.005)
Konstante	-4.622**	(0.367)
N		4198
Log-likelihood		-1816.692
$\chi^2_{(46)}$		499.568
Signifikanzlevel : † : 10% * : 5% ** : 1%		

Tabelle 7.5: Hazardratenmodell 3, Männer

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realschule	-0.100	(0.206)
Mutter: min. Realschule	0.200	(0.201)
Alter 19-20	0.968**	(0.293)
21-22	1.427**	(0.312)
23-24	1.747**	(0.344)
> 24	1.510**	(0.403)
Jahrg. 70-73	0.068	(0.165)
74-76	-0.101	(0.221)
77-79	0.723*	(0.284)
Haushaltsgröße	-0.094	(0.058)
Erwerbsstat.: Ausbildung	0.277	(0.203)
Arbeitslos	0.595	(0.473)
Beschäftigt	0.303	(0.208)
sonstiges	0.307	(0.208)
besch.*arbeitsl.	-0.152	(0.523)
Jahre seit ersten Job	0.088*	(0.036)
Bildungsabschl.: Haupt-, Realsch.	0.463*	(0.234)
Abitur	0.715**	(0.272)
sonstiges	-0.319	(0.560)
Eigenes Einkommen	0.007	(0.005)
Arbeitslosengeld/-hilfe	0.014	(0.040)
Einkommen Eltern	0.004	(0.004)
Konstante	-5.562**	(0.441)
Mit Partnerin		
Vater: min. Realschule	-0.305	(0.260)
Mutter: min. Realschule	0.114	(0.257)
Alter 19-20	2.163**	(0.619)
21-22	3.105**	(0.624)
23-24	3.299**	(0.648)
> 24	3.357**	(0.680)
Jahrg. 70-73	-0.400*	(0.185)
74-76	-0.659*	(0.286)
77-79	-0.183	(0.492)
Haushaltsgröße	0.128**	(0.049)
Erwerbsstat.: Ausbildung	0.425 <sup>†</sup>	(0.219)
Arbeitslos	-1.184	(1.040)
Beschäftigt	0.411 <sup>†</sup>	(0.239)
sonstiges	-0.144	(0.259)
besch.*arbeitsl.	1.495	(1.062)
Jahre seit ersten Job	0.061	(0.039)
Bildungsabschl.: Haupt-, Realsch.	0.514 <sup>†</sup>	(0.286)
Abitur	0.439	(0.338)
sonstiges	0.112	(0.513)
Eigenes Einkommen	0.011 <sup>†</sup>	(0.005)
Arbeitslosengeld/-hilfe	-0.009	(0.056)
Einkommen Eltern	0.005	(0.005)
Konstante	-7.654**	(0.704)
N		5623
Log-likelihood		-1547.394
$\chi^2_{(48)}$		497.122
Signifikanzlevel : <sup>†</sup> : 10%   * : 5%   ** : 1%		

Tabelle 7.6: Hazardratenmodell 3, Frauen

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realschule	-0.048	(0.198)
Mutter: min. Realschule	0.697**	(0.203)
Alter 19-20	0.754**	(0.196)
21-22	0.878**	(0.242)
23-24	0.936**	(0.308)
> 24	0.570	(0.396)
Jahrg. 70-73	-0.233	(0.164)
74-76	-0.122	(0.187)
77-79	-0.209	(0.236)
Haushaltsgröße	-0.079	(0.057)
Erwerbsstat.: Ausbildung	0.343*	(0.174)
Arbeitslos	0.462	(0.547)
Beschäftigt	0.381*	(0.186)
sonstiges	0.791**	(0.203)
besch.*arbeitsl.	0.238	(0.593)
Jahre seit ersten Job	0.006	(0.045)
Bildungsabschl.: Haupt-, Realsch.	0.305	(0.192)
Abitur	0.552*	(0.238)
sonstiges	0.193	(0.629)
Eigenes Einkommen	0.021**	(0.008)
Arbeitslosengeld/-hilfe	0.036	(0.072)
Einkommen Eltern	0.007†	(0.004)
Konstante	-4.503**	(0.373)
Mit Partner		
Vater: min. Realschule	-0.223	(0.222)
Mutter: min. Realschule	0.290	(0.233)
Alter 19-20	0.605**	(0.203)
21-22	0.876**	(0.239)
23-24	0.905**	(0.297)
> 24	0.711†	(0.375)
Jahrg. 70-73	-0.100	(0.156)
74-76	-0.160	(0.196)
77-79	-0.101	(0.252)
Haushaltsgröße	0.022	(0.050)
Erwerbsstat.: Ausbildung	0.036	(0.170)
Arbeitslos	0.956*	(0.422)
Beschäftigt	1.050**	(0.185)
sonstiges	1.366**	(0.166)
besch.*arbeitsl.	-0.330	(0.467)
Jahre seit ersten Job	0.059	(0.041)
Bildungsabschl.: Haupt-, Realsch.	0.330†	(0.191)
Abitur	-0.047	(0.269)
sonstiges	0.349	(0.472)
Eigenes Einkommen	0.012†	(0.007)
Arbeitslosengeld/-hilfe	-0.050	(0.081)
Einkommen Eltern	-0.003	(0.005)
Konstante	-4.622**	(0.367)
N		4198
Log-likelihood		-1816.301
$\chi^2_{(48)}$		500.35
Signifikanzlevel : † : 10%    * : 5%    ** : 1%		

## 7 Schätzergebnisse

Punkte. Wenn Männer von einer Ausbildung in Arbeitslosigkeit wechseln, verringert sich die Wahrscheinlichkeit, dass sie mit einer Partnerin zusammen ziehen. Bei Frauen hingegen erhöht sie sich. Möglicherweise deutet dies darauf hin, dass das Familienmodell, worin der Mann der Haupternährer ist, in der hier untersuchten jungen Generation noch von großer Bedeutung ist. Für den Mann ist Arbeitslosigkeit dann ein Hindernis, mit einer Partnerin einen Haushalt zu gründen. Für die Frau hingegen ist dies kein Hindernis, sofern der Mann als Hauptverdiener fungiert.

In den bisher diskutierten Modellvarianten wurde unbeobachtete Heterogenität zwischen den Individuen nicht berücksichtigt. Wenn es zwischen den Personen unbeobachtete Unterschiede gibt, welche die Hazardrate beeinflussen, können die geschätzten Parameter verzerrt sein (vgl. Kiefer, 1988). Es werden daher noch Modelle mit proportionalen Hazardraten geschätzt, die auf unbeobachtete Heterogenität kontrollieren. Bei diesem Modell werden nicht mehr zwei verschiedene Zielzustände verglichen, sondern es gibt am Ende der Beobachtung nur die Möglichkeit „Auszug“ oder „kein Auszug“.

Unbeobachtete Heterogenität wird berücksichtigt, indem, wie von Meyer (1990) vorgeschlagen, ein nicht beobachteter gammaverteilter Term  $\nu_i$  eingebaut wird. Die Hazardrate lautet in diesem Fall

$$h_{it} = \nu_i \exp \{ \alpha \mathbf{x}_{it} \}. \quad (7.1)$$

Zur Schätzung wird die von Stephen Jenkins für Stata programmierte Prozedur `pgmhaz` verwendet.

Die Tabellen 7.9 und 7.10 zeigen die Ergebnisse der Schätzungen. Tests zeigen, dass hier unbeobachtete Heterogenität vorliegt. Die Unterschiede zwischen den Modellen sind aber sehr gering, so dass die Aussagen aus dem Modellen ohne die Berücksichtigung von unbeobachteter Heterogenität erhalten bleiben.

Tabelle 7.7: Prognose der Auszugswahrscheinlichkeit

	Ausbildung	Beschäftigt	Arbeitslos
Männer			
Bei Eltern	0,9384	0,9380	0,9463
Single	0,0348	0,0355	0,0484
Mit Partnerin	0,0268	0,0265	0,0053
Frauen			
Bei Eltern	0,9126	0,8684	0,8726
Single	0,0572	0,0551	0,0607
Mit Partner	0,0302	0,0765	0,0667

7 Schätzergebnisse

Tabelle 7.8: Veränderung der Auszugswahrscheinlichkeit

	Ausbildung → Arbeitslos	Beschäftigt → Arbeitslos
Männer		
Bei Eltern	0,0079	0,0083
Single	0,0136	0,0129
Mit Partnerin	-0,0215	-0,0212
Frauen		
Bei Eltern	-0,0401	0,0041
Single	0,0036	0,0056
Mit Partner	0,0365	-0,0098

Tabelle 7.9: Modell mit unbeobachteter Heterogenität: Frauen

Variable	ohne unbeob. Heterogenität (1)		mit unbeob. Heterogenität (2)	
	Koeffizient	(Std. Abw. )	Koeffizient	(Std. Abw. )
Vater: min. Realsch.	-0,119	0,139	-0,062	0,161
Mutter: min. Realsch.	0,460	0,143	0,428	0,170
Alter 19-20	0,661	0,136	0,950	0,154
21-22	0,844	0,162	1,197	0,200
23-24	0,891	0,202	1,323	0,280
≥24	0,641	0,256	1,196	0,399
Jahrg. 70-73	-0,149	0,106	-0,124	0,129
74-76	-0,126	0,127	-0,076	0,147
77-79	-0,154	0,164	-0,183	0,187
Haushaltsgröße	-0,021	0,036	-0,020	0,042
Erwerbsstat. Ausbildung	0,138	0,115	0,139	0,124
Arbeitslos	0,688	0,312	0,824	0,331
Erwerbstätig	0,679	0,124	0,779	0,136
anderes	0,997	0,118	1,023	0,148
Arbeitslos*Erwerbst.	-0,085	0,342	-0,166	0,361
Jahre seit Arbeitsbeginn	0,030	0,028	0,060	0,044
Bildungsabschl.: Haupt-, Realsch.	0,305	0,129	0,326	0,153
Abitur	0,265	0,167	0,277	0,190
sonst. Bildungsabschl.	0,289	0,359	0,136	0,459
Eigenes Einkommen	0,015	0,005	0,016	0,005
Arbeitslosengeld/-hilfe	-0,009	0,050	0,000	0,055
Einkommen Eltern	0,002	0,003	0,003	0,003
Konstante	-3,828	0,247	-4,244	0,306
Likelihood-ratio-Statistik für Test Modell (1) vs (2)			161,54012	
Pr(Test-Statistik $\chi^2$ )			0,000	

## 7 Schätzergebnisse

Tabelle 7.10: Modell mit unbeobachteter Heterogenität: Männer

Variable	ohne unbeob. Heterogenität (1)		mit unbeob. Heterogenität (2)	
	Koeffizient	(Std. Abw. )	Koeffizient	(Std. Abw. )
Vater: min. Realsch.	-0,166	0,155	-0,173	0,156
Mutter: min. Realsch.	0,147	0,152	0,192	0,152
Alter 19-20	1,261	0,256	1,822	0,318
21-22	1,899	0,264	2,480	0,325
23-24	2,143	0,284	2,711	0,343
≥24	2,063	0,317	2,600	0,373
Jahrg. 70-73	-0,132	0,117	-0,148	0,119
74-76	-0,310	0,168	-0,347	0,173
77-79	0,418	0,234	0,399	0,246
Haushaltsgröße	0,023	0,036	0,014	0,037
Erwerbsstat. Ausbildung	0,302	0,144	0,329	0,145
Arbeitslos	0,069	0,412	-0,288	0,476
Erwerbstätig	0,373	0,152	0,338	0,154
anderes	0,102	0,157	0,082	0,159
Arbeitslos*Erwerbst.	0,250	0,439	0,562	0,501
Jahre seit Arbeitsbeginn	0,072	0,025	0,076	0,026
Bildungsabschl.: Haupt-, Realsch.	0,447	0,176	0,397	0,178
Abitur	0,545	0,205	0,477	0,206
sonst. Bildungsabschl.	-0,064	0,361	-0,443	0,417
Eigenes Einkommen	0,008	0,003	0,009	0,003
Arbeitslosengeld/-hilfe	0,003	0,031	0,009	0,030
Einkommen Eltern	0,005	0,003	0,005	0,003
Konstante	-5,566	0,340	-6,030	0,390
Likelihood-ratio-Statistik für Test Modell (1) vs (2)			98,459	
Pr(Test-Statistik $\chi^2$ )			0,000	

## 7.2 Semiparametrisches Modell

Um die  $\lambda$ -Werte für die Schätzung des semiparametrischen Modells zu bestimmen wurde wie oben erläutert ein genetischer Algorithmus durchgeführt. Eine Population, auf welche der genetische Algorithmus angewandt wurde, bestand aus 47 Strings. Der Algorithmus wurde abgebrochen, wenn der am besten angepasste String (das fitteste Genom) 15 Generationen lang nicht durch einen besseren Genotyp verdrängt wurde. Bei den Männern war dies nach 40 Runden der Fall, bei den Frauen nach 90 Runden.

In einer ersten Variante der Schätzungen wurden die Variablen, bei denen ein nicht-linearer Zusammenhang denkbar ist, flexibel modelliert: das eigene Einkommen, das Elterneinkommen sowie die Jahre seit dem ersten Job. Einige der  $\lambda$  liefen hierbei gegen unendlich. Aus der Formulierung des Penalisierungsterms in Gleichung 5.17 wird klar, dass für den Fall, dass  $\lambda$  sehr groß wird, der Ausdruck  $(\Delta^{d_{kj}} \gamma_{kjs})^2$  klein werden muss und im Extremfall gegen Null geht. Die Likelihoodfunktion wird ja maximiert, womit der Penalisierungsterm minimiert wird. Die Unterschiede in der Steigung an zwei benachbarten Punkten müssen also Null werden. Für Variablen, deren  $\lambda$  gegen unendlich läuft, ist die beste Anpassung damit eine lineare Funktion.

Bei dem Modell für die Männer liefen die  $\lambda$ -Werte für das Einkommen der Eltern und für die Jahre seit dem ersten Job gegen unendlich, bei dem Modell für die Frauen nur das Einkommen der Eltern. In der Schätzung, deren Ergebnisse hier vorgestellt werden, wurden die Variablen, deren  $\lambda$  gegen unendlich lief, linear modelliert.

Die Ergebnisse der Schätzungen finden sich in den Abbildungen 7.1 bis 7.4 sowie in den Tabellen 7.11 und 7.12.

Ein steigendes Einkommen von jungen Männern, beginnend bei keinem Einkommen, erhöht zunächst die Auszugswahrscheinlichkeit. Bei einem mittleren Einkommen führt zusätzliches Einkommen nicht mehr zu einer höheren Auszugswahrscheinlichkeit. Ist das Einkommen sehr hoch, findet sich wieder ein positiver Zusammenhang zwischen Einkommen und Auszugswahrscheinlichkeit. Der Einfluss des Einkommens auf die Auszugswahrscheinlichkeit ist in der Kategorie ‚mit Partnerin‘ deutlicher als in der Kategorie ‚Single‘. Wenn junge Männer mit einer Partnerin zusammenziehen wollen, spielt die eigene Einkommenssituation also eine wichtige Rolle.

Bei Frauen erhöht sich mit zunehmender Berufserfahrung die Wahrscheinlichkeit, mit einem Partner einen Haushalt zu gründen (Abbildung 7.4). Beim eigenen Einkommen der Frauen steigt im Bereich der unteren Einkommen die Funktion zwar den Erwartungen entsprechend an, jedoch sind die Konfidenzbänder hier sehr breit, so dass eine Interpretation schwierig ist.

## 7 Schätzergebnisse

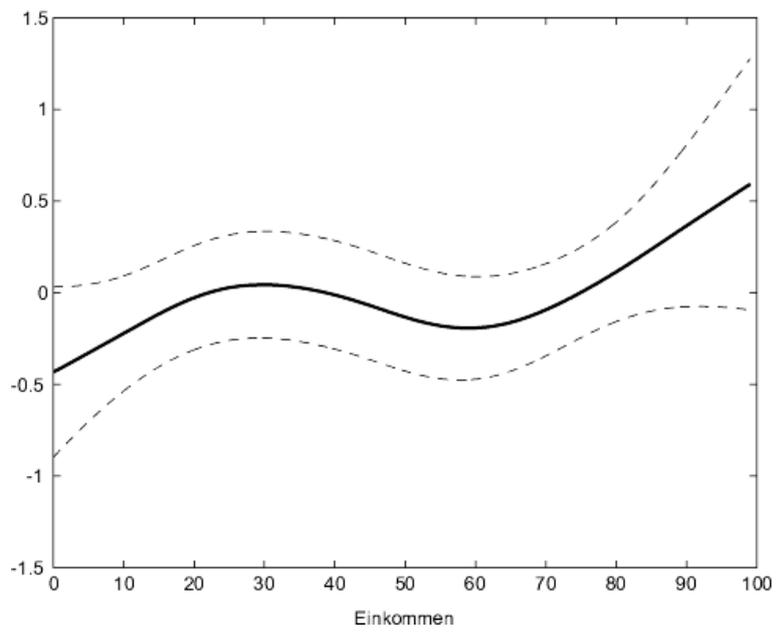


Abbildung 7.1: Geschätzter Effekt - Single, Männer

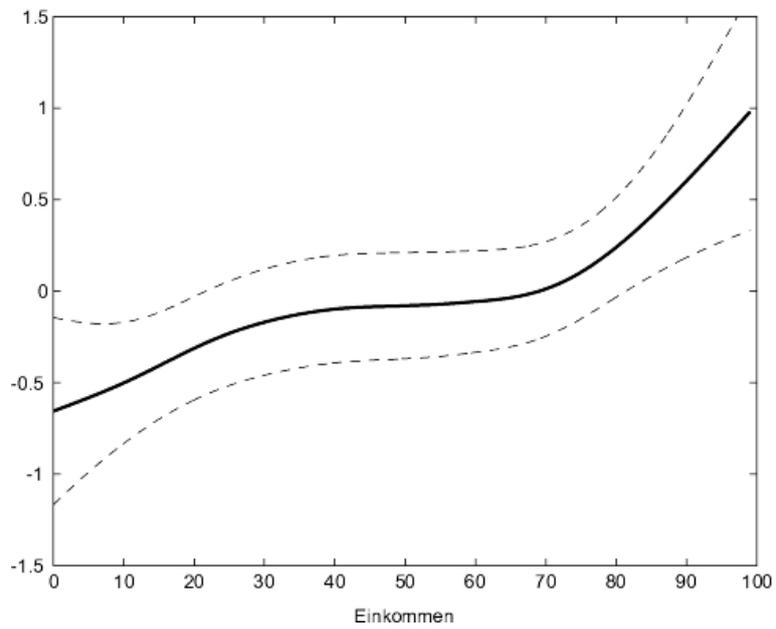


Abbildung 7.2: Geschätzter Effekt - mit Partnerin, Männer

## 7 Schätzergebnisse

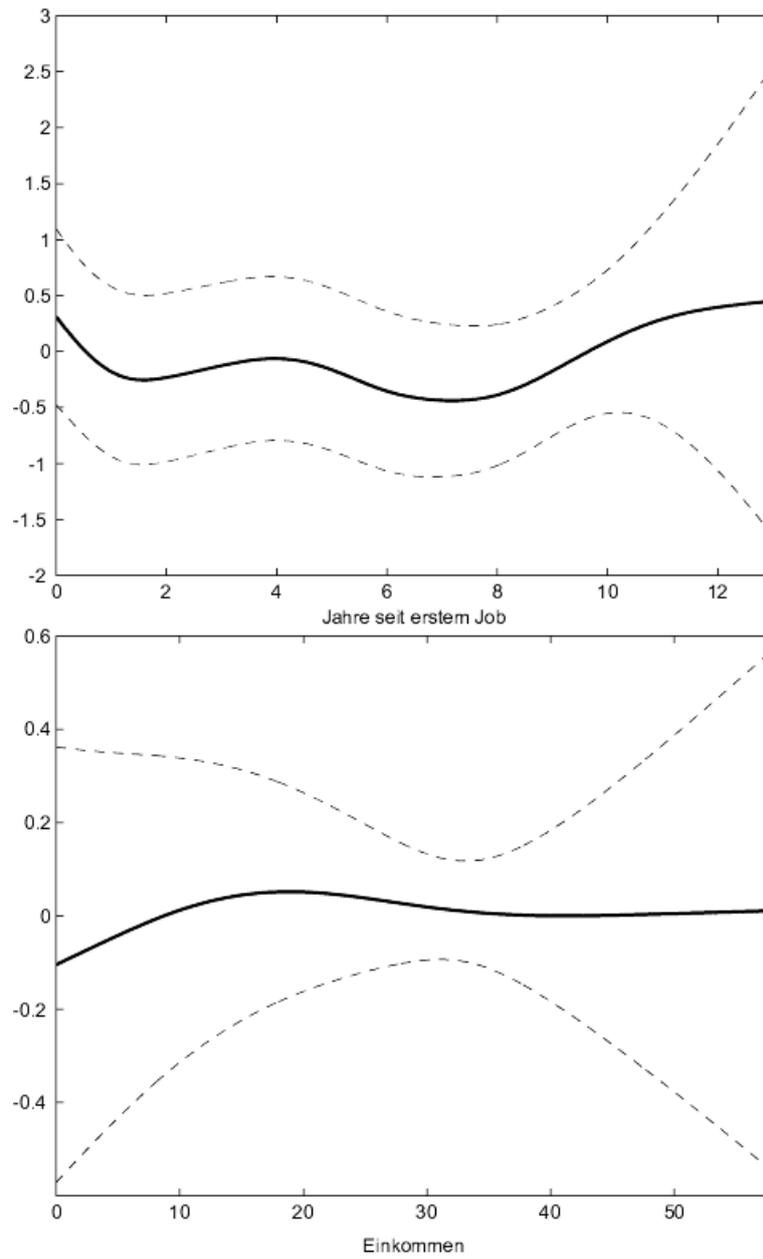


Abbildung 7.3: Geschätzte Effekte - Single, Frauen

## 7 Schätzergebnisse

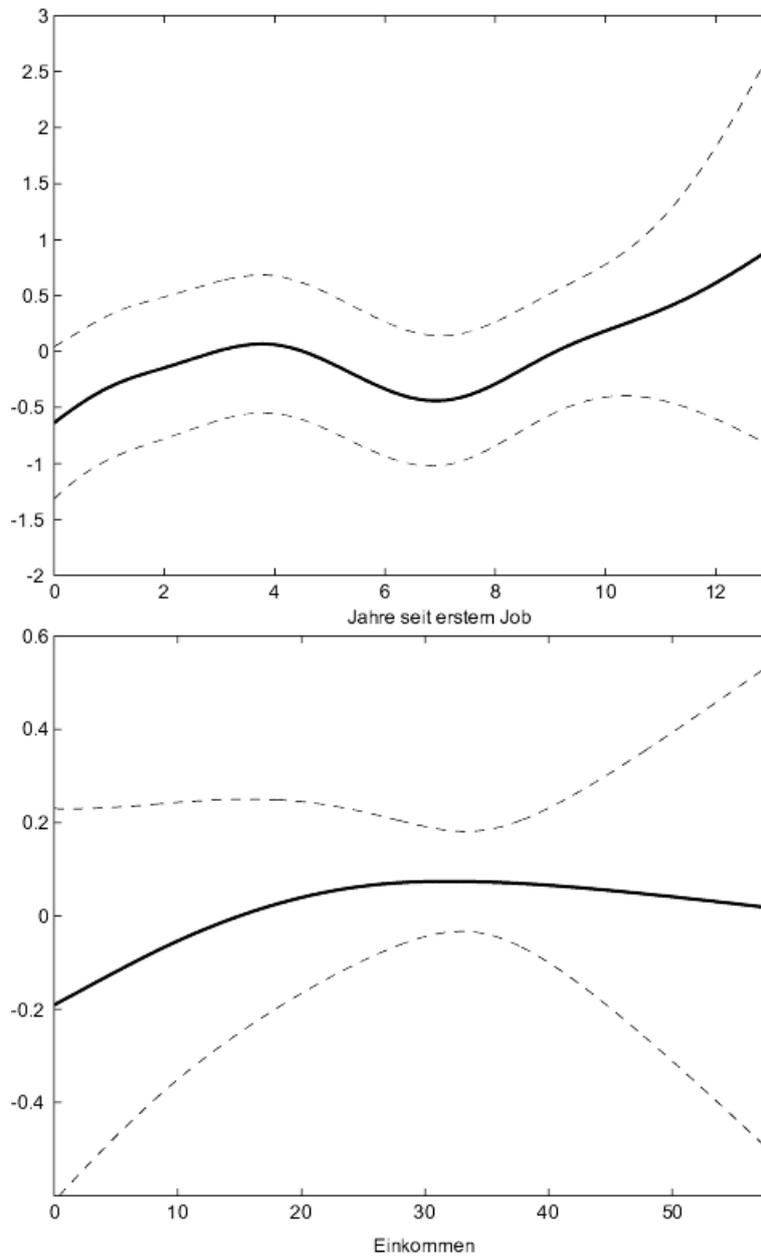


Abbildung 7.4: Geschätzte Effekte - mit Partner, Frauen

Tabelle 7.11: Semiparametrisches Modell - Koeffizienten: Männer

Variable	Koeffizient	(Std. Abw.)
Single		
Vater: min. Realsch.	-0,1475	0,2311
Mutter: min. Realsch.	0,2143	0,2259
Alter j 19	-4,7804	0,5184
19-21	-4,2198	0,4993
22-24	-3,9354	0,5129
j24	-4,2885	0,5842
Jahrg. 70-73	0,1105	0,191
74-76	0,061	0,2542
77-79	0,7888	0,3075
Haushaltsgröße	-0,0785	0,0703
Erwerbsstat. Ausbildung	-0,674	0,2291
Arbeitslos	-0,0004	0,2856
Erwerbstätig	0,9838	0,2787
anderes	0,2078	0,4025
Jahre seit Arbeitsbeginn	0,0921	0,0439
Bildungsabschl.: Haupt-, Realsch.	0,2139	0,257
Abitur	0,5484	0,3055
sonst. Bildungsabschl.	-0,2307	0,6357
Einkommen Eltern	0,0061	0,0057
Mit Partnerin		
Vater: min. Realsch.	-0,2776	0,2732
Mutter: min. Realsch.	0,0673	0,2795
Alter j 19	-6,3406	0,645
19-21	-4,5693	0,5743
22-24	-4,0708	0,5825
j24	-3,9524	0,6325
Jahrg. 70-73	-0,3051	0,199
74-76	-0,4734	0,2995
77-79	-0,7103	0,6154
Haushaltsgröße	0,1031	0,062
Erwerbsstat. Ausbildung	0,3191	0,2388
Arbeitslos	-0,0958	0,3171
Erwerbstätig	0,2797	0,3261
anderes	0,131	0,5249
Jahre seit Arbeitsbeginn	0,0532	0,0442
Bildungsabschl.: Haupt-, Realsch.	0,8569	0,361
Abitur	0,7799	0,4071
sonst. Bildungsabschl.	0,0367	0,8038
Einkommen Eltern	0,0051	0,0066

Tabelle 7.12: Semiparametrisches Modell - Koeffizienten: Frauen

Variable	Koeffizient (Std. Abw.)	
Single		
Vater: min. Realsch.	-0,0357	0,2128
Mutter: min. Realsch.	0,6285	0,213
Alter j 19	-4,1977	0,6004
19-21	-3,13	0,5789
22-24	-3,0589	0,5594
j24	-3,0822	0,5567
Jahrg. 70-73	-0,2561	0,1836
74-76	0,0076	0,2048
77-79	0,0266	0,2463
Haushaltsgröße	-0,1315	0,067
Erwerbsstat. Ausbildung	-0,8296	0,1876
Arbeitslos	0,3142	0,2556
Erwerbstätig	1,2037	0,2283
anderes	0,1557	0,3042
Bildungsabschl.: Haupt-, Realsch.	0,2966	0,2132
Abitur	0,3404	0,2619
sonst. Bildungsabschl.	-0,0902	0,7586
Einkommen Eltern	0,0064	0,0051
Mit Partner		
Vater: min. Realsch.	-0,3038	0,2288
Mutter: min. Realsch.	0,3304	0,2317
Alter j 19	-3,0061	0,5403
19-21	-2,2832	0,5202
22-24	-2,0705	0,4967
j24	-2,5817	0,5155
Jahrg. 70-73	-0,2303	0,1627
74-76	-0,152	0,197
77-79	-0,3023	0,2614
Haushaltsgröße	0,0025	0,0553
Erwerbsstat. Ausbildung	-0,4378	0,18
Arbeitslos	0,684	0,2068
Erwerbstätig	0,6569	0,2256
anderes	0,7931	0,2425
Bildungsabschl.: Haupt-, Realsch.	0,2105	0,1997
Abitur	-0,0962	0,2761
sonst. Bildungsabschl.	-0,2453	0,636
Einkommen Eltern	-0,0063	0,0054

## 8 Schlussfolgerung

In dieser Arbeit wurde untersucht, ob Arbeitslosigkeit einen Einfluss auf das Auszugsverhalten junger Menschen in der Bundesrepublik (West) hat. Hierzu wurden mit Daten des Sozioökonomischen Panels Hazardratenmodelle geschätzt. Als mögliche Zielzustände nach dem Auszug wurden das Wohnen in einem Single-Haushalt und das Wohnen mit einem Partner unterschieden. Es hat sich gezeigt, dass bei Arbeitslosigkeit entgegen der theoretischen Erwartung die Wahrscheinlichkeit des Auszuges aus dem Elternhaus nicht verringert wird. Lediglich die Wahrscheinlichkeit, dass arbeitslose junge Männer mit einer Partnerin zusammenziehen nimmt ab. Die Institution Familie fungiert also für junge Menschen nur in geringem Maße als Versicherung gegen Arbeitslosigkeit. Wichtige Determinante für das Auszugsverhalten ist das eigene Einkommen. Ein höheres Einkommen der Eltern erhöht entgegen der theoretischen Erwartung die Auszugswahrscheinlichkeit. Die Schätzung eines semi-parametrischen Modelles zeigt, dass der Einfluss des Elterneinkommens am besten linear zu modellieren ist. Beim Einfluss des eigenen Einkommens zeigt sich bei den Männern ein ansteigender S-förmiger Verlauf. Insbesondere bei jungen Männern, die mit einer Partnerin einen Haushalt gründen wollen, spielen ökonomische Bedingungen eine Rolle für die Auszugswahrscheinlichkeit. Insgesamt zeigt sich jedoch, dass ökonomische Faktoren wie das eigene Einkommen und das Elterneinkommen eine geringere Bedeutung für die Auszugsentscheidung haben als in angelsächsischen oder süd-europäischen Ländern.

# Literaturverzeichnis

- Aassve, A., Billari, F. C., Mazzuco, S. und Ongaro, F.** (2001). Leaving home ain't easy – a comparative longitudinal analysis of echn data. MPIDR Workingpaper 2001-038.
- Allison, P. D.** (1982). Discrete-time methods for the analysis of event histories. In: S. Leinhardt, Hg., *Sociological Methodology*, Kap. 2, S. 61–98. San Fransisco.
- Avery, R., Goldschneider, F. und Speare, A.** (1992). Feathered nest/gilded cage: Parental income and leaving home in the transition to adulthood. *Demography*, 29(3), 375–388.
- Becker, G. S.** (1981). *A Treatise on the Family*. Harvard University Press, Cambridge.
- Billari, F. C., Philipov, D. und Baizan, P.** (2001). Leaving home in Europe: The experience of cohorts born around 1960. MPIDR Workingpaper 2001-014.
- Blanco, A. und Kluve, J.** (2002). Why not stay home? determinants of nest-leaving behavior in germany and spain. Workingpaper, präsentiert auf der ESPE 2002 in Bilbao, Spanien.
- Börsch-Supan, A.** (1986). Household formation, housing prices, and policy impacts. *Journal of Public Economics*, 30, 146–164.
- Buck, N. und Scott, J.** (1993). She's leaving lome: but why? an analysis of young people leaving the parental home. *Journal of Marriage and the Family*, 55, 863–874.
- Card, D. und Lemieux, T.** (1997). Adapting to circumstances: The evolution of work, school, and living arrangements among north american youth. NBER Working Paper No. W5487.
- Di, Z. X., Yang, Y. und Liu, X.** (2002). Young american adults living in parental homes. Workingpaper des Joint Center for Housing Studies, Harvard University.

## Literaturverzeichnis

- Eilers, P. H. C.** und **Marx, B. D.** (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2), 89–102.
- Ermisch, J.** (1999). Prices, parents, and young people's household formation. *Journal of Urban Economics*, 45, 47–71.
- Ermisch, J.** und **DiSalvo, P.** (1997). The economic determinants of young people's household formation. *Economica*, 64, 627–644.
- Faik, J.** (1997). Institutionelle Äquivalenzskalen als Basis von Verteilungsanalysen. In: *Einkommensverteilung und Armut. Deutschland auf dem Weg zur Vierfünftel-Gesellschaft?* Irene Becker and Richard Hauser.
- Fogli, A.** (2000). Endogenous labour market rigidities and family ties. mimeo. University of Pennsylvania.
- Ghidoni, M.** (2001). Determinants of young Europeans' decision to leave the parental household. Workingpaper präsentiert auf der Royal Economic Society Annual Conference 2002. Mimeo.
- Goldberg, D. E.** (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley Publishing Company, Inc.
- Goldscheider, F., Thornton, A.** und **Young-DeMarco, L.** (1993). A portrait of the nest-leaving process in the early adulthood. *Demography*, 30(4), 683–699.
- Haurin, D. R., Hendershott, P. H.** und **Kim, D.** (1993). The impact of real rents and wages on household formation. *The Review of Economics and Statistics*, 75(2), 284–293.
- Haurin, D. R., Hendershott, P. H.** und **Kim, D.** (1994). Housing decision of american youth. *Journal of Urban Economics*, 35, 28–45.
- Haurin, R. J., Haurin, D. R., Hendershott, P. H.** und **Bourassa, S. C.** (1997). Home or alone: The costs of independent living for youth. *Social Science Research*, 26, 135–152.
- Holland, J.** (1975). *Adaption in natural and artificial systems*. University of Michigan Press.
- Härdle, W.** und **Linton, O.** (1994). Applied nonparametric methods. In: *Handbook of Econometrics*, Bd. 4, Kap. 38. Elsevier Science, Amsterdam,.

- Hu, W.-Y.** (2000). Welfare and family stability: Do benefits affect when children leave the nest? Discussionpaper, Department of Economics, University of California, Los Angeles.
- Hullen, G.** (1995). Der Auszug aus dem Elternhaus im Vergleich von West- und Ostdeutschland. *Zeitschrift für Bevölkerungswissenschaft*, 20(2), 141–158.
- Iacovou, M.** (1998). Young people in Europe: Two models of household formation. Working Papers of the ESRC Research Centre on Micro-social Change, Paper 98-13. Colchester, University of Essex.
- Iacovou, M.** und **Berthoud, R.** (2001). Young people's lives: a map of Europe. Report des Institute for Social and Economic Research.
- Jenkins, S. P.** (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1), 129–137.
- Kiefer, N. M.** (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, 26(2), 646–679.
- Klasen, S.** und **Woolard, I.** (1999). Surviving unemployment without state support: Unemployment and household formation in South Africa. SFB 386, Ludwig-Maximilians-Universität München, Discussion Paper 213.
- Mayer, K. U.** und **Wagner, M.** (1986). Der Auszug von Kindern aus dem elterlichen Haushalt – ein Erklärungsmodell für die Geburtsjahrgänge 1929-31, 1939-41 und 1994-51. In: K. F. Zimmermann, Hg., *Demographische Probleme der Haushaltsökonomie*, Bd. IV von *Contributions to Quantitative Economics*. Brockmeyer, Bochum.
- McElroy, M. B.** (1985). The joint determination of household membership and market work: The case of young man. *Journal of Labour Economics*, 3(3), 293–315.
- Meyer, B. D.** (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4), 757–782.
- Murphey, M.** und **Wang, D.** (1998). Family and sociodemographic influences on patterns of leaving home in postwar Britain. *Demography*, 35(3), 293–305.
- Nelder, J. A.** und **Wedderburn, R. W. M.** (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3), 370–384.

- Ott, N.** (1986). Ausscheiden erwachsener Kinder aus dem elterlichen Haushalt. In: K. F. Zimmermann, Hg., *Demographische Probleme der Haushaltsökonomie*, Bd. IV von *Contributions to Quantitative Economics*. Brockmeyer, Bochum.
- Richards, T., Withe, M. J. und Tsui, A. O.** (1987). Changing living arrangements: A hazard model of transitions among household types. *Demography*, 24(1), 77–97.
- Ronning, G.** (1991). *Mikroökometrie*. Springer-Verlag, Berlin.
- Rosenzweig, M. R. und Wolpin, K. I.** (1993). Intergenerational support and the life-cycle incomes of young men and their parents: Human capital investment s, core-sidence, and intergenerational financial transfers. *Journal of Labor Economics*, 11(1), 84–112.
- Schwarz, K.** (1989). Wann verlassen die Kinder das Elternhaus? *Zeitschrift für Bevölkerungswissenschaft*, 15(1), 39–58.
- Tutz, G. und Scholz, T.** (2000). Semiparametric modelling of multicategorical data. SFB 386, Ludwig-Maximilians-Universität München, Discussion Paper 209.
- Wagner, M. und Huinink, J.** (1991). Neuere Trends beim Auszug aus dem Elternhaus. *Acta Demographica*, S. 39–62.
- Wall, R.** (1989). Leaving home and living alone: An historical perspective. *Population Studies*, 43(3), 369–389.
- Weick, S.** (1993). Determinanten des Auszugs aus der elterlichen Wohnung. In: A. Dieckmann und S. Weick, Hg., *Der Familienzyklus als sozialer Prozess*. Drucker & Humbolt.
- White, L.** (1994). Coresidence and leaving home: Young adults and their parents. *Annual Review of Sociology*, 20, 81–102.
- Wolff, J.** (1998). *Essays in Unemployment Duration in two Economies in Transition: East Germany and Hungary*. Dissertation, European University Institute, Department of Economics, Florenz.
- Yi, Z., Coale, A., Choe, M. K., Zhiwu, L. und Li, L.** (1994). Leaving the parental home: Census-based estimation for china, japan, south korea, united states, france, and sweden. *Population Studies*, 48, 65–80.

*Literaturverzeichnis*

**Ziegler, R.** und **Schladt, D.** (1993). Auszug aus dem Elternhaus und Hausstandsgründung. In: A. Dieckmann und S. Weick, Hg., *Der Familienzyklus als sozialer Prozess*. Drucker & Humbolt.

## **Teil II**

# **Wann verlassen junge Menschen in Italien das Elternhaus?**

## **Zusammenfassung**

Ziel der Untersuchung ist, den Einfluss ökonomischer Faktoren auf die Wahrscheinlichkeit zu bestimmen, mit der junge Menschen in Italien das Elternhaus verlassen. Hierzu werden mit den Wellen von 1991 bis 1998 aus dem Householdsurvey der Banca d' Italia mehrere Logit-Modelle geschätzt. Es zeigt sich, dass mit der Beendigung der Schulausbildung die Auszugswahrscheinlichkeit zunimmt. Arbeitslose haben gegenüber Beschäftigten eine geringere Auszugswahrscheinlichkeit. Diese ist auf deren geringeres Einkommen zurückzuführen. Ein höheres Einkommen des Vaters erhöht die Auszugswahrscheinlichkeit. Ökonomische Faktoren haben in Italien somit einen hohen Einfluss auf das Auszugsverhalten junger Menschen. Die Bedeutung ökonomischer Größen deutet auf die Versicherungsfunktion hin, die familiäre Netzwerke in Italien einnehmen.

JEL-Klassifikation: D19, C25

## 9 Einführung

Rigiditäten am Arbeitsmarkt, welche durch sozialstaatliche Einrichtungen oder gesetzliche Regulierungen entstehen, sind seit längerer Zeit Gegenstand intensiver Forschung (vergleiche hierzu Layard et al., 1991). In jüngster Zeit wird jedoch zunehmend die Rolle der Institution Familie für den Arbeitsmarkt erkannt. So entstanden in den vergangenen Jahren eine Reihe von Arbeiten, welche die Wechselwirkungen zwischen dem Arbeitsmarkt, sozialstaatlichen Einrichtungen und dem Haushaltsformationsverhalten analysieren.

Von Interesse ist hier ein Vergleich zwischen verschiedenen Ländern Europas. Untersucht man, wann junge Menschen das Elternhaus verlassen, finden sich im Europäischen Raum zwei verschiedene Muster des Auszugsverhaltens (vgl. Iacovou, 1998). In skandinavischen Ländern, in denen der Sozialstaat weiter ausgebaut ist, verlassen junge Menschen bereits früh (im Durchschnitt mit 20 bis 21 Jahren) das Elternhaus. In den Mittelmeerländern, in denen häufiger familiäre Bindungen die Absicherung vor Risiken wie Arbeitslosigkeit übernehmen, bleiben Heranwachsende länger bei den Eltern. Dort ist das durchschnittliche Auszugsalter etwa 24 Jahre. Länder wie Deutschland oder Frankreich sind dazwischen anzusiedeln (siehe Tabelle 9.1).

Es lässt sich auch beobachten, dass in den Ländern im Mittelmeerraum die Arbeitslosigkeit unter Jugendlichen höher ist, als in anderen Europäischen Ländern (siehe Tabelle 9.2). Die Unterschiede in der Arbeitslosigkeit lassen sich nur zum Teil durch Faktoren wie das Steuersystem oder Lohnersatzleistungen erklären (Bean, 1994; Blanchard und Wolfers, 1999; Wenzel, 1995).

Die Lohnersatzleistungen betragen 1997 in den südeuropäischen Ländern im Durchschnitt 63%, in den skandinavischen Ländern hingegen 80%. In Kontinentaleuropa betragen sie 77% (OECD, 1999). In Italien, das in der vorliegenden Arbeit untersucht wird, erhalten Jugendliche, die einen Ausbildungsplatz haben und arbeitslos werden, keine Arbeitslosenunterstützung. In den südlichen Ländern ist die Arbeitslosenunterstützung also eher gering, es gibt dort dennoch eine hohe Arbeitslosigkeit.

Auch andere Formen der öffentlichen Unterstützung für Jugendliche, die sich ungünstigen Arbeitsmarktbedingungen gegenüber sehen, sind in den Mittelmeer-Ländern weniger

## 9 Einführung

ausgebaut. Tabelle 9.3 zeigt, dass in Südeuropa trotz hoher Jugendarbeitslosigkeit die Ausgaben für Arbeitsmarktprogramme sehr gering sind.

Zu vermuten ist, dass in den süd-europäischen Ländern in höherem Maße familiäre Strukturen als Versicherungsinstanz dienen. Dies kann aber auch ein Erklärungsfaktor für die hohe Jugendarbeitslosigkeit sein, da die Versicherung, die familiäre Netzwerke zur Verfügung stellen, zu ineffizienten Arbeitsmarktallokationen führt. Wenn aufgrund von Arbeitslosigkeit junge Menschen länger im Elternhaus wohnen bleiben, sind sie damit weniger mobil. Sie ziehen also nicht in Regionen, in welchen die Arbeitsnachfrage höher ist.

Tabelle 9.1: Durchschnittliches Auszugsalter 1996

Land	Durchschnitt	
	Männer	Frauen
Griechenland	24,5	22,8
Italien	24,3	23,5
Spanien	24,1	23,6
Deutschland	22,8	21,6
Frankreich	22,5	21,7
Dänemark	20,8	20,2
Finnland	21,8	20,2

*Quelle: Ghidoni (2001)*

Tabelle 9.2: Jugendarbeitslosigkeit

Land	Arbeitslosenquote 14-25jährige		
Jahr	1990	1997	2000
Griechenland	23,3	31,0	29,5
Italien	28,9	33,6	31,5
Spanien	30,1	37,1	25,5
Deutschland	5,6	10,2	7,7
Frankreich	19,1	28,1	20,7
Dänemark	11,5	8,1	6,7
Finnland	9,1	25,3	21,7

*Quelle: OECD (2001)*

Tabelle 9.3: Arbeitsmarktprogramme für Jugendliche

<b>Land</b>	<b>Ausgaben</b>
Griechenland	0,02
Italien	0,01
Spanien	0,06
Deutschland	0,06
Frankreich	0,09
Dänemark	0,10
Finland	0,11

*Ausgaben für öffentliche Maßnahmen für arbeitslose und benachteiligte Jugendliche in % des BIP im Jahre 1997; Quelle: OECD (2001)*

Klasen und Woolard (1999) haben dieses Problem für Südafrika analysiert. Dort ziehen Menschen, die in Städten leben und arbeitslos werden, häufig zu ihren Eltern, die oft in ländliche Regionen leben, zurück. Sie passen sich an Arbeitslosigkeit also an, indem sie gerade dorthin ziehen, wo weniger Arbeitsplätze angeboten werden.

Das Phänomen des späten Auszugs aus dem Elternhaus ist auch aus entwicklungspsychologischer Sicht von Interesse. In der Literatur wird die Zunahme an Freiheit und Eigenständigkeit betont, die mit dem Auszug aus dem Elternhaus verbunden ist. Findet der Auszug erst sehr spät statt, könne das für die Entwicklung der Persönlichkeit ein Nachteil sein (Papastefanou, 2000). Der Psychologe M. Cicogna kommt zu dem Ergebnis, dass der späte Auszug aus dem Elternhaus die Emanzipation behindere und ein Karrierehindernis sei (BdW, 2001). Für den späten Auszug junger italienischer Männer macht er vor allem die Ängstlichkeit deren Mütter um ihre Söhne verantwortlich.

Aus ökonomischer Sicht gehören neben den Präferenzen der Eltern auch die Präferenzen der Kinder und deren Budgetrestriktionen zu den Bestimmungsfaktoren der Auszugswahrscheinlichkeit. In dieser Arbeit ist der Fokus auf die Budgetrestriktionen gerichtet. Es soll untersucht werden, ob der Erwerbsstatus und das Einkommen das Auszugsverhalten junger Menschen in Italien beeinflusst. Zunächst wird ein Überblick über bisherige Untersuchungen zum Thema gegeben. Im darauf folgenden Abschnitt wird ein theoretisches Modell der Auszugsentscheidung dargestellt. Im Abschnitt 12 wird der verwendete Datensatz beschrieben. Im Abschnitt 13 werden die Ergebnisse des ökonometrischen Modells diskutiert.

## 10 Bisherige Untersuchungen

In den vergangenen Jahren sind eine Reihe von Arbeiten zu den ökonomischen Einflussfaktoren auf das Haushaltsformationsverhalten entstanden.

Fogli (2000) untersucht die Unterschiede der Arbeitsmarktinstitutionen und Familienstrukturen zwischen verschiedenen Ländern. Im Rahmen eines overlapping generation models versucht sie, Familienstrukturen als endogene Antwort auf Kreditmarktbeschränkungen zu erklären. Sehen sich junge Menschen Kreditmarktrestriktionen gegenüber, versuchen sie ihren Konsum zu glätten, indem sie länger bei den Eltern bleiben. Dadurch werden Arbeitsmarktregulierungen wie Kündigungsschutz attraktiver, die zu einer Umverteilung der Arbeitsplätze zur alten Generation führen. Im empirischen Teil arbeitet sie mit Makro-Daten aus den EU-Ländern, Kanada und den USA. In Ländern mit stärkeren Kreditmarktbeschränkungen (dies sind die Mittelmeer-Länder) herrscht ein höherer Arbeitsplatzschutz. Die Arbeitslosigkeit ist dort zur jüngeren Generation hin verteilt und junge Menschen leben länger bei ihren Eltern.

Bentolila und Ichino (2000) untersuchen die Rolle von Arbeitslosigkeit und familiären Bindungen aus einer anderen Perspektive. Sie befassen sich mit der Frage, wie sich das Konsumverhalten aufgrund von Arbeitslosigkeit in den Ländern USA, Spanien, Italien, Deutschland und England ändert. Sie arbeiten mit Panel-Daten aus den jeweiligen Ländern und argumentieren mit deskriptiven Statistiken. Auch sie stellen fest, dass in Südeuropa die Kreditmarktbeschränkungen am stärksten sind. Der Kreditmarkt kann also dort in geringerem Maße zur Konsumglättung genutzt werden als in anderen Ländern. Da es in den südeuropäischen Ländern geringere Sozialtransfers gibt, fallen auch diese zur Konsumglättung weg. In Südeuropa spielen Familiennetzwerke eine größere Rolle als Versicherung gegen Arbeitsmarktrisiken. Der Konsum muss bei Arbeitslosigkeit nicht um den ganzen Einkommensausfall eingeschränkt werden, weil auf Ressourcen der Familienmitglieder zurückgegriffen werden kann.

Ghidoni (2001) analysiert das Auszugsverhalten junger Menschen in Europa. Er verwendet hierzu das Haushaltspanel der Europäischen Gemeinschaft. Als erklärende Variablen verwendet er das Einkommen der Eltern, das eigene Einkommen und den Erwerbsstatus in einem Random-Effects-Probit-Modell. Im Vergleich der EU-Länder stellt

er fest, dass das aktuelle Einkommen und der Erwerbsstatus in den südlichen Ländern einen stärkeren Einfluss auf die Auszugswahrscheinlichkeit haben als in den nördlichen Ländern. Das Haushaltsformationsverhalten wird also in den Mittelmeerländern stärker von Arbeitsmarktereignissen bestimmt, als in den nördlichen Ländern Europas. Zu dem selben Ergebnis kommen Aassve et al. (2001), die mit dem gleichen Datensatz arbeiten.

Cantó und Mercader-Prats (2001) untersucht für Spanien die Auswirkungen von Arbeitslosigkeit auf junge Menschen. Sie verwenden Daten des Spanischen Haushalts-Panel ‚Encuesta Continua de Presupuestos Familiares‘ (ECPF) für die Jahre 1985-1991. Das Auszugsverhalten analysieren sie mit deskriptiven Statistiken. Trotz hoher Jugendarbeitslosigkeit sind junge Menschen weniger von Armut betroffen, als andere Altersgruppen, weil Eltern und Geschwister als Sicherheitsnetz fungieren.

Auch speziell für Italien gibt es bereits mehrere Untersuchungen, welche die Wechselwirkung von familiären Bindungen und ökonomischen Einflussgrößen untersuchen. Ichino et al. (2002) versuchen den späten Auszug junger Menschen in Italien durch die Job-Unsicherheit des Vaters zu erklären. Sie verwenden ein Probit-Modell auf Basis von Daten des Householdsurveys der Banca d'Italia. Sie stellen fest, dass Arbeitslosigkeit des Vaters bzw. eine höhere Wahrscheinlichkeit, dass dieser arbeitslos wird, die Wahrscheinlichkeit bei den Eltern zu bleiben verringert. Sie erklären das Ergebnis damit, dass die Kinder bei Arbeitslosigkeit des Vaters nicht mehr auf dessen Einkommen zur Glättung ihres eigenen Konsumes zurückgreifen können.

Giannelli und Monfardini (2000) untersuchen mit den gleichen Daten die simultane Entscheidung über Humankapitalakkumulation und Auszug aus dem Elternhaus. Sie nutzen jedoch nicht den Panelcharakter des Datensatzes, sondern schätzen ein statisches Logitmodell für die Situation im Jahre 1995. Um den Einfluss des erwarteten Lohnes zu quantifizieren, schätzen sie eine Lohngleichung nach der Heckman-Methode. Ihr Ergebnis ist, dass das erwartete Lebenseinkommen einen starken Einfluss auf die Haushaltsformationsentscheidung und die Humankapitalinvestitionsentscheidung hat. Wenn junge Menschen eine schlechte Berufsperspektive aufgrund regional hoher Arbeitslosigkeit haben, reagieren sie darauf, indem sie länger bei den Eltern bleiben und nach der Schule ein Studium anschließen.

Manacorda und Moretti (2001) arbeiten ebenfalls mit den Daten der Banca d'Italia. Sie gehen von einem non-cooperating bargaining Modell aus, in welchem sie unterstellen, dass das Zusammenleben der Kinder mit den Eltern für die Eltern ein Gut und für die Kinder ein „Schlecht“ ist. Sie untersuchen die Auswirkung einer Erhöhung der Sozialtransfers an die Eltern. Höhere Transfers erhöhen die Verhandlungsmacht der Eltern und

## 10 Bisherige Untersuchungen

sollten damit die Auszugswahrscheinlichkeit der Kinder verringern. Die Autoren kommen zu dem Ergebnis, dass eine Erhöhung des Jahres-Einkommens der Eltern um 1 Mio. Lire die Bleibewahrscheinlichkeit um 1%-5% erhöht. Sie verwenden in ihrer ökonometrischen Analyse ein OLS-Modell mit nach Jahrgang und Alter aggregierten Daten.

In den genannten Untersuchungen wird also festgestellt, dass in den Südeuropäischen Ländern familiäre Netzwerke eine wichtige Funktion als Versicherung bei Einkommensausfall haben.

# 11 Modell

Ökonomische Modelle der Haushaltsformation gehen zurück auf Becker (1981). Er hat das Heirats-, Scheidungs- und Fertilitätsverhalten analysiert. Einen allgemeineren Ansatz zur Analyse des Haushaltsformationsverhaltens diskutiert Ermisch (1986).

Die erste formale Modellierung der Auszugsentscheidung stammt von McElroy (1985). Sie analysiert im Rahmen eines Nash (1953)-Verhandlungsmodelles die simultane Entscheidung über den Arbeitsmarktstatus und den Auszug aus dem Elternhaus. Das Entscheidungsverfahren hat zwei Stufen. Zunächst optimiert das Kind seinen Nutzen über die Wahl der Freizeit und Gütermenge für den Fall, dass es in einem eigenen Haushalt lebt. Dieses Optimum wird dann verglichen mit dem Fall, dass das Kind gemeinsam mit den Eltern im Haushalt lebt. Im zweiten Schritt wird dann über den Haushaltsstatus entschieden. Das Nutzenniveau, wenn das Kind allein lebt, bestimmt den Drohpunkt im Nash-Verhandlungsmodell. Ist der Drohpunkt unterhalb der Nutzenmöglichkeitenkurve des Haushaltes, bleibt das Kind bei den Eltern. Die Entscheidung über das Konsumniveau und das Arbeitsangebot wird dann simultan von den Haushaltsmitgliedern in einer Nash-Verhandlung getroffen (vgl. McElroy und Horney, 1981). Der Vorteil des Zusammenlebens entsteht, weil es innerhalb des Haushaltes öffentliche Güter gibt (gemeinsam genutzter Wohnraum, Fernseher, Waschmaschine usw.), die gemeinsam genutzt werden. Ein Ergebnis des Modells ist, dass Eltern ihren Kindern ein Mindest-Nutzenniveau zusichern, wenn die Kinder mit schlechten Arbeitsmarktbedingungen konfrontiert sind. Eine weniger formale Diskussion eines Nash-Verhandlungs-Modelles zur Analyse der Auszugsentscheidung findet sich in Ott (1986). Sie berücksichtigt in ihrer Argumentation auch psychologische Faktoren, wie den Wunsch der Kinder nach Privatsphäre oder den Wunsch nach sozialen Kontakten innerhalb der Familie.

Während McElroy ein kooperatives Spiel zugrunde legt, gehen Rosenzweig und Wolpin (1993, 1994) von einem zweistufigen nicht-kooperativen Spiel aus. Sie nehmen an, dass die Eltern altruistisch sind und Ressourcen an ihre Kinder transferieren wollen. Da Haushaltsgüter ein öffentliches Gut im Haushalt sind, ist es für Eltern günstiger, die Transfers über diese öffentlichen Güter in einem gemeinsamen Haushalt zu leisten, als über monetäre Transfers in dem Fall, dass die Kinder nicht mehr im elterlichen Haus-

halt leben. Eltern haben somit einen Anreiz, dass die Kinder länger im Haushalt verweilen. Eltern beeinflussen die Entscheidung der Kinder, indem sie die Transfers entsprechend ausgestalten. Rosenzweig und Wolpin (1993, 1994) formulieren für ihren Ansatz ein Modell mit sich überlappenden Generationen, in welchem auch die Humankapital-Investitionsentscheidung der jungen Generation berücksichtigt wird.

Ermisch und DiSalvo (1997) berücksichtigen wie Rosenzweig und Wolpin (1993) ebenfalls die Möglichkeiten von Transfers, verwenden jedoch ein statisches Modell. Sie erweitern aber den Ansatz, indem sie explizit die Nachfrage nach Wohnraum modellieren. Manacorda und Moretti (2001, 2002) verwenden einen ähnlichen Ansatz wie Ermisch und DiSalvo (1997). Sie gehen in ihrem Modell davon aus, dass es für die Eltern ein Gut ist, wenn die Kinder im Haushalt leben. Der Nutzen der Eltern erhöht sich dadurch also. Für die Kinder ist es hingegen ein Gut, wenn sie alleine leben. Die Bereitschaft der Eltern, die Kinder länger im Haushalt zu halten, ist also bei ihnen nicht altruistisch motiviert.

Das im Folgenden dargestellte Modell orientiert sich an Ermisch und DiSalvo (1997). Die Nutzenfunktion der Kindes lautet

$$U^c(C_c, h_c, d_{aus}), \quad (11.1)$$

welche maximiert wird unter der Nebenbedingung

$$y_c + T(d_{aus}) = C_c + d_{aus}h_c. \quad (11.2)$$

$h_c$  ist der Konsum von Gütern, die innerhalb eines Haushaltes ein öffentliches Gut sind. Solange das Kind bei den Eltern lebt, muss es hierfür nicht aufkommen. Die öffentlichen Güter werden von den Eltern finanziert. Hierzu zählen insbesondere die Kosten des Wohnens.  $y_c$  ist das Einkommen des Kindes,  $C_c$  ist der sonstige Konsum des Kindes.  $d_{aus}$  ist eine Dummyvariable, die den Wert 1 bei einem Auszug annimmt und sonst auf 0 gesetzt ist.  $T(d_{aus})$  ist der finanzielle Transfer der Eltern an die Kinder. Er hängt davon ab, ob die Kinder bei den Eltern leben. Wenn Kinder noch bei ihren Eltern leben, sind Transaktionskosten geringer, so dass zu erwarten ist, dass die Transfers dann höher sind. Es ist auch zu vermuten, dass Eltern bereit sind höhere Transfers zu zahlen, solange die Kinder in der unmittelbaren Umgebung der Eltern sind.

Es wird davon ausgegangen, dass die Eltern gegenüber ihren Kindern altruistisch sind, was sich mit Modellen aus der Evolutionstheorie begründen lässt (hierzu Bergstorm, 1997). Der Nutzen der Eltern hängt also auch vom Nutzen der Kinder ab. Die Nutzenfunktion der Eltern lautet

$$U^p = U^p((c_p, h_p, d_{aus}, U^c(C_c, h_c, d_{aus}))). \quad (11.3)$$

Deren Nebenbedingung ist

$$y_p = C_p + h_p + T(d_{aus}). \quad (11.4)$$

$y_p$  ist das Einkommen,  $C_p$  und  $h_p$  sind der Konsum der Eltern. Es findet ein zweistufiges Spiel statt. In der ersten Stufe entscheiden die Eltern über ihren Konsum und den Transfer an die Kinder. In der zweiten Stufe entscheiden die Kinder über ihren Konsum, und ob sie bei den Eltern bleiben. Der indirekte Nutzen des Kindes zum Zeitpunkt  $t$ , wenn es noch bei den Eltern lebt, lautet

$$V_t^{in} = V_t^{in}(y_{ct}, y_{pt}, \mathbf{x}_t, u^{in}), \quad (11.5)$$

wenn es ausgezogen ist

$$V_t^{aus} = V_t^{aus}(y_{ct}, y_{pt}, \mathbf{x}_t, u^{aus}), \quad (11.6)$$

$\mathbf{x}_t$  ist ein Vektor aus Charakteristika, welche die Präferenzen bestimmen (z.B. Alter, Geschlecht).  $u^i$  (mit  $i = in, aus$ ) ist eine stochastische Nutzenkomponente. Der Auszug findet statt, wenn

$$V_t^{in} < V_t^{aus}. \quad (11.7)$$

Die Nutzenfunktion wird wie üblich (vgl. Ermisch, 1999) linear spezifiziert:

$$V_t^i = \beta_{0i} + \beta_{1i}y_{ct} + \beta_{2i}y_{pt} + \beta_{4i}\mathbf{x}_t + u_t^i. \quad (11.8)$$

Die Bedingung für den Auszug lautet damit

$$\alpha_0 + \alpha_1y_{ct} + \alpha_2y_{pt} + \alpha_3\mathbf{x}_t + \epsilon_t > 0, \quad (11.9)$$

wobei gilt  $\alpha_j = \beta_{j,aus} - \beta_{j,in}$  und  $\epsilon_t = u_t^{in} - u_t^{aus}$ . Die Koeffizienten der Indexfunktion

$$I = \alpha_0 + \alpha_1y_{ct} + \alpha_2y_{pt} + \alpha_3\mathbf{x}_t + \epsilon_t \quad (11.10)$$

lassen sich, sofern  $\epsilon_t$  logistisch verteilt ist, mit einem Logit-Modell schätzen.

Die Wahrscheinlichkeit, dass ein Kind das Elternhaus verlässt, ist dann

$$p(y_{ct}, y_{pt}, \mathbf{x}_t) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1y_{ct} + \alpha_2y_{pt} + \alpha_3\mathbf{x}_t)}}. \quad (11.11)$$

Manacorda und Moretti (2001, 2002) formulieren ihr Modell mit einer Stone-Geary-Nutzenfunktion. Ermisch und DiSalvo (1997) legen eine CES-Nutzenfunktion zugrunde. Beide Varianten kommen zu dem Ergebnis, dass ein höheres Einkommen des Kindes die Auszugswahrscheinlichkeit erhöht. Ein höheres Einkommen der Eltern verringert (mit Ausnahme unwahrscheinlicher Parameterkonstellationen) die Auszugswahrscheinlichkeit.

## 12 Daten

In dieser Untersuchung werden Daten des Householdsurveys der Banca d'Italia verwendet. Ein Nachteil des Panels ist, dass die Fluktuation zwischen den Wellen sehr groß ist. So wurden zum Beispiel 1995 23.924 Personen befragt, in der darauffolgenden Welle 1998 wurden 20.901 Personen befragt. Von diesen waren nur 7.772 bereits in der vorhergehenden Welle. Die Daten des Panels wurden meist alle zwei Jahre erhoben. Zwischen den beiden neuesten Wellen liegt jedoch ein Zeitraum von drei Jahren. In dieser Untersuchung werden die Wellen von 1991, 1993, 1995, 1998 verwendet. Mehrere Wellen mit verschiedenen Zeitabständen in einer Untersuchung zu verwenden stößt auf methodische Schwierigkeiten. Es werden hier daher zwei Modelle gebildet. Das eine schätzt die Auszugswahrscheinlichkeit auf Basis der Daten aus den Jahren 1991 und 1993. Das andere wird mit den Daten aus dem Jahre 1995 geschätzt.

Die abhängige Variable lautet somit in einem Fall ‚verlässt in den nächsten zwei Jahren das Elternhaus‘ und im anderen Fall ‚verlässt in den nächsten drei Jahren das Elternhaus‘. Die beiden Modelle lassen sich daher nicht zu einem Modell poolen.

Im Jahre 1991 wurden 24.930 Personen in 8.188 Haushalten befragt, im Jahre 1993 24.013 Personen in 8089 Haushalten. Im Jahre 1991 gab es in den befragten Haushalten 5794 Kinder über 14 Jahren. 1993 waren es 5341. Die in der Welle des Jahres 1995 befragten 23.924 Individuen teilen sich auf 8.135 Haushalte auf. In diesen Haushalten leben 5.364 Kinder, die älter als 14 Jahre sind.

Es wurden für den Datensatz der vorliegenden Untersuchung jene Personen ausgewählt, die im Beobachtungsjahr bei den Eltern lebten, die älter als 14 Jahre waren und deren Elternhaushalt in der darauf folgenden Welle wieder befragt wurde. Dies sind in der Welle von 1991 (1993) 2.593 (2.562) Personen, 1995 waren dies 1.957 Personen. Von 1991 bis 1993 sind 12,11 % der Personen aus dem Elternhaus ausgezogen, von 1993 bis 1995 waren es 8,78 %. Zwischen 1995 und 1998 haben 15,94% der beobachteten Personen den Haushalt der Eltern verlassen. Von nur knapp der Hälfte der ausgezogenen Kinder liegen aus den Befragungen genaue Informationen vor, warum sie den Haushalt verlassen haben. Jene, die aufgrund Wehrdienst, Gefängnis oder Tod das Haus verlassen haben, wurden aus dem Datensatz ausgeschlossen.

Die Befragungen wurde jeweils im Frühjahr durchgeführt und bezogen sich rückblickend auf das jeweilige Vorjahr. Gefragt wurde hierbei auch nach der hier interessierenden Variable: dem vorherrschenden Erwerbsstatus. In dieser Arbeit wird unterschieden zwischen abhängig Beschäftigten, Selbständigen, Arbeitslosen und sonstigen. Zu den Sonstigen zählen vor allem Hausfrauen und Wehr-/Zivildienstleistende.

Die Tabellen 12.1 und 12.2 geben einen deskriptiven Überblick über die verwendeten Variablen. Das Einkommen der betrachteten Person (eigenes Einkommen) und das der Eltern ist das verfügbare Jahres-Netto-Einkommen in 100 Mio. Lire. Zusätzlich wurden die Ausgaben der Eltern für dauerhafte Konsumgüter aufgenommen. Die meisten dieser Ausgaben (Möbel, Haushaltsgeräte usw.) haben innerhalb eines Haushaltes den Charakter eines öffentlichen Gutes.

Die Qualifikation der Kinder wurde aufgenommen, da sie Einfluss auf das permanente Einkommen hat. Entscheidungen, wie die, einen eigenen Haushalt zu gründen, hängen in hohem Maße vom permanenten Einkommen ab, nicht nur vom aktuellen Einkommen. Die Referenzgruppe in der Schätzung setzt sich zusammen aus Personen, die keinen Schulabschluss oder nur den Hauptschulabschluss haben. Für höher qualifizierte wurden Dummyvariablen gebildet. Es wird das Niveau Mittelstufe unterschieden sowie Abitur oder höher.

Tabelle 12.1: Variablenbeschreibung: Daten 1991/1993

Variable	Durchschnitt	Std.-Abw.
Auszug	0,105	0,306
weiblich	0,448	0,497
Alter	22,813	6,681
Einkommen	0,05	0,084
Ausgaben f. dauerh. Konsumgüter	0,052	0,123
Einkommen Vater	0,297	0,27
Einkommen Mutter	0,088	0,122
Erwerbsstatus		
Schule	0,437	0,496
Beschäftigt	0,266	0,442
Selbständig	0,039	0,193
arbeitslos	0,219	0,414
sonstiges	0,04	0,195
Bildungsabschluss		
kein Abschl. od. Haupts.	0,056	0,23
Mittelstufe	0,664	0,472
Abitur	0,279	0,449
Region		
Nord	0,371	0,483
Mitte	0,174	0,379
Süd	0,455	0,498
Jahr: 1993	,499	0,500
N	5155	

*Monetäre Größen in 100 Mio. Lire; die Regionen Mitte umfasst Toskana, Umbrien, Latium und Marken.*

Tabelle 12.2: Variablenbeschreibung: Daten 1995

Variable	Durchschnitt	Std.-Abw.
Auszug	0,159	0,366
weiblich	0,435	0,496
Alter	23,162	6,408
Einkommen	0,058	0,098
Ausgaben f. dauerh. Konsumgüter	0,068	0,185
Einkommen Vater	0,334	0,36
Einkommen Mutter	0,105	0,132
Erwerbsstatus		
Schule	0,420	0,494
Beschäftigt	0,265	0,442
Selbständig	0,061	0,240
arbeitslos	0,211	0,408
sonstiges	0,042	0,202
Bildungsabschluss		
kein Abschl. od Haupts.	0,044	0,205
Mittelstufe	0,458	0,498
Abitur	0,498	0,500
Region		
Nord	0,324	0,468
Mitte	0,176	0,381
Süd	0,500	0,500
N		1957

## 13 Ergebnisse

Wie oben bereits angemerkt, werden Logit-Modelle geschätzt (Gleichung 11.11), um die Determinanten der Auszugsentscheidung zu bestimmen.

Für das Modell der Jahre 1991/93 wurde auch ein Modell mit zufälligen Effekten für die Individuen geschätzt. Die Indexfunktion wird hierbei um einen stochastischen Term  $\nu_i$  ergänzt:

$$I_i = \alpha_0 + \alpha_1 y_{cit} + \alpha_2 y_{pit} + \alpha_3 \mathbf{x}_{it} + \nu_i + \epsilon_{it}. \quad (13.1)$$

Es zeigte sich, dass die Varianz von  $\nu_i$  nicht signifikant von Null verschieden ist. Das Modell kann also auch mit einem einfachen Logit-Ansatz geschätzt werden. Die Personen wurden dennoch zu Cluster zusammengefasst. Dies heißt, dass zwar die Beobachtungen zwischen den Personen unabhängig voneinander sind, aber mehrere Beobachtungen einer Person sind nicht unabhängig voneinander.

Zur Kontrolle regionaler Unterschiede wurden Dummies für die Regionen Nord und Süd aufgenommen. Die Referenzregionen wird durch die Bezirke in der Mitte Italiens (Toskana, Umbrien, Latium und Marken) gebildet.

Zentrale Frage in dieser Untersuchung ist, ob junge Menschen auf eigene Arbeitslosigkeit damit reagieren, dass sie länger bei ihren Eltern wohnen bleiben. Das eigene Einkommen wird in der Variante I (Tabelle 13.1 bis 13.6 und 13.11 bis 13.14) nicht im Modell aufgenommen. In Variante II wird es mit berücksichtigt. In Variante I wurde es ausgeschlossen, da es naturgemäß vom Erwerbsstatus abhängt und damit Endogenitätsprobleme vorliegen können. Zur besseren Interpretation der Ergebnisse folgen den Tabellen mit den geschätzten Koeffizienten noch Tabellen, in denen die marginalen Effekte angegeben sind. Für Dummy-Variablen wurde angegeben, wie sich die Auszugswahrscheinlichkeit ändert, wenn die Variable von null auf eins wechselt.

Zunächst werden die Ergebnisse der Schätzungen mit dem Datensatz der Wellen 1991 und 1993 erläutert. Die erste geschätzte Modellvariante findet sich in Tabelle 13.1. Hier wurde hier ein gemeinsames Modell für Frauen und Männer geschätzt. Der Koeffizient ‚weiblich‘ ist signifikant positiv. Frauen verlassen mit einer größeren Wahrscheinlichkeit das Elternhaus als Männer. Dieses Ergebnis finden sich auch in Studien anderer eu-

ropäischen Länder. Die Dummy-Variable für das Beobachtungsjahr 1993 ( $d_{1993}$ ) ist in allen Schätzungen signifikant negativ. Das Auszugsalter hat also in der ersten Hälfte der 90er Jahre zugenommen. Dies ist ein Trend, der in den meisten westlichen Industrieländern zu finden ist.

Die weiteren Modelle wurden wie in der Literatur üblich für Frauen und Männer getrennt geschätzt. Bei einer Reihe von Kovariablen zeigen sich deutliche Unterschiede zwischen den Geschlechtern. Bei der Schätzung für die Männer (Tabelle 13.3) sind die Erwerbsstatusvariablen ‚Arbeitslos‘, ‚Beschäftigt‘ und ‚Selbständig‘ signifikant positiv (Referenzgruppe sind Schüler). Bei den Frauen (Tabelle 13.5) ist auch die Kategorie ‚sonstiges‘, was meist Hausfrauen sind, signifikant. Auch wenn das durchschnittliche Auszugsalter in Italien höher ist als in anderen Ländern und der Auszug meist nicht an die Beendigung der Schulausbildung gekoppelt ist, hat die Beendigung der Schulausbildung dennoch eine deutliche Erhöhung der Auszugswahrscheinlichkeit zur Folge.

Bei den Männern sind die Koeffizienten ‚Arbeitslos‘ und ‚Beschäftigt‘ mit einer Fehlerwahrscheinlichkeit von 0,04% signifikant voneinander verschieden. Junge arbeitslose Männer bleiben also länger im Elternhaus als erwerbstätige. Dies ist konsistent mit der Annahme, dass familiäre Strukturen eine Versicherungsfunktion bei Arbeitslosigkeit übernehmen. Bei den Frauen kann die Hypothese, dass diese beiden Koeffizienten ‚Arbeitslos‘ und ‚Beschäftigt‘ gleich sind, jedoch nicht abgelehnt werden.

Die Regional-Variable ‚Süd‘ ist in den Schätzungen signifikant positiv bezüglich der Referenzregion Mittelitalien. Bei den Männern ist der Koeffizient ‚Süd‘ auch vom Koeffizient ‚Nord‘ signifikant verschieden. Im Süden verlassen junge Männer also früher die elterliche Wohnung.

Für die Schätzung wurden Altersgruppen gebildet. Die Referenzgruppe bilden die 15 bis 16-jährigen (bzw. die 15-jährigen bei den Schätzungen mit der 1995er Welle). Mit zunehmendem Alter nimmt erwartungsgemäß die Auszugswahrscheinlichkeit zu. In der Altersgruppe der 22-25-jährigen ist der Koeffizient der Frauen erheblich höher als bei Männern. In der Altersgruppe der über 30-jährigen ist dies umgekehrt. Auch hier zeigt sich, dass Frauen in jüngeren Alter das Elternhaus verlassen als Männer. Zurückzuführen ist dies auf das jüngere Heiratsalter der Frauen.

Anders als in Manacorda und Moretti (2001) zeigt sich hier, dass ein höheres Einkommen der Eltern die Auszugswahrscheinlichkeit bei Männern erhöht. Die Ergebnisse sind jedoch nur begrenzt vergleichbar, da Manacorda und Moretti (2001) kein mikroökonomisches Modell schätzen.

Das Einkommen der Mutter hat im Gegensatz zum Einkommen des Vaters ein nega-

tives Vorzeichen. In Modellvarianten, in denen die beiden Einkommen zusammengefasst wurden, hatte das Elterneinkommen ein positives Vorzeichen. Das negative Vorzeichen beim Einkommen der Mutter kann damit erklärt werden, dass ein höheres Einkommen der Mutter ein Indiz dafür ist, dass der Vater kein hohes Einkommen hat und die Mutter deshalb einer Erwerbstätigkeit nachgehen muss. Das Einkommen der Väter ist nur bei den Modellen für die Männern signifikant.

Der eigene Bildungsabschluss spielt keine signifikante Rolle für die Auszugswahrscheinlichkeit. Dies war auch bei Modellspezifikationen der Fall, in denen nur die Bildungsvariablen aufgenommen wurden.

In den weiteren Varianten des Modells (Tabellen 13.7 und folgende) wurde das *eigene Einkommen* als Variable mit berücksichtigt. Dieses hat im Modell für Frauen einen signifikanten positiven Einfluss. Hier nicht präsentierte Spezifikationen haben gezeigt, dass das eigene Einkommen auch bei den Männern signifikant positiv wird, wenn die Erwerbsstatus-Variablen aus der Regression ausgeschlossen werden.

In den Tabellen 13.11 und folgende finden sich die Schätzungen mit dem Datensatz aus der 1995er Welle. Da hier weniger Beobachtungen vorliegen, sind weniger signifikante Effekte zu erwarten. Die Regional-Dummies etwa sind hier nicht signifikant. Die Koeffizienten der Altersgruppen zeigen einen ähnlichen Verlauf wie bei den Schätzungen für die Jahre 1991/93, es sind aber nicht alle signifikant. Bei den Erwerbsstatus-Variablen finden sich die selben Ergebnisse, wie in den anderen Schätzungen. Für das Einkommen der Eltern lässt sich kein signifikanter Einfluss feststellen. Die Vorzeichen beim Einkommen des Vaters sind aber wieder positiv. Die Ausgaben der Eltern für dauerhafte Konsumgüter haben hier das erwartete positive Vorzeichen und sind bei den Männern auch signifikant. Eine allgemeine Aussage lässt sich hier jedoch nicht treffen, da der entsprechende Koeffizient im Datensatz der Jahre 1991/1993 ein entgegengesetztes Vorzeichen hatte. Das eigene Einkommen (Tabellen 13.15 und 13.17) hat hier wieder einen positiven, bei den Männern auch signifikanten Einfluss auf die Auszugswahrscheinlichkeit.

13 Ergebnisse

Tabelle 13.1: Auszugswahrscheinlichkeit 1991/93: Logit-Modell 1

Variable	Koeffizient	(Std. Abw.)
Alter 17-21	1.154**	(0.405)
Alter 22-25	2.039**	(0.399)
Alter 26-30	2.554**	(0.406)
Alter > 30	2.098**	(0.419)
weiblich	0.389**	(0.098)
Nord	0.215	(0.149)
Süd	0.425**	(0.149)
dauerhafte Konsumgüter	0.648 <sup>†</sup>	(0.379)
Einkommen Vater	0.460**	(0.176)
Einkommen Mutter	-0.593	(0.448)
Beschäftigt	1.426**	(0.158)
Selbständig	1.649**	(0.231)
arbeitslos	1.147**	(0.161)
sonstiges	0.804**	(0.286)
Mittelstufe	0.118	(0.215)
Abitur	0.022	(0.225)
d_1993	-0.492**	(0.119)
Konstante	-5.386**	(0.461)
<hr/>		
N	5155	
Log-likelihood	-1500.484	
$\chi^2_{(17)}$	452.732	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

*Bemerkung: Monetäre Größen in 100 Mio. Lire Netto-Jahreseinkommen bzw. jährliche Ausgaben, Ausgelassene Kategorien sind bei den Regionen Mittelitalien, beim Alter die 15-jährigen, beim Erwerbsstatus Schüler, beim Bildungsabschluss die mit Hauptschulabschluss und keinen Abschluss*

### 13 Ergebnisse

Tabelle 13.2: Marginale Effekte

Variable	marginaler Effekt	P
Alter 17-21	0,088	0,016
Alter 22-25	0,187	0,000
Alter 26-30	0,329	0,000
Alter $\geq$ 30	0,264	0,001
weiblich	0,024	0,000
Nord	0,014	0,160
Süd	0,027	0,006
dauerhafte Konsumgüter	0,040	0,088
Einkommen Vater	0,028	0,009
Einkommen Mutter	-0,036	0,186
Beschäftigt	0,121	0,000
Selbständig	0,193	0,000
arbeitslos	0,095	0,000
sonstiges	0,068	0,032
Mittelstufe	0,007	0,578
Abitur	0,001	0,922
Vorhergesagte Auszugswahrscheinlichkeit 0,065		

13 Ergebnisse

Tabelle 13.3: Auszugswahrscheinlichkeit 1991/93: Logit-Modell 1: Männer

Variable	Koeffizient	(Std. Abw.)
Alter 17-21	0.940 <sup>†</sup>	(0.543)
Alter 22-25	1.538**	(0.536)
Alter 26-30	2.259**	(0.542)
Alter > 30	2.254**	(0.556)
Nord	0.151	(0.209)
Süd	0.440*	(0.210)
dauerhafte Konsumgüter	0.371	(0.494)
Einkommen Vater	0.788**	(0.275)
Einkommen Mutter	-0.811	(0.639)
Beschäftigt	1.515**	(0.232)
Selbständig	1.734**	(0.296)
arbeitslos	1.079**	(0.242)
sonstiges	0.523	(0.570)
Mittelstufe	0.464	(0.322)
Abitur	0.373	(0.332)
d_1993	-0.614**	(0.163)
Konstante	-5.457**	(0.626)
<hr/>		
N	2843	
Log-likelihood	-770.962	
$\chi^2_{(16)}$	251.847	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

### 13 Ergebnisse

Tabelle 13.4: Marginale Effekte: Männer

Variable	marginale Effekt	P
Alter 17-21	0,063	0,138
Alter 22-25	0,115	0,021
Alter 26-30	0,248	0,006
Alter $\geq$ 30	0,275	0,009
Nord	0,009	0,475
Süd	0,025	0,043
dauerhafte Konsumgüter	0,021	0,453
Einkommen Vater	0,044	0,005
Einkommen Mutter	-0,045	0,204
Beschäftigt	0,117	0,000
Selbständig	0,190	0,000
arbeitslos	0,080	0,000
sonstiges	0,036	0,451
Mittelstufe	0,024	0,125
Abitur	0,022	0,300
Vorhergesagte Auszugswahrscheinlichkeit 0,059		

13 Ergebnisse

Tabelle 13.5: Auszugswahrscheinlichkeit 1991/93: Logit-Modell 1: Frauen

Variable	Koeffizient	(Std. Abw.)
Alter 17-21	1.393*	(0.611)
Alter 22-25	2.507**	(0.603)
Alter 26-30	2.834**	(0.615)
Alter > 30	1.730**	(0.643)
Nord	0.309	(0.217)
Süd	0.438*	(0.217)
dauerhafte Konsumgüter	1.083 <sup>†</sup>	(0.599)
Einkommen Vater	0.149	(0.280)
Einkommen Mutter	-0.268	(0.630)
Beschäftigt	1.364**	(0.218)
Selbständig	1.452**	(0.433)
arbeitslos	1.207**	(0.218)
sonstiges	1.007**	(0.351)
Mittelstufe	-0.196	(0.298)
Abitur	-0.267	(0.320)
d_1993	-0.359*	(0.177)
Konstante	-5.036**	(0.688)
<hr/>		
N	2312	
Log-likelihood	-712.75	
$\chi^2_{(16)}$	229.109	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

### 13 Ergebnisse

Tabelle 13.6: Marginale Effekte: Frauen

Variable	marginaler Effekt	P
Alter 17-21	0,119	0,054
Alter 22-25	0,272	0,001
Alter 26-30	0,413	0,001
Alter $\geq$ 30	0,207	0,060
Nord	0,021	0,172
Süd	0,029	0,049
dauerhafte Konsumgüter	0,072	0,071
Einkommen Vater	0,010	0,595
Einkommen Mutter	-0,018	0,671
Beschäftigt	0,126	0,000
Selbständig	0,172	0,026
arbeitslos	0,109	0,000
sonstiges	0,098	0,035
Mittelstufe	-0,013	0,522
Abitur	-0,017	0,384
Vorhergesagte Auszugswahrscheinlichkeit		0,071

Tabelle 13.7: Auszugswahrscheinlichkeit 1991/93: Logit-Modell 2, Männer

Variable	Koeffizient	(Std. Abw.)
Alter 17-21	0.937 <sup>†</sup>	(0.543)
Alter 22-25	1.541**	(0.535)
Alter 26-30	2.270**	(0.543)
Alter > 30	2.270**	(0.558)
Nord	0.158	(0.209)
Süd	0.440*	(0.210)
eigenes Einkommen	-0.423	(1.085)
dauerhafte Konsumgüter	0.366	(0.495)
Einkommen Vater	0.800**	(0.276)
Einkommen Mutter	-0.803	(0.640)
Beschäftigt	1.583**	(0.289)
Selbständig	1.803**	(0.344)
arbeitslos	1.079**	(0.242)
sonstiges	0.536	(0.570)
Mittelstufe	0.467	(0.322)
Abitur	0.378	(0.332)
d.1993	-0.616**	(0.163)
Konstante	-5.469**	(0.627)
<hr/>		
N	2843	
Log-likelihood	-770.886	
$\chi^2_{(17)}$	252	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

13 Ergebnisse

Tabelle 13.8: Marginale Effekte: Männer

Variable	marginaler Effekt	P
Alter 17-21	0,063	0,139
Alter 22-25	0,115	0,021
Alter 26-30	0,250	0,005
Alter $\geq$ 30	0,278	0,009
Nord	0,009	0,457
Süd	0,025	0,043
Einkommen	-0,023	0,697
dauerhafte Konsumgüter	0,020	0,459
Einkommen Vater	0,044	0,004
Einkommen Mutter	-0,044	0,209
Beschäftigt	0,124	0,000
Selbständig	0,202	0,001
arbeitslos	0,080	0,000
sonstiges	0,037	0,443
Mittelstufe	0,024	0,124
Abitur	0,023	0,295
Vorhergesagte Auszugswahrscheinlichkeit		0,059

Tabelle 13.9: Auszugswahrscheinlichkeit 1991/93: Logit-Modell 2: Frauen

Variable	Koeffizient	(Std. Abw.)
Alter 17-21	1.419*	(0.611)
Alter 22-25	2.517**	(0.603)
Alter 26-30	2.814**	(0.616)
Alter > 30	1.637*	(0.646)
Nord	0.277	(0.218)
Süd	0.450*	(0.218)
eigenes Einkommen	3.116 <sup>†</sup>	(1.619)
dauerhafte Konsumgüter	1.023 <sup>†</sup>	(0.604)
Einkommen Vater	0.132	(0.283)
Einkommen Mutter	-0.347	(0.634)
Beschäftigt	0.923**	(0.320)
Selbständig	1.013*	(0.497)
arbeitslos	1.199**	(0.218)
sonstiges	0.969**	(0.352)
Mittelstufe	-0.176	(0.299)
Abitur	-0.275	(0.321)
d.1993	-0.375*	(0.177)
Konstante	-5.027**	(0.688)
<hr/>		
N	2312	
Log-likelihood	-710.894	
$\chi^2_{(17)}$	232.819	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

13 Ergebnisse

Tabelle 13.10: Marginale Effekte: Frauen

Variable	marginaler Effekt	P
Alter 17-21	0,121	0,050
Alter 22-25	0,274	0,001
Alter 26-30	0,409	0,001
Alter $\geq$ 30	0,190	0,074
Nord	0,019	0,220
Süd	0,030	0,043
Einkommen	0,206	0,057
dauerhafte Konsumgüter	0,068	0,091
Einkommen Vater	0,009	0,642
Einkommen Mutter	-0,023	0,584
Beschäftigt	0,076	0,017
Selbständig	0,102	0,144
arbeitslos	0,108	0,000
sonstiges	0,093	0,041
Mittelstufe	-0,012	0,566
Abitur	-0,017	0,372
Vorhergesagte Auszugswahrscheinlichkeit 0,071		

13 Ergebnisse

Tabelle 13.11: Auszugswahrscheinlichkeit 1995: Logit-Modell 1, Männer

Variable	Koeffizient	(Std. Abw.)
Alter 16-20	-0.340	(0.793)
Alter 21-25	0.859	(0.769)
Alter 26-30	1.655*	(0.778)
Alter > 30	1.702*	(0.797)
Nord	-0.084	(0.274)
Süd	-0.113	(0.269)
dauerhafte Konsumgüter	-1.029	(0.676)
Einkommen Vater	0.325	(0.371)
Einkommen Mutter	0.511	(0.754)
Beschäftigt	1.183**	(0.324)
Selbständig	1.421**	(0.377)
arbeitslos	1.046**	(0.341)
sonstiges	0.727	(0.616)
Mittelstufe	0.774	(0.563)
Abitur	0.789	(0.563)
Konstante	-4.399**	(0.949)
<hr/>		
N		1106
Log-likelihood		-387.47
$\chi^2_{(15)}$		135.814
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

13 Ergebnisse

Tabelle 13.12: Marginale Effekte: Männer

Variable	marginaler Effekt	P
Alter 16-20	-0,028	0,653
Alter 21-25	0,085	0,328
Alter 26-30	0,209	0,111
Alter ≥30	0,244	0,129
Nord	-0,007	0,756
Süd	-0,010	0,674
dauerhafte Konsumgüter	-0,088	0,128
Einkommen Vater	0,028	0,381
Einkommen Mutter	0,044	0,497
Beschäftigt	0,126	0,002
Selbständig	0,193	0,006
arbeitslos	0,116	0,011
sonstiges	0,082	0,346
Mittelstufe	0,067	0,179
Abitur	0,070	0,182
Vorhergesagte Auszugswahrscheinlichkeit		0,095

13 Ergebnisse

Tabelle 13.13: Auszugswahrscheinlichkeit 1995: Logit-Modell 1, Frauen

Variable	Koeffizient	(Std. Abw.)
Alter 16-20	1.251	(1.044)
Alter 21-25	1.873 <sup>†</sup>	(1.052)
Alter 26-30	2.512*	(1.069)
Alter > 30	2.004 <sup>†</sup>	(1.085)
Nord	0.263	(0.298)
Süd	0.150	(0.287)
dauerhafte Konsumgüter	-0.627	(0.701)
Einkommen Vater	0.071	(0.251)
Einkommen Mutter	-1.304	(0.795)
Beschäftigt	1.232**	(0.295)
Selbständig	1.239*	(0.490)
arbeitslos	1.156**	(0.280)
sonstiges	0.646	(0.520)
Mittelstufe	-0.314	(0.445)
Abitur	-0.355	(0.448)
Konstante	-3.795**	(1.116)
<hr/>		
N	851	
Log-likelihood	-352.013	
$\chi^2_{(15)}$	97.739	
<hr/>		
Signifikanz Niveaus :	† : 10%	* : 5%    ** : 1%

13 Ergebnisse

Tabelle 13.14: Marginale Effekte: Frauen

Variable	marginaler Effekt	P
Alter 16-20	0,171	0,282
Alter 21-25	0,282	0,125
Alter 26-30	0,468	0,039
Alter $\geq 30$	0,374	0,131
Nord	0,032	0,396
Süd	0,018	0,601
dauerhafte Konsumgüter	-0,074	0,369
Einkommen Vater	0,008	0,779
Einkommen Mutter	-0,154	0,100
Beschäftigt	0,184	0,000
Selbständig	0,213	0,048
arbeitslos	0,171	0,000
sonstiges	0,093	0,294
Mittelstufe	-0,036	0,472
Abitur	-0,042	0,436
Vorhergesagte Auszugswahrscheinlichkeit		0,137

13 Ergebnisse

Tabelle 13.15: Auszugswahrscheinlichkeit 1995: Logit-Modell 2, Männer

Variable	Koeffizient	(Std. Abw.)
Alter 16-20	-0.343	(0.793)
Alter 21-25	0.857	(0.771)
Alter 26-30	1.535*	(0.783)
Alter > 30	1.503 <sup>†</sup>	(0.803)
Nord	-0.116	(0.279)
Süd	-0.082	(0.273)
eigenes Einkommen	3.618**	(1.194)
dauerhafte Konsumgüter	-1.581*	(0.764)
Einkommen Vater	0.355	(0.378)
Einkommen Mutter	0.336	(0.767)
Beschäftigt	0.575	(0.386)
Selbständig	0.800 <sup>†</sup>	(0.435)
arbeitslos	1.042**	(0.342)
sonstiges	0.629	(0.620)
Mittelstufe	0.780	(0.562)
Abitur	0.767	(0.562)
Konstante	-4.327**	(0.950)
<hr/>		
N	1106	
Log-likelihood	-381.754	
$\chi^2_{(16)}$	147.246	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

13 Ergebnisse

Tabelle 13.16: Marginale Effekte: Männer

Variable	marginaler Effekt	P
Alter 16-20	-0,028	0,651
Alter 21-25	0,085	0,331
Alter 26-30	0,188	0,137
Alter ≥30	0,204	0,178
Nord	-0,010	0,674
Süd	-0,007	0,763
Einkommen	0,309	0,004
dauerhafte Konsumgüter	-0,135	0,037
Einkommen Vater	0,030	0,347
Einkommen Mutter	0,029	0,661
Beschäftigt	0,054	0,169
Selbständig	0,089	0,138
arbeitslos	0,115	0,012
sonstiges	0,068	0,407
Mittelstufe	0,067	0,176
Abitur	0,068	0,193
Vorhergesagte Auszugswahrscheinlichkeit		0,094

13 Ergebnisse

Tabelle 13.17: Auszugswahrscheinlichkeit 1995: Logit-Modell 2, Frauen

Variable	Koeffizient	(Std. Abw.)
Alter 16-20	1.260	(1.043)
Alter 21-25	1.898 <sup>†</sup>	(1.053)
Alter 26-30	2.508*	(1.069)
Alter > 30	1.925 <sup>†</sup>	(1.087)
Nord	0.230	(0.300)
Süd	0.145	(0.287)
eigenes Einkommen	2.667	(1.891)
dauerhafte Konsumgüter	-0.650	(0.712)
Einkommen Vater	0.048	(0.264)
Einkommen Mutter	-1.446 <sup>†</sup>	(0.810)
Beschäftigt	0.818 <sup>†</sup>	(0.421)
Selbständig	0.947 <sup>†</sup>	(0.538)
arbeitslos	1.142**	(0.280)
sonstiges	0.605	(0.523)
Mittelstufe	-0.327	(0.446)
Abitur	-0.395	(0.450)
Konstante	-3.741**	(1.116)
<hr/>		
N	851	
Log-likelihood	-351.018	
$\chi^2_{(16)}$	99.73	
<hr/>		
Signifikanz Niveaus : † : 10% * : 5% ** : 1%		

13 Ergebnisse

Tabelle 13.18: Marginale Effekte: Frauen

Variable	marginaler Effekt	P
Alter 16-20	0,172	0,279
Alter 21-25	0,286	0,120
Alter 26-30	0,467	0,040
Alter $\geq$ 30	0,355	0,151
Nord	0,028	0,458
Süd	0,017	0,612
Einkommen	0,314	0,160
dauerhafte Konsumgüter	-0,077	0,360
Einkommen Vater	0,006	0,856
Einkommen Mutter	-0,170	0,073
Beschäftigt	0,113	0,088
Selbständig	0,150	0,160
arbeitslos	0,168	0,000
sonstiges	0,086	0,324
Mittelstufe	-0,038	0,454
Abitur	-0,047	0,389
Vorhergesagte Auszugswahrscheinlichkeit		0,137

## 14 Schlussfolgerung

Ziel der Untersuchung war, den Einfluss ökonomischer Faktoren auf die Auszugswahrscheinlichkeit junger Menschen in Italien zu bestimmen. Es zeigt sich, dass mit der Beendigung der Schulausbildung die Wahrscheinlichkeit, aus dem Elternhaus auszuziehen zunimmt.

Der Erwerbsstatus und das Einkommen junger Menschen sind wesentliche Einflussfaktoren für die Wahrscheinlichkeit, das Elternhaus zu verlassen. Arbeitslose junge Männer bleiben mit einer höheren Wahrscheinlichkeit im Haushalt der Eltern als beschäftigte. Junge Menschen mit höherem Einkommen verlassen mit einer größeren Wahrscheinlichkeit das Elternhaus. Auch ein höheres Einkommen des Vaters erhöht die Wahrscheinlichkeit eines Auszugs. Bei den Ausgaben für dauerhafte Konsumgüter durch die Eltern lies sich kein eindeutiger Zusammenhang finden.

Für künftige Forschungen ist vor allem von Interesse, welche Auswirkungen die Rolle der familiären Netzwerke auf den Arbeitsmarkt hat. Familiäre Beziehungen, die Versicherungsfunktionen übernehmen, können zu ähnlichen Anreizproblemen führen, wie Mindestlöhne oder staatliche Transfers. Um zu klären, ob die familiären Strukturen einen Teil der Arbeitslosigkeit in den Mittelmeerländern erklären können, sind also noch weitere Untersuchungen erforderlich.

# Literaturverzeichnis

- Aassve, A., Billari, F. C., Mazzuco, S. und Ongaro, F.** (2001). Leaving home ain't easy – a comperative longitudinal analysis of echp data. MPIDR Workingpaper 2001-038.
- BdW** (2001). Bild der Wissenschaft Online, newsticker vom 4.7.2001. <http://www.wissenschaft.de/sixcms/detail.php?id=95827>.
- Bean, C. R.** (1994). European unemployment: A survey. *Journal of Economic Literature*, 32, 573–619.
- Becker, G. S.** (1981). *A Treatise on the Family*. Harvard University Press, Cambridge.
- Bentolila, S. und Ichino, A.** (2000). Unemployment and consumption: Are job losses less painful in the Mediterranean? Discussionpaper des Centro de Estudios Monetarios Y Financieros.
- Bergstorm, T. C.** (1997). A survey of theories of the family. In: *Handbook of Population Economics*, Bd. 1a.
- Blanchard, O. und Wolfers, J.** (1999). The role of shocks and institutions in the rise of european unemployment: The aggregate evidence. Harry Johnson Lecture.
- Cantó, O. und Mercader-Prats, M.** (2001). Young people leaving home: the impact on poverty in Spain. In: B. Bradbury, S. Jenkins und J. Micklewright, Hg., *The Dynamics of Child Poverty in Industrialised Countries*. Cambridge University Press, Cambridge.
- Ermisch, J.** (1986). An economic perspective on household modelling. In: N. Keilman, A. Kuijsten und A. Vossen, Hg., *Modelling Household Formation and Dissolution*. Oxford University Press, Oxford.

- Ermisch, J.** (1999). Prices, parents, and young people's household formation. *Journal of Urban Economics*, 45, 47–71.
- Ermisch, J.** und **DiSalvo, P.** (1997). The economic determinants of young people's household formation. *Economica*, 64, 627–644.
- Fogli, A.** (2000). Endogenous labour market rigidities and family ties. mimeo. University of Pennsylvania.
- Ghidoni, M.** (2001). Determinants of young Europeans' decision to leave the parental household. Workingpaper präsentiert auf der Royal Economic Society Annual Conference 2002. Mimeo.
- Giannelli, G. C.** und **Monfardini, C.** (2000). Joint decisions on household membership and human capital accumulation of youth. IZA Discussion Paper No. 191.
- Iacovou, M.** (1998). Young people in Europe: Two models of household formation. Working Papers of the ESRC Research Centre on Micro-social Change, Paper 98-13. Colchester, University of Essex.
- Ichino, A., Bentolila, S.** und **Becker, S. O.** (2002). Parental job insecurity and children's emancipation: The Italian puzzle. Workingpaper präsentiert auf der Jahrestagung der European Economic Association 2002 in Venedig, Italien.
- Klasen, S.** und **Woolard, I.** (1999). Surviving unemployment without state support: Unemployment and household formation in South Africa. SFB 386, Ludwig-Maximilians-Universität München, Discussion Paper 213.
- Layard, R., Nickell, S.** und **Jackman, R.** (1991). *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford University Press, Oxford.
- Manacorda, M.** und **Moretti, E.** (2001). The evolution of living arrangements among young italian men. mimeo, LSE-UCLA.
- Manacorda, M.** und **Moretti, E.** (2002). Intergenerational transfers and household structure: Why do most italian youths live with their parents? Workingpaper präsentiert auf der ESPE 2002 in Bilbao, Spanien.
- McElroy, M. B.** (1985). The joint determination of household membership and market work: The case of young man. *Journal of Labour Economics*, 3(3), 293–315.

- McElroy, M. B.** und **Horney, M. J.** (1981). Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review*, 22(2), 333–349.
- Nash, J.** (1953). Two-person cooperative games. *Econometrica*, 21(1), 128–140.
- OECD** (1999). *Benefit systems and work incentives*. Paris.
- OECD** (2001). *Employment Outlook 2001*. Paris.
- Ott, N.** (1986). Ausscheiden erwachsener Kinder aus dem elterlichen Haushalt. In: K. F. Zimmermann, Hg., *Demographische Probleme der Haushaltsökonomie*, Bd. IV von *Contributions to Quantitative Economics*. Brockmeyer, Bochum.
- Papastefanou, C.** (2000). Der Auszug aus dem Elternhaus – Ein vernachlässigter Gegenstand der Entwicklungspsychologie. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 20(1), 55–69.
- Rosenzweig, M. R.** und **Wolpin, K. I.** (1993). Intergenerational support and the life-cycle incomes of young men and their parents: Human capital investment s, core-sidence, and intergenerational financial transfers. *Journal of Labor Economics*, 11(1), 84–112.
- Rosenzweig, M. R.** und **Wolpin, K. I.** (1994). Parental and public transfers to young women and their children. *American Economic Review*, 84(5), 1195–1212.
- Wenzel, J.-O.** (1995). Arbeitslosigkeit in der OECD und die Suche nach ihren Ursachen. In: J. Flemmig, Hg., *Moderne Makroökonomik – Eine kritische Bestandsaufnahme*, S. 135–166. Metropolis-Verlag.

## **Teil III**

# **Multiple Imputation von fehlenden Werten mit Daten über Unterernährung und Kindersterblichkeit**

## Zusammenfassung

In dieser Arbeit werden die Auswirkungen einer Ersetzung von fehlenden Werten auf das Ergebnis einer Regressionsanalyse untersucht. Grundlage ist eine Untersuchung von Klasen (2000) über die Unterschiede im Zusammenhang zwischen Unterernährung und Kindersterblichkeit in Afrika und Südasien. In dem Makro-Datensatz, welcher 101 Entwicklungsländer umfasst, fällt etwa ein Drittel der 273 Beobachtungen weg, da für verschiedene verwendete Variablen die Werte fehlen. Die so verloren gegangenen Informationen sollen in diese Untersuchung genutzt werden um die Schätzergebnisse zu verbessern. Hierzu wird ein Verfahren zur multiplen Imputation verwandt, in welchem mit einem Markov-Chain-Monte-Carlo-Verfahren mehrere vervollständigte Datensätze generiert werden, mit welchen dann getrennt Schätzungen durchgeführt werden. Die Ergebnisse der Schätzungen werden dann miteinander kombiniert. Durch die Auswertung mehrerer vervollständigter Datensätze wird eine höhere Effizienz der Schätzer erreicht.

Ein Vergleich von Regressionsanalysen, die mit dem vervollständigten Daten durchgeführt wurden, mit einer Complete-case-Analyse hat gezeigt, dass sich bestimmte Koeffizienten in ihrer Größenordnung geändert haben. Bei manchen Koeffizienten sind unplausible Vorzeichen aus der Complete-case Analyse verschwunden. Es ist also vorteilhaft, bei Problemen mit fehlenden Werten moderne Imputationsverfahren zu verwenden. Die wesentlichen Ergebnisse aus der Untersuchung von Klasen (2000) konnten dennoch bestätigt werden.

Durch die Ersetzung der fehlenden Werte konnten noch eine Reihe von Variablen zugänglich gemacht werden, die in den bisherigen Untersuchungen nicht verwendet wurden, da dadurch auf noch mehr Beobachtungen hätte verzichtet werden müssen.

JEL-Klassifikation: C15, O11,

## 15 Einführung

In den meisten empirischen Arbeiten zu ökonomischen Fragen wird auf das Problem fehlender Werte nicht eingegangen. Wenn bei einigen Variablen Werte fehlen, wird meist die entsprechende Beobachtung aus der Untersuchung ausgeklammert. Damit gehen Informationen verloren, die effizienter genutzt werden könnten. Insbesondere bei Querschnittsdaten über mehrere Länder, wie sie auch hier verwendet werden, ist das Problem der fehlenden Werte oft gravierend. So zum Beispiel auch bei der Querschnittsuntersuchung von Mankiw, Romer und Weil (1992) zur Wachstumstheorie, in welcher eine Reihe von Beobachtungen (insbesondere von Entwicklungsländern) wegfallen, weil Werte fehlen.

Auf Grundlage einer Untersuchung von Klasen (2000) über den Zusammenhang von Kindersterblichkeit und Unterernährung sollen hier die aufgrund fehlender Werte verloren gegangenen Informationen genutzt werden und geprüft werden, ob sich die Ergebnisse aus der genannten Untersuchung bestätigen. Darüber hinaus werden nun Variablen aus dem Datensatz zugänglich, die bei der Complete-Case-Analyse nicht verwendet werden konnten.

Um alle Informationen im Datensatz zu nutzen, ist es erforderlich die fehlenden Werte durch Schätzungen zu ersetzen. In dieser Arbeit wird ein multiple-Imputation Verfahren angewandt, welches mehrere vervollständigte Datensätze mittels eines Markov-Chain-Monte-Carlo-Prozesses generiert.

Die Arbeit ist wie folgt aufgebaut. Zunächst werden wichtige Ergebnisse aus der Untersuchung, worauf sich diese Arbeit bezieht, vorgestellt. Danach werden Probleme verschiedener Ersetzungsverfahren diskutiert und das multiple-Imputation Verfahren erläutert. Schließlich werden die Ergebnisse der Regressionen vorgestellt und mit der Untersuchung aus Klasen (2000) verglichen.

# 16 Unterernährung und Kindersterblichkeit

## 16.1 Fragestellung

Unterernährung und Kindersterblichkeit sind zentrale Indikatoren, an denen der Entwicklungsstand eines Landes abgelesen werden kann (Sen, 1999). Um die knappen Ressourcen, die für Entwicklungshilfe verwendet werden, effizient einsetzen zu können, ist es erforderlich, die genauen Erklärungsfaktoren der Unterernährung und der Kindersterblichkeit zu bestimmen.

In Ländern Südasiens lässt sich eine höhere Unterernährung beobachten, als in Ländern Afrikas. Auf der anderen Seite zeigt sich, dass die Kindersterblichkeit in Afrika höher ist als in Südasiens.

Es stellt sich die Frage, wie dieses Muster zu erklären ist. Sowohl Unterernährung, als auch Kindersterblichkeit werden von den ähnlichen Faktoren beeinflusst und bedingen sich gegenseitig. Zu erwarten wäre, dass in Ländern mit hoher Unterernährung auch die Kindersterblichkeit besonders hoch ist. Wenn in einem Land die Ernährungssituation schlecht ist, führt dies zu Unterernährung, aber auch zu einer höheren Anfälligkeit gegenüber Erkrankungen, was auch zu einer höheren Mortalitätsrate führt.

## 16.2 Datensatz

Es liegen Daten aus insgesamt 101 Ländern vor, die aus einem Zeitraum von 1965 bis 1996 stammen. Für einen Teil der Länder liegen Beobachtungen aus mehreren Jahren vor. Der Datensatz hat also eine Panelstruktur. Welche Länder dies sind und wie viel Beobachtungen aus den jeweiligen Ländern vorliegen, ist in einer Tabelle im Anhang zu finden.

Zur Beschreibung der Unterernährungssituation dienen folgende Variablen. Der Anteil der Säuglinge mit geringem Geburtsgewicht an allen Säuglingen ist kennzeichnend für die Ernährungssituation der Mutter. Als gering gilt ein Geburtsgewicht von weniger als 2.500 Gramm.

Der Ernährungszustand der Kinder wird mit anthropometrischen Indikatoren gemessen, welche die WHO verwendet (WHO, 1983). Hierzu dient ein Z-Score. Der Z-Score ist ein standardisierter Messwert, welcher die Distanz einer Beobachtung vom Mittelwert einer Referenzgruppe misst.

Ein Maß für das Untergewicht des Individuums  $i$  im Alter  $t$  ist das gemessene Gewicht abzüglich der durchschnittlichen Gewichtes der Kinder  $\mu_t$  im gleichen Alter  $t$  geteilt durch die Standardabweichung  $\sigma_t$ .

$$Z_i = \frac{x_i - \mu_t}{\sigma_t} \quad (16.1)$$

Kinder die mehr als zwei Standardabweichungen vom Mittelwert abweichen sind leicht untergewichtig. Starkes Untergewicht liegt vor bei drei Standardabweichungen oder mehr.

Mit zwei weiteren Z-Scores lässt sich zwischen akuter und chronischer Unterernährung unterscheiden. Ein Indikator für chronische Unterernährung ist die Wachstumsverzögerung (stunting). Sie liegt vor bei einer zu geringe Größe der Kinder für ihr Alter. Kinder die dauerhaft zu wenig Nahrungszufuhr erhalten, passen ihr Wachstum an und sind somit bei gegebenem Alter kleiner.

Auszehrung (wasting) wird gemessen in Gewicht bei gegebener Körpergröße. Sie ist meist Ergebnis von kürzlich durchlebten Hunger- oder Krankheitsperioden und misst die akute Unterernährung.

Die Variablen, die hier verwendet werden (Tabelle 17.1), geben den Anteil der Kinder mit akuter Unterernährung (Auszehrung) bzw. mit verzögertem Wachstum an allen Kindern an.

Indikator für die Kindersterblichkeit ist die Mortalitätsrate der under 5-jährigen. Sie ist definiert als die Wahrscheinlichkeit, zwischen der Geburt und dem vollendeten 5. Lebensjahr zu sterben, bezogen auf 1000 Lebendgeburten.

Zur Erklärung der Unterernährung und Kindersterblichkeit stehen eine Reihe von Charakteristika zur Verfügung. Ein Teil der Variablen wurden in der bisherigen Untersuchung von Klasen (2000) nicht verwendet, weil dadurch noch mehr Beobachtungen verloren gegangen wären, als es ohnehin der Fall war. Einige Variablen werden für das Imputationsverfahren hinzugezogen, da sie mit den Variablen, bei welchen Werte fehlen, korreliert sind. Die Tabelle 17.1 gibt einen Überblick über die verwendeten Variablen.

Eine Diskussion der Ursachen der Unterernährung findet sich in UNICEF (1998). Dort wird unterschieden zwischen grundlegenden (indirekten) Ursachen und direkten Ursachen.

Zu den grundlegenden Ursachen gehören die Ressourcen, die durch die Umwelt, die

technischen Möglichkeiten und durch die Menschen zur Verfügung gestellt werden. Um diese Grundlegenden Faktoren zu erfassen wird das Bruttoinlandsprodukt pro Kopf aufgenommen. Ein höheres Sozialprodukt erhöht die Ressourcen, über welche die Haushalte verfügen, und senkt so Unterernährung und Kindersterblichkeit.

Um auch zu berücksichtigen, wie das Einkommen verteilt ist, wird der Gini-Koeffizient in den Untersuchungen aufgenommen. Er misst die Ungleichverteilung des Einkommens in dem Land. In einem Land, in welchem ein großer Teil der Menschen vom Reichtum ausgeschlossen sind, ist eine höhere Unterernährung und Kindersterblichkeit zu erwarten. Auf diesen Zusammenhang hat Sen hingewiesen (z. B. Drèze und Sen (1995)).

Die Bevölkerungsdichte wirkt nachteilig auf die Unterernährung und Kindersterblichkeit weil Krankheiten sich leichter ausbreiten und weil die Landfläche relativ zur Bevölkerungsgröße knapper wird, der Preis für Nahrungsmittel und Energie ist damit höher.

Die Variable Tropenregion gibt an, welcher Flächenanteil des Landes in tropischen Gebieten liegt. In diesen Gebieten ist die Wahrscheinlichkeit einer Malariaerkrankung besonders hoch. Malaria ist in Afrika die häufigste Todesursache der unter 5-jährigen.

Direkte Einflussfaktoren sind der Zugang zu sanitären Anlagen, der Zugang zu Wasser sowie der Zugang zu Gesundheitsversorgung. Diese verbessern die Gesundheitssituation der Kinder und senken damit die Mortalitätsrate und die Unterernährung. Ein weiterer Indikator für die Gesundheitssituation ist der Anteil der Kinder, die gegen Tuberkulose bzw. gegen Tetanus, Diphtherie und Polio geimpft wurden.

Bei einer höheren Fruchtbarkeitsrate der Frauen ist zu erwarten, dass sie Unterernährung und Kindersterblichkeit erhöht, da bei einer höheren Fruchtbarkeitsrate die Zahl der Kindern, die eine Mutter zu versorgen hat, größer ist und somit die Mittel des Haushaltes sich auf mehr Kinder aufteilen. Die gesamte Fruchtbarkeitsrate, die hier verwandt wird, ist definiert als die Zahl der Kinder, welche eine Frau, die wenigstens bis zum Ende des gebärfähigen Alters lebt, im Durchschnitt gebiert.

Eine höhere Bildung der Mutter führt dazu, dass deren Kinder besser versorgt und bei Krankheit besser gepflegt werden. Es wird daher erwartet, dass sie sich positiv auf die Unterernährung und die Kindersterblichkeit auswirkt. Die Qualifikation der Mutter wird durch die Alphabetisierungsquote der Frauen erfasst.

Mehrere Variablen liegen vor, die das Stillverhalten der Mütter beschreiben. Die Stilldauer ist zum einen von Interesse, weil eine lange Stilldauer auf eine schlechte Nahrungsversorgung hindeutet. Wenn genügend andere Nahrung für das Kind vorhanden wäre, würde die Muttermilch viel früher abgesetzt.

Zum anderen führt das optimale Stillverhalten der Mütter zu einer besseren Gesundheits-

und Ernährungssituation des Kindes. Dieses wird beschrieben durch den Anteil der Kinder von der Geburt bis zum 3. Monat ausschließlich gestillt wurden sowie durch den Anteil der Kinder, die vom 6. bis 9. Monat gestillt und mit Zusatznahrung gefüttert wurden.

Als weitere Variablen stehen zur Verfügung die Kalorienzufuhr pro Kopf der Bevölkerung, die Malariafälle in % der Bevölkerung sowie die durchschnittliche Lebenserwartung. Das Verhältnis der Lebenserwartung der Frauen zu jener der Männer steht für das Ausmaß der Geschlechterdiskriminierung im Land. Die Geschlechterdiskriminierung korreliert mit mehreren Faktoren, wie z.B. dem Sozialprodukt oder der Fruchtbarkeitsrate (vgl. Klasen, 1999).

Tabelle 17.1 gibt auch einen Überblick über die Zahl der Beobachtungen, die fehlen. Von der Variable Stilldauer fehlt über die Hälfte der Beobachtungen. Von den Unterernährungs- und Sterblichkeits-Variablen, also die zu erklären Variablen, fehlen 10-30%. Von allen Variablen, die in der Tabelle aufgelistet sind, fehlen 24 % der Werte. Dennoch fallen bei der Schätzung der Regressions-Modelle zum Teil über die Hälfte der Beobachtungen weg. Ein großer Teil der Informationen bleibt also ungenutzt.

## 17 Die Complete-case-Analyse in Klasen (2000)

In Klasen (2000) wird untersucht, weshalb in Südasien relativ hohe Unterernährung zu beobachten ist, die Kindersterblichkeit jedoch vergleichsweise gering ist, in Afrika südlich der Sahara aber eine höhere Kindersterblichkeit mit einer geringeren beobachteten Unterernährung einher geht.

Hierzu wird im Rahmen einer Regressionsanalyse versucht, die Erklärungsfaktoren für Variablen, welche Unterernährung und Kindersterblichkeit messen, zu finden.

Zentrale Ergebnisse der Regressionen aus Klasen (2000) sind in den Tabellen 17.2 und 17.3 zu finden. Fixe Effekte der Regionen werden durch Dummyvariablen erfasst. Unterschieden werden die Regionen: Ostasien/Pazifik, Osteuropa/Zentralasien, Mittlerer Osten/Nordafrika, Karibik, Südasien, Subsahara und Lateinamerika, wobei letztere als Referenzkategorie Verwendung findet. Um zeitliche Veränderungen aufzufangen, werden die Perioden vor 1985, 1985 bis 1989, 1990 bis 1994 und 1995 und später unterschieden.

In Tabelle 17.2 werden Unterernährungsmodelle geschätzt. Die Ergebnisse entsprechen den theoretischen Erwartungen. In den ersten drei Spalten ist die abhängige Variable der Anteil der untergewichtigen Neugeborenen an allen Neugeborenen. Ein höheres Einkommen verringert den Anteil der Kinder mit geringem Geburtsgewicht.

In Spalte (3) wurde ein reduziertes Modell geschätzt, bei dem die Fruchtbarkeitsrate ausgelassen wurde, da diese stark von der Bildung der Frauen abhängt. Bei der Regression in Spalte (1) fallen fast 90 Beobachtungen von den potentiell 273 Beobachtungen durch fehlende Werte weg. Durch die Hinzunahme der Stilldauer in Spalte (2) fällt über die Hälfte der Beobachtungen weg. Aus Osteuropa und Zentralasien sind dann keine Beobachtungen mehr vorhanden.

	N	Mittelwert	Standard- abweichung	Anzahl	Fehlende Prozent
Variablen verwendet in Klasen(2000)					
Geringes Geburtsgewicht (%)	207	13,59	8,11	66	24,18
Mod. + starkes Untergewicht (%)	188	5,76	6,04	85	31,14
Starkes Untergewicht (%)	243	20,99	15,00	30	10,99
Mod. + starke Wachstums- verzögerung (%)	216	30,57	15,87	57	20,88
Sterblichkeit der unter 5jährigen	220	107,64	71,12	53	19,41
Bevölkerungs- dichte	268	87,30	131,86	5	1,83
ln(BIP)	258	7,44	0,76	15	5,49
Alphabetisierung Frauen (%)	247	60,03	28,68	26	9,52
Stilldauer 20-23 Mon. (%)	133	41,17	20,34	140	51,28
Fruchtbarkeitsra- te	268	4,77	1,68	5	1,83
Sanitärzugang (%)	194	54,00	29,03	79	28,94
Impfung Tetanus (%)	205	67,99	24,21	68	24,91
zusätzliche Variablen					
Tropenregion (%)	271	0,75	0,40	2	0,73
Kalorien pro Kopf	226	2398,87	380,89	47	17,22
Stilldauer 3 Mon. (%)	153	34,86	23,47	120	43,96
Stilldauer 6-9 Mon. (%)	139	55,14	21,28	134	49,08
Malaria (%)	221	2,18	6,23	52	19,05
Gini	192	45,62	8,93	81	29,67
Lebenserwartung	230	59,99	9,72	43	15,75
Lebenserwartung Frauen/Männer	273	1,07	0,03	0	0,00
Impfung Tuberkulose (%)	201	79,51	21,46	72	26,37
Wasserzugang (%)	228	62,37	23,87	45	16,48
Gesundheitsver- sorgung (%)	154	66,91	20,65	119	43,59
Jahr	273	1988,41	6.339	0	0,00

Tabelle 17.1: Beschreibung des Datensatzes

Abhängige Variable	1 geringes Ge- burtsgewicht	2 geringes Ge- burtsgewicht	3 geringes Ge- burtsgewicht	4 leichte + starke Wachstums- verzögerung	5 leichte + starke Wachstums- verzögerung	6 starkes Untergewicht	7 leichtes + starkes Untergewicht	8 leichtes + starkes Untergewicht
Ostasien/Pazifik	2,500 (1,03)	2,753 (1,39)	2,331 (1,03)	6,411 (2,58)	6,031 (2,59)	2,702 (0,82)	15,650 (1,73)	15,580 (1,75)
Osteurope/Zentralasien	-3,817 (2,51)		-4,389 (2,46)		-8,633 (3,47)	0,320 (1,55)		-2,905 (2,80)
Mittlerer Osten/Nordafrika	-2,029 (1,26)	1,211 (1,55)	-2,158 (1,26)	0,432 (2,88)	-3,533 (2,72)	-1,409 (1,06)	0,164 (2,28)	-3,083 (2,15)
Karibik	-2,023 (1,75)	0,874 (2,66)	-2,251 (1,74)	-8,821 (3,56)	-11,601 (2,95)	-0,360 (1,08)	-1,294 (3,99)	-1,549 (2,35)
Südasien	15,304 (1,89)	13,305 (2,04)	15,001 (1,89)	9,683 (3,99)	10,620 (3,74)	10,103 (1,44)	20,642 (4,13)	20,677 (2,96)
Subsahara	-2,866 (1,14)	-2,150 (1,39)	-2,800 (1,14)	-7,936 (2,57)	-5,079 (2,31)	-0,337 (0,88)	5,946 (2,17)	2,957 (1,81)
d9094	0,088 (0,86)	-0,411 (1,00)	0,148 (0,86)	-3,737 (2,02)	-1,724 (1,86)	-0,347 (0,69)	0,563 (1,53)	-0,204 (1,44)
d8589	1,616 (0,96)	2,150 (1,11)	1,837 (0,95)	-3,850 (2,26)	-2,199 (2,09)	-0,759 (0,79)	-1,136 (1,76)	-0,620 (1,65)
pre1985	1,978 (1,26)	4,037 (3,99)	2,154 (1,25)	-1,686 (2,39)	2,765 (2,15)	0,520 (0,82)	-0,190 (2,33)	0,382 (1,66)
Bevölkerungsdichte	0,019 (0,00)	0,018 (0,00)	0,018 (0,00)	0,023 (0,01)	0,017 (0,01)	0,007 (0,00)	0,022 (0,01)	0,021 (0,00)
Fruchtbarkeitsrate	0,354 (0,38)	0,180 (0,52)		3,259 (0,84)			1,704 (0,66)	
Alphabetisierung d. Frauen	-0,045 (0,02)	0,005 (0,03)	-0,055 (0,02)	0,026 (0,05)	-0,121 (0,04)	-0,066 (0,01)	-0,030 (0,04)	-0,150 (0,03)
ln(BIP)	-3,136 (0,68)	-3,302 (1,20)	-3,338 (0,61)	-7,602 (1,74)	-12,191 (1,31)	-2,722 (0,46)	-2,876 (1,49)	-6,960 (0,97)
Stillddauer20-23		0,068 (0,03)						
Sanitärzugang				-0,140 (0,04)			-0,090 (0,04)	
geringes Geburtsgewicht							0,209 (0,15)	
Konstante	35,429 (6,50)	31,378 (11,64)	39,231 (4,56)	78,426 (15,60)	128,598 (9,57)	28,639 (3,43)	31,172 (13,88)	76,564 (7,19)
Angepasstes $R^2$	0,773	0,760	0,770	0,725	0,702	0,746	0,828	0,776
Ramsey Test P:	0,000	0,000	0,000	0,831	0,591	0,000	0,402	0,321
N	186	114	187	138	190	170	136	217

Tabelle 17.2: Unternährungsmodelle mit unvollständigen Datensatz

Für alle Tabellen gilt: Ausgelassene Kategorien sind Lateinamerika und 1995+; Standardabweichungen in Klammern

Die Spalten (4) und (5) zeigen das Ergebnis eines Regressionsmodells, das die Wachstumsverzögerung, also chronische Unterernährung, erklären soll. In Regression (5) wurden die Variablen Sanitärzugang und Fruchtbarkeitsrate weg gelassen, wodurch sich die Zahl der verwendbaren Beobachtung um 22 auf 190 erhöhte. Auch hier haben die Koeffizienten das erwartete Vorzeichen. Der Regionalkoeffizient für Südasien ist signifikant und hat den höchsten Betrag. Die gemessene Wachstumsverzögerung ist dort also besonders groß. Die letzten drei Regressionen versuchen den Anteil derer, die für ihre Größe zu wenig wiegen, zu erklären. Die Ergebnisse sind ähnlich wie bei den Regressionen zur Wachstumsverzögerung. Auch hier fallen zahlreiche Beobachtungen weg, wenn die zusätzlichen Variablen Sanitärzugang und geringes Geburtsgewicht hinzugezogen werden.

Ein Ergebnis von Klasen (2000) ist, dass die Ursachen der höheren Kindersterblichkeit in Afrika unter anderem in höheren Fruchtbarkeitsraten, geringerem Einkommen, geringerer Bevölkerungsdichte und schlechterem Impfschutz zu finden sind. Der Einfluss dieser Faktoren auf die Sterblichkeit der unter 5-jährigen wurde mit einer linearen Regression geschätzt, deren Ergebnisse in Tabelle 17.3 zu sehen sind.

Auch hier fallen durch die Hinzunahme der Regressoren geringes Geburtsgewicht und Untergewicht gegenüber dem Modell in Spalte (1) zahlreiche Beobachtungen weg. Die Hinzunahme dieser Variablen ändert nicht die Ergebnisse der Regression und deren Einfluss ist nicht signifikant. Ein Regressionsmodell, das nur die Unterernährungsvariablen enthält ist in Spalte (4) angegeben. Hier ist nur der Koeffizient Untergewicht signifikant. In Spalte (5) sind wieder die Ergebnisse eines Modells in reduzierter Form angegeben, in welchem die Fruchtbarkeitsrate ausgelassen wurde.

Eine Vermutung in Klasen (2000) zur Erklärung des beobachteten Musters der Unterernährung und der Kindersterblichkeit ist, dass die beobachtete Unterernährung durch Messfehler zu erklären ist. Er argumentiert, dass kleine Messfehler bereits zu großen Unterschieden in den gemessenen Anteilen der unterernährten Kinder führen können. Wenn keine Messfehler vorlägen, wäre zu erwarten, dass die Unterschiede in der Unterernährung durch die vorhandenen Kovariablen erklärt werden könnten. Es zeigt sich jedoch, dass es sehr große regionenspezifische Effekte gibt, die nicht durch die Regressoren erklärt werden können.

So zeigt der Koeffizient *Südasien* den höchsten Wert der Regionendummies. Diese Beobachtung ist konsistent mit der These, dass die Population in Südasien für die Körpergröße ein geringeres genetisches Potential hat, als die amerikanische Referenzgruppe, bezüglich deren die Unterernährung gemessen wird. Es ist also zu vermuten, dass die

17 Die Complete-case-Analyse in Klassen (2000)

	1	2	3	4	5
Abhängige Variable	Sterblichkeit der unter 5jährigen				
Ostasien/Pazifik	-15,504 (8,64)	-10,236 (8,97)	-15,656 (11,32)		-18,300 (8,72)
Osteurope/Zentralasien	-0,619 (13,09)	4,399 (20,22)	6,471 (20,23)		-11,610 (12,70)
Mittlerer Osten/Nordafrika	-24,417 (10,10)	-13,807 (10,65)	-14,714 (10,93)		-25,911 (10,24)
Karibik	-23,053 (12,19)	-20,066 (14,23)	-17,403 (14,31)		-26,772 (12,31)
Südasien	-18,682 (13,98)	-17,309 (18,17)	-28,026 (21,46)		-25,106 (14,01)
Subsahara	4,605 (9,19)	19,747 (9,86)	20,833 (10,01)		7,747 (9,27)
d9094	1,799 (6,59)	-0,922 (6,80)	-2,326 (6,92)		3,708 (6,66)
d8589	12,263 (7,46)	1,453 (7,86)	2,082 (8,12)		15,377 (7,48)
pre1985	16,007 (10,68)	36,488 (24,30)	35,071 (24,58)		23,424 (10,52)
Bevölkerungsdichte	0,011 (0,02)	-0,030 (0,03)	-0,043 (0,03)		-0,001 (0,02)
Fruchtbarkeitsrate	8,669 (3,06)	6,549 (3,24)	5,863 (3,39)		
Alphabeteisierung d. Frauen	-0,923 (0,17)	-0,733 (0,18)	-0,767 (0,19)		-1,154 (0,15)
ln(BIP)	-34,238 (5,38)	-28,516 (6,14)	-25,713 (6,61)		-40,617 (4,94)
Impfung		-0,300 (0,16)	-0,295 (0,17)		
geringes Geburtsgewicht		1,361 (0,65)	1,373 (0,67)	-0,249 (0,77)	
leichtes Untergewicht			0,341 (0,42)	2,665 (0,42)	
Konstante	374,991 (50,76)	334,246 (61,52)	314,667 (66,88)	48,936 (8,54)	475,826 (36,53)
Angepasstes $R^2$	0,786	0,811	0,811	0,306	0,779
Ramsey Test P:	0,014	0,362	0,448	0,000	0,077
N	198	167	156	176	199

Tabelle 17.3: Sterblichkeitsmodelle mit unvollständigen Datensatz

Unterernährung in Südasien nicht so groß ist, wie es die Messungen vermuten lassen.

Ein Problem bei dieser Schätzung ist, wie bereits angemerkt, dass Beobachtungen wegfallen, die vorhandenen Informationen also nicht effizient genutzt werden. Ein weiteres Problem besteht jedoch auch darin, dass die wegfallenden Beobachtungen, wie Homogenitätstests zeigen, anders verteilt sind als die verwendeten Beobachtungen. Es treten also Schichtungseffekte auf, die zu verzerrten Schätzern führen (vgl. Little und Rubin, 1986, S. 41).

# 18 Die Ersetzung fehlender Werte

## 18.1 Klassische Verfahren

Es gibt verschiedene Verfahren, um die fehlenden Werte eines Datensatzes zu ersetzen, die jedoch mit unterschiedlichen Problemen verbunden sind (zur ausführlicheren Diskussion siehe z. B. Toutenburg, 1992).

Das einfachste Verfahren ist eine Zero-order Regression. Es besteht darin, die fehlenden Werte durch den Mittelwert der entsprechenden beobachteten Variablenwerte zu ersetzen.

Es sei  $\Phi_j$  die Menge der  $i$ , für welche die Beobachtung fehlt.

$$\Phi_j = \{i | y_{ij} \text{ fehlt}\} \quad (18.1)$$

$M_j$  ist die Anzahl der fehlenden Werte in der  $j$ -ten Variable. Der Schätzer für die fehlenden Werte lautet dann

$$\hat{y}_{ij} = \frac{1}{T - M_j} \sum_{i \notin \Phi_j} y_{ij}. \quad (18.2)$$

Die Verwendung dieses Schätzers hat zur Folge, dass die Varianz der Variable geringer ist, als jene der wirklichen Werte. Wenn jedoch eine Variable mit anderen Variablen korreliert ist oder einen Trend aufweist, ist der Mittelwert keine gute Schätzung für die fehlenden Werte.

Ein besseres Verfahren ist eine First-order Regression. Hierbei werden alle Variablen, bei denen eine Beobachtung fehlt, auf die anderen Variablen regressiert. Die fehlenden Werte werden dann ersetzt durch die Prognosen, die sich aus der Regression ergeben. Die eingesetzten Werte liegen dann jedoch alle auf der Regressionsgeraden. Dies führt dazu, dass die Korrelation der Variablen untereinander zu groß ist. Bei einer weiteren Analyse mit ökonometrischen Methoden würden Schätzer produziert, deren Signifikanz überschätzt wird.

Es wird daher häufig ein Störterm  $\mathbf{u}$  ergänzt, welcher zufällig aus einer Normalverteilung gezogen wird. Die Varianz des Störtermes wird mit Hilfe des Datensatzes geschätzt.

Ein Nachteil bleibt allerdings auch hier noch bestehen. Die Informationen, die in jenen Zeilen enthalten sind, bei denen Werte fehlen, können nicht genutzt werden. Die Regression wird auf Basis einer Complete-case Analyse durchgeführt. Das Problem der Complete-case Analyse, dass die Informationen im Datensatz nicht effizient genutzt werden, taucht auf dieser Ebene erneut auf.

## 18.2 Multiple Imputation mit Markov-Chain-Monte-Carlo Verfahren

Ein Verfahren, das von Joseph L. Schafer propagiert wird, ist die multiple Imputation mit Markov-Chain-Monte-Carlo (MCMC) Verfahren. Mit Hilfe eines Markov-Ketten-Monte-Carlo-Prozesses werden mehrere vollständige Datensätze erzeugt, die getrennt voneinander analysiert werden. Im zweiten Schritt werden die Ergebnisse dann verknüpft. Das Verfahren wird im Folgenden beschrieben<sup>1</sup>.

### 18.2.1 Annahmen

Zur Beschreibung der erforderlichen Annahmen sind einige Definitionen erforderlich. Der datengenerierende Prozess laute

$$P(Y, \theta), \quad (18.3)$$

wobei  $Y$  aus zwei Komponenten, den beobachteten und der fehlenden Werten, besteht:  $Y = \{Y_{obs}, Y_{mis}\}$ .

An welcher Position Werte fehlen, wird durch die Matrix  $R$  angezeigt. Sie hat den Eintrag 1, wenn der entsprechende Wert in  $Y$  fehlt, ansonsten den Wert 0. Der unbeobachtbare Parametervektor des Mechanismus, der zu fehlenden Werten führt, (Missing-Prozess) lautet  $\xi$ . Der Missing-Prozess sei

$$P(R|Y_{obs}, Y_{mis}, \xi). \quad (18.4)$$

Vorausgesetzt wird in diesem Verfahren, dass die Variablen gemeinsam normalverteilt sind und dass die Werte zufällig fehlen (*missing at random*, MAR). Damit ist gemeint, dass die Tatsache, ob ein bestimmter Wert fehlt oder nicht, nur von den beobachteten Werten und von den Parametern des Missing-Mechanismus abhängt, nicht aber von den

---

<sup>1</sup>Zur Diskussion und weiter gehenden Beschreibung dieses Verfahrens wird auf Schafer (1997, 1998) verwiesen

fehlenden Werten selber. Die *missing at random*-Bedingung lautet also:

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|Y_{obs}, \xi) \quad (18.5)$$

Eine weiter gehende Annahme wäre Missing completely at random (MCAR). Diese setzt voraus, dass der Missing-Prozess völlig zufällig ist, also auch nicht von  $Y_{obs}$  abhängt.

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|\xi) \quad (18.6)$$

Wenn MAR erfüllt ist und  $\theta$  und  $\xi$  distinkt sind (das heißt, in  $\theta$  sind keine Informationen über  $\xi$  enthalten, und umgekehrt) kann in den hier vorgestellten Ersetzungsverfahren der Missing-Mechanismus ignoriert werden.

## 18.2.2 Multiple Imputation

Bei den klassischen Imputationsverfahren taucht ein Problem auf, das durch die mehrfache Ersetzung der fehlenden Werte gelöst werden kann.

Werden in einem Datensatz die fehlende Werte ersetzt und wird dieser ein Datensatz dann zur Schätzung von Parametern verwendet, werden die Varianzen der Schätzer als zu gering eingeschätzt. Da nicht die wahren Werte für die fehlenden eingesetzt wurden, sondern Werte, die mit einer bestimmten Wahrscheinlichkeit von den tatsächlichen Werten abweichen, weicht auch Parameter, die in der weiteren Analyse geschätzt werden, mit einer entsprechend höheren Wahrscheinlichkeit vom wahren Wert des Parameters ab. Die Testverfahren, welche auf den geschätzten Varianzen basieren, geben damit eine zu hohe Signifikanz an.

Um die höhere Varianz der Schätzer ermitteln zu können, wird das Multiple-Imputation-Verfahren angewandt. Hierbei werden  $m$  mal die fehlenden Werte aus einer Verteilung gezogen und damit  $m$  verschiedene vervollständigte Datensätze erzeugt.

$\hat{q}_t$  (mit  $t = 1, \dots, m$ ) seien die  $m$  Schätzer auf Basis jeweils eines vervollständigten Datensatzes.  $\widehat{U}_t$  sei die Schätzung der Varianz der Schätzer.

Nach Rubin (1987) lautet dann der zusammengeführte Schätzer für den Parameter:

$$\hat{q} = \frac{1}{m} \sum_{t=1}^m \hat{q}_t \quad (18.7)$$

Die Varianz des Schätzers setzt sich zusammen aus der durchschnittlichen Varianz der einzelnen Schätzer und der Varianz zwischen den Schätzern. Die durchschnittliche Varianz ist

$$\bar{U} = \frac{1}{m} \sum_{t=1}^m \widehat{U}_t. \quad (18.8)$$

Die Varianz zwischen den Schätzern ist

$$B = \frac{1}{m-1} \sum_{t=1}^m (\hat{q}_t - \hat{q})' (\hat{q}_t - \hat{q}). \quad (18.9)$$

Die Varianz des Schätzers  $\hat{q}$  ist dann

$$\widehat{Var}(\hat{q}) = \bar{U} + \left(1 + \frac{1}{m}\right)B. \quad (18.10)$$

Da die Zahl der verwendeten Datensätze  $m$  endlich ist, wird  $B$  mit dem Faktor  $(1 + \frac{1}{m})$  gewichtet. Die angegebenen Formeln sind gültig für skalare und vektorwertige Schätzungen.

Die Effizienz einer Schätzung kann approximativ angegeben werden mit (vgl. Rubin, 1996)

$$\left(1 + \frac{\gamma}{m}\right)^{-\frac{1}{2}} \quad (18.11)$$

$\gamma$  ist der Anteil der fehlenden Werte am gesamten Datensatz,  $m$  ist die Zahl der Imputationen. Daher sind in den meisten Fällen nach Rubin (1996) 3 bis 5 Imputationen ausreichend, um befriedigende Ergebnisse zu erhalten. Im hier vorliegenden Datensatz beträgt  $\gamma$  23%. Bei 5 Imputationen ist eine Effizienz von 98% erreicht.

### 18.2.3 Markov-Chain-Monte-Carlo (MCMC) Verfahren

Die fehlenden Werte werden mit einem Markov-Chain-Monte-Carlo Verfahren (MCMC) ersetzt. Es vermeidet das Problem der Imputation mittels First-order Regression, worin (wie in der Complete-case Analyse) ein Teil der Beobachtungen nicht genutzt wird.

Das MCMC-Verfahren wurde bereits in den 50er Jahren entwickelt und von Hastings (1970) verallgemeinert. Verbreitung hat diese Methode erst in den 90er Jahren durch die Entwicklung der Computertechnologie gefunden. Die Grundgedanken werden hier kurz beschrieben. Eine tiefer gehende Diskussion findet sich z. B. in Besag et al. (1995). Eine Einführung hierzu gibt Jerak (2000)

Verwendet wird das MCMC-Verfahren vor allem für die bayesianische Schätzung von unbekanntem Parametern. Anders, als in frequentistischen Verfahren, in welchen die unbekanntem Parameter als fix aufgefasst werden, ist in bayesianischen Verfahren der unbekanntem Parameter stochastisch.

Es existiert also eine gemeinsame Verteilung der Parameter  $\theta$  und der Variablen  $y$  lautet

$$p(y, \theta) = p(y|\theta)p(\theta) \quad (18.12)$$

Die Verteilung  $p(y|\theta)$  der Variablen  $y$ , gegeben den Parametervektor  $\theta$ , wird als bekannt angenommen. Die Verteilung von  $\theta$  ist hingegen unbekannt. Es ist jedoch eine Priori-Verteilung  $p(\theta)$  bekannt, in die Vorwissen über den Parameter  $\theta$  eingeht. Gesucht ist die Posteriori-Verteilung  $p(\theta|y)$ , aus welcher die Schätzer für die gesuchten Parameter ermittelt werden können. Diese ergibt sich nach dem Bayesschen Theorem aus

$$p(\theta|y) = \frac{p(y, \theta|x)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \quad (18.13)$$

Die Posteriori-Verteilung lässt sich in den meisten Fällen weder analytisch noch numerisch finden. Mit dem Markov-Chain-Monte-Carlo Verfahren wurde eine Methode entwickelt, mit der sich die Posteriori-Verteilung approximativ finden lässt.

Eine Markov-Kette ist eine Folge von Ereignissen, bei der die Wahrscheinlichkeit für das Eintreten eines Ereignisses zum Zeitpunkt  $t$  nur davon abhängt, welches Ereignis in  $t - 1$  eingetreten ist. Von früheren Ereignissen ist sie unabhängig. Eine Einführung zu Markov-Ketten findet sich in Gnedenko (1997).

Die Idee der Markov-Ketten wird nun auf ein Monte-Carlo-Experiment angewandt, in welchem aus einer gegebenen Verteilung Zufallszahlen gezogen werden.

Eine Variante des MCMC-Verfahrens, die in dieser Arbeit Anwendung findet, ist der Gibbs-Sampler. Es wird hier kurz formal vorgestellt. Die gesuchte Posteriori-Verteilung laute  $P(\theta)$ . Es wird angenommen, der Vektor  $\theta$  bestehe aus mehreren Blöcken.

$$\theta = (\theta_1, \theta_2, \dots, \theta_K) \quad (18.14)$$

$t$  sei der aktuelle Durchlauf. Dann sind

$$\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_K^{(t)}) \quad (18.15)$$

die aktuellen Werte von  $\theta$ . Gesucht ist die unbekannte Verteilung  $p(\theta)$ . Von einem beliebigen Startwert ausgehend werden neue Werte für  $\theta$  sukzessive aus der Priori-Verteilung  $q(\theta^{(t+1)}, \theta^{(t)})$  gezogen, die nur vom aktuellen Wert von  $\theta$  abhängt.

$$\theta_1^{(t+1)} \sim q(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_K^{(t)}) \quad (18.16)$$

$$\theta_2^{(t+1)} \sim q(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_K^{(t)}) \quad (18.17)$$

$$\vdots \quad (18.18)$$

$$\theta_K^{(t+1)} \sim q(\theta_K|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{K-1}^{(t)}) \quad (18.19)$$

Nach dem Satz über Grenzwahrscheinlichkeiten bei Markov-Ketten hängt die Wahrscheinlichkeit der Ausprägung der gezogenen Zufallsvariable nach genügend Ziehungen

nicht mehr vom Anfangswert ab und nähert sich ihrem Grenzwert an. Die  $\theta^{(t)}$  bilden also eine Markov-Kette, deren Verteilung gegen die gesuchte Verteilung  $P(\theta)$  konvergiert. Nach einer Burn-In-Phase können die generierten Werte als Zufallszahlen aus der gesuchten Posteriori-Verteilung aufgefasst werden. Aus den so generierten Werten lassen sich dann Kenngrößen wie der Erwartungswert und die Varianz ermitteln.

#### 18.2.4 MCMC-Verfahren in einem Modell mit gemischten Variablen

Das MCMC-Verfahren wird nun angewandt um im vorliegenden Datensatz die fehlenden Werte zu ersetzen. Im hier verwendeten Datensatz sind sowohl kontinuierliche als auch kategoriale Variablen vorhanden. Es wird daher das Modell für gemischte Daten aus Schafer (1997) angewandt. Die kategoriale Variable bezeichnet hier das Land, aus dem die Beobachtungen kommen.

$y_{ij}$  sei der Vektor der  $j$ -ten Beobachtung aus dem Land bzw. der Region  $i$ . Der Vektor sei multivariat-normalverteilt:

$$y_{ij} \sim N(\mu_i, \Sigma), i = 1, \dots, d; j = 1, \dots, n_i; \sum_i n_i = i \quad (18.20)$$

Die Verteilung der Variablen lautet dann

$$P(y_{obs}, y_{mis} | \Theta) \quad (18.21)$$

Wobei gilt  $\Theta = (\mu_1, \dots, \mu_d, \Sigma)$  mit  $\mu_i$  als Erwartungswert und  $\Sigma$  als Kovarianzmatrix von  $y_i$ .

Es wird nun, wie in bayesianischen Verfahren üblich, eine Annahme über die Priori-Verteilung der Parameter benötigt. Für die Vektoren  $\mu$  wird angenommen, sie seien uniform-verteilt.  $\Sigma$  habe eine inverse Wishart-Verteilung. Die Wishart-Verteilung ist eine multivariate Verallgemeinerung der  $\chi^2$ -Verteilung.

Für die Posteriori-Verteilungen folgt daraus (siehe Schafer, 1997, S.347)

$$\Sigma | Y \sim W^{-1}(n - d, S^{-1}) \quad (18.22)$$

$$\mu_i | \Sigma, Y \sim N(\hat{\mu}_i, n_i^{-1} \Sigma). \quad (18.23)$$

$S$  bezeichnet die empirische Kovarianzmatrix der Residuen der  $y_{ij}$ . Es werden nun Schätzer für die fehlenden Werte aus der Verteilung gezogen. Als Startwert wurden in dieser Arbeit die Schätzer für  $\theta$  aus einem Expectation-Maximation-Verfahren verwendet. Der Imputation-Schritt lautet:

$$Y_{mis}^{(t+1)} \sim P(Y_{mis} | Y_{obs}, \theta^t) \quad (18.24)$$

## 18 Die Ersetzung fehlender Werte

Im Probability-Schritt werden dann die neuen Parameter aus der Verteilung, gegeben die beobachteten *und* die geschätzten Daten, gezogen.

$$\theta^{(t+1)} \sim P(\theta | Y_{mis}^{(t+1)}, Y_{obs}) \quad (18.25)$$

Mit den neuen Parameterschätzungen wird dieser Prozess wiederholt. Die so entstehende Markov-Kette, stabilisiert sich (unter wenig restriktiven Bedingungen) nach einer genügend großen Zahl von Iterationen. Die Verteilung von  $Y_{mis}^{(t)}, \Theta^{(t)}$  konvergiert gegen die gesuchte posteriori Verteilung  $P(Y_{mis}, \theta | y_{obs})$ . Es werden dann die fehlenden Werte  $Y_{mis}$  aus der Verteilung gezogen. Diese bilden mit  $Y_{obs}$  den ersten vervollständigten Datensatz.

## 19 Ergebnisse

Eine Reihe von Variablen im hier verwendeten Datensatz können nur Werte zwischen null und eins annehmen. Die Annahme der Normalverteilung, die dem Imputationsverfahren zugrundegelegt wird, ist bei diesen Variablen nicht erfüllt. Sie wurden für das Imputationsverfahren mit der logistischen Funktion

$$f(x) = \frac{e^{-x}}{1 + e^{-x}} \quad (19.1)$$

transformiert. Q-Q-Diagramme zeigen, dass die Normalverteilungsannahme eine vertretbare Annäherung für die transformierten Variablen darstellt. Schafer (1997) weist darauf hin, dass das Imputationsverfahren sehr robust ist und auch bei annähernder Normalverteilung noch funktioniert.

Zur Schätzung der fehlenden Werte wurde der Panel-Charakter des Datensatzes genutzt. Gibt es für ein Land zwei oder mehr Beobachtungen aus verschiedenen Jahren und fehlt beispielsweise in einem Jahr der Wert des Sozialproduktes, liegt der fehlende wahre Wert vermutlich in der Nähe der Werte, die beobachtet wurden. Um diese Informationen zu nutzen, wurde eine kategoriale Ländervariable im Imputations-Modell aufgenommen. Die Länder, für die nur eine Beobachtung vorliegt, wurden zu Regionen, wie sie in den Unterernährungs- und Sterblichkeits-Modellen verwendet wurden, zusammengefasst.

Um die fehlenden Werte des Datensatzes zu ersetzen wurden 1000 Iterationen des oben beschriebenen Prozesses durchgeführt und dann die fehlenden Werte aus der Verteilung gezogen. Mit diesem Verfahren wurden 5 vervollständigte Datensätze generiert, mit denen verschiedene Regressionen durchgeführt wurden. Die Zusammenführung der Ergebnisse ist in den Tabellen 19.1 bis 19.4 dargestellt.

In der Spalte (1) wird jeweils noch einmal das Ergebnis der Complete-Case-Analyse präsentiert, um einen besseren Vergleich zu ermöglichen. Spalte (2) zeigt die gleiche Regression mit dem vervollständigten Datensatz. In Spalte (3) wurden weitere Variablen für die Regression hinzu genommen.

In der Tabelle 19.1 finden sich die Ergebnisse des Modelles für geringes Geburtsgewicht. Der Koeffizient zur Alphabetisierungsrate hat nun anders als in Klasen(2000) das

erwartete Vorzeichen. Ist in einem Land die Qualifikation der Mutter höher, verringert dies den Anteil der Kinder mit zu geringem Geburtsgewicht. Der Einfluss der Fruchtbarkeitsrate erhöht sich deutlich. In Ländern, in denen die Fruchtbarkeitsrate größer ist, ist auch das Geburtsgewicht der Kinder geringer. Der Koeffizient  $\ln(\text{BIP})$  wird bei der Verwendung des vervollständigten Datensatzes kleiner. Der Einfluss des Sozialproduktes halbiert sich also nahezu. Entgegen den Erwartungen zeigen der Zugang zu Sanitäreinrichtungen, Trinkwasser und zu Gesundheitseinrichtungen keinen signifikanten Einfluss. Auch der Ginikoeffizient ist insignifikant. Der Koeffizient Malaria ist signifikant positiv. Eine höhere Zahl von Malariafällen erhöht den Anteil der Kinder mit geringem Geburtsgewicht.

Die Modelle zum verzögertem Wachstum (Tabelle 19.2) kommen zu ähnlichen Ergebnissen. Auch hier wechselt das Vorzeichen bei der Alphabetisierungsrate. Der Einfluss des Zugangs zu Sanitäreinrichtungen verringert sich im Modell mit dem vervollständigten Daten. In Tabelle 19.3 finden sich die Schätzung des Modells für leichtes und starkes akutes Untergewicht. Die Regression in Spalte (3) zeigt, dass der Zugang zu Gesundheitseinrichtungen einen signifikant negativen Effekt auf die Unterernährung hat.

Bei den Sterblichkeitsmodellen (Tabelle 19.4) ändert sich im Vergleich zur Complete-Case-Analyse das Vorzeichen des Dummies für die Region Osteuropa/Zentralasien. Allerdings ist deren Standardabweichung auch sehr hoch. Die Größenordnungen und die Vorzeichen der anderen Koeffizienten bleiben bei der Verwendung der vervollständigten Datensätze erhalten. Mit dem Multiple-Imputation-Verfahren können also die Ergebnisse von Klasen (2000) gestützt werden.

Ein ungewöhnliches Ergebnis ist, dass sich das Bestimmtheitsmaß bei den Schätzungen mit den vervollständigten Datensätzen verringert. Zu erwarten wäre eigentlich eine Erhöhung. Eine Verringerung des Bestimmtheitsmaß zeigt sich auch, wenn andere Imputationsverfahren, wie z.B. eine Regressionsanalyse, eingesetzt werden. Eine mögliche Ursache hierfür ist, dass das Datenmaterial aus Ländern, bei denen Werte fehlen, allgemein schlechter ist. Dort können die Messfehler besonders groß sein. Die Daten aus diesen Ländern sind aber in der Complete-Case-Analyse weggefallen. Mit dem vervollständigten Datensatz werden diese Beobachtungen in die Analyse mit eingeschlossen und können zu einer schlechteren Anpassung führen.

19 Ergebnisse

Abhängige Variable	Geringes Geburtsgewicht		
	(1)	(2)	(3)
Ostasien/Pazifik	2,753 (1,39)	2,766 (1,27)	2,643 (1,51)
Osteurope/Zentralasien		-1,351 (2,14)	-4,917 (5,01)
Mittlerer Osten/Nordafrika	1,211 (1,55)	-3,076 (1,64)	-2,506 (2,49)
Karibik	0,874 (2,66)	-1,237 (2,45)	-0,768 (2,19)
Südasien	13,305 (2,04)	13,808 (4,95)	13,490 (4,19)
Subsahara	-2,150 (1,39)	-2,439 (1,38)	-2,681 (1,88)
d9094	-0,411 (1,00)	0,169 (1,57)	0,710 (1,79)
d8589	2,150 (1,11)	1,439 (1,80)	1,618 (1,74)
pre1985	4,037 (3,99)	1,911 (1,74)	2,757 (1,89)
Bevölkerungsdichte	0,018 (0,00)	0,018 (0,01)	0,018 (0,00)
Fruchtbarkeitsrate	0,180 (0,52)	0,665 (0,46)	0,432 (0,45)
Alphabetisierung d. Frauen	0,005 (0,03)	-0,043 (0,03)	-0,038 (0,03)
ln(BIP)	-3,302 (1,20)	-1,862 (0,92)	-1,220 (0,99)
Stillen 20-23 Mon.	0,068 (0,03)	0,035 (0,03)	0,015 (0,04)
Sanitär			-0,036 (0,02)
Gesundheit			-0,023 (0,07)
Wasser			0,005 (0,05)
Gini			-0,033 (0,06)
Malaria			0,106 (0,05)
Konstante	31,378 (11,64)	23,226 (9,26)	24,03 (10,35)
N	114	273	273
$R^2$	0,760	0,627	0,660

Tabelle 19.1: Regression: Geburtsgewicht

19 Ergebnisse

Abhängige Variable	leichte + starke Wachstumsverzögerung		
	(1)	(2)	(3)
Ostasien/Pazifik	6,411 (2,58)	8,771 (2,38)	6,453 (2,48)
Osteurope/Zentralasien		-0,537 (3,99)	-3,395 (4,43)
Mittlerer Osten/Nordafrika	0,432 (2,88)	0,662 (3,42)	0,066 (4,91)
Karibik	-8,821 (3,56)	-7,537 (3,34)	-8,663 (3,38)
Südasien	9,683 (3,99)	11,965 (4,15)	9,859 (4,18)
Subsahara	-7,936 (2,57)	-6,487 (2,32)	-6,885 (2,45)
d9094	-3,737 (2,02)	-2,845 (1,91)	-2,387 (1,93)
d8589	-3,850 (2,26)	-4,681 (2,25)	-4,034 (2,18)
pre1985	-1,686 (2,39)	0,838 (2,57)	1,420 (2,66)
Bevölkerungsdichte	0,023 (0,01)	0,025 (0,01)	0,027 (0,01)
Fruchtbarkeitsrate	3,259 (0,84)	3,603 (1,01)	3,287 (0,92)
Alphabetisierung d. Frauen	0,026 (0,05)	-0,012 (0,06)	-0,025 (0,06)
ln(BIP)	-7,602 (1,74)	-9,646 (1,76)	-7,627 (2,22)
Sanitär	-0,140 (0,04)	-0,078 (0,05)	-0,060 (0,05)
Gesundheit			-0,014 (0,05)
Wasser			-0,088 (0,05)
Gini			-0,149 (0,10)
Malaria			0,124 (0,09)
Konstante	78,426 (15,60)	89,546 (17,34)	88,660 (18,73)
N	138	273	273
$R^2$	0,831	0,689	0,700

Tabelle 19.2: Regression: Wachstumsverzögerung

19 Ergebnisse

Abhängige Variable	leichtes + starkes Untergewicht		
	(1)	(2)	(3)
Ostasien/Pazifik	15,650 (1,73)	15,465 (1,90)	14,794 (1,97)
Osteurope/Zentralasien		1,375 (2,99)	-10,194 (6,23)
Mittlerer Osten/Nordafrika	0,164 (2,28)	-0,218 (2,45)	-0,084 (2,62)
Karibik	-1,294 (3,99)	-0,598 (2,77)	-0,289 (2,63)
Südasien	20,642 (4,13)	17,172 (4,46)	17,791 (3,59)
Subsahara	5,946 (2,17)	0,927 (2,14)	2,287 (2,04)
d9094	0,563 (1,53)	-1,597 (1,51)	-0,758 (1,66)
d8589	-1,136 (1,76)	-2,779 (1,70)	-2,146 (1,71)
pre1985	-0,190 (2,33)	0,469 (1,93)	1,880 (2,01)
Bevölkerungsdichte	0,022 (0,01)	0,021 (0,01)	0,021 (0,01)
Fruchtbarkeitsrate	1,704 (0,66)	2,356 (0,74)	1,879 (0,66)
Alphabetisierung d. Frauen	-0,030 (0,04)	-0,044 (0,04)	-0,035 (0,04)
ln(BIP)	-2,876 (1,49)	-4,838 (1,50)	-3,647 (1,74)
geringes Geburtsgewicht	-0,090 (0,04)	-0,052 (0,03)	0,137 (0,11)
Sanitär	0,209 (0,15)	0,242 (0,16)	-0,057 (0,04)
Gesundheit			-0,114 (0,03)
Wasser			0,017 (0,04)
Gini			-0,207 (0,07)
Malaria			0,082 (0,09)
Konstante	31,172 (13,88)	44,862 (13,12)	54,118 (14,04)
N	136	273	273
$R^2$	0,830	0,741	0,771

Tabelle 19.3: Regression: Untergewicht

19 Ergebnisse

Abhängige Variable	Mortalitätsrate bis 5j.		
	(1)	(2)	(3)
Ostasien/Pazifik	-15,504 (8,64)	-10,586 (8,93)	-20,549 (10,86)
Osteurope/Zentralasien	-0,619 (13,09)	10,600 (13,69)	-11,505 (22,56)
Mittlerer Osten/Nordafrika	-24,417 (10,10)	-12,592 (11,02)	-8,668 (20,68)
Karibik	-23,053 (12,19)	-17,208 (12,64)	-18,651 (13,64)
Südasien	-18,682 (13,98)	-10,730 (15,17)	-21,173 (16,46)
Subsahara	4,605 (9,19)	3,869 (11,00)	3,359 (11,85)
d9094	1,799 (6,59)	4,080 (7,34)	8,685 (7,55)
d8589	12,263 (7,46)	11,537 (8,35)	15,138 (8,13)
pre1985	16,007 (10,68)	21,965 (10,64)	28,536 (11,71)
Bevölkerungsdichte	0,011 (0,02)	0,003 (0,02)	0,013 (0,03)
Fruchtbarkeitsrate	8,669 (3,06)	10,262 (3,56)	8,189 (3,33)
Alphabetisierung d. Frauen	-0,923 (0,17)	-0,968 (0,18)	-0,942 (0,19)
ln(BIP)	-34,238 (5,38)	-37,220 (5,36)	-26,846 (6,31)
Sanitär			-0,111 (0,12)
Gesundheit			-0,141 (0,39)
Wasser			-0,343 (0,20)
Gini			-0,612 (0,40)
Malaria			0,601 (0,39)
Konstante	374,991 (50,76)	388,590 (54,70)	380,145 (52,96)
N	198	273	273
$R^2$	0,786	0,764	0,748

Tabelle 19.4: Regression: Kindersterblichkeit

## 20 Schlussfolgerung

In dieser Arbeit wurde ein Makro-Datensatz, in dem eine Reihe von Werten fehlten, mit einem Multiple-Imputation-Verfahren vervollständigt. Ein Vergleich von Regressionsanalysen, die mit dem vervollständigten Daten durchgeführt wurden, mit einer Complete-case-Analyse hat gezeigt, dass sich bestimmte Koeffizienten in ihrer Größenordnung geändert haben. Bei manchen Koeffizienten sind unpdlausible Vorzeichen aus der Complete-case Analyse verschwunden. Es ist also vorteilhaft, bei Problemen mit fehlenden Werten moderne Imputationsverfahren zu verwenden, welche die in den Daten vorhandenen Informationen besser ausnutzen können. Schichtungseffekte, die gegebenenfalls bei einer Complete-case-Analyse auftreten und zu verzerrten Schätzern führen, können vermieden werden. Die wesentlichen Ergebnisse aus der Untersuchung von Klasen (2000) konnten dennoch bestätigt werden.

Durch die Ersetzung der fehlenden Werte konnten noch eine Reihe von Variablen zugänglich gemacht werden, die in den bisherigen Untersuchungen nicht verwendet wurden, da dadurch auf noch mehr Beobachtungen hätte verzichtet werden müssen. So wurden noch Informationen über die Einkommensverteilung, den Zugang zu Wasser und Gesundheitseinrichtungen verfügbar. Ein Ergebnis ist, dass eine bessere Gesundheitsversorgung die akute Unterernährung eines Landes verringert.

# Literaturverzeichnis

- Besag, J., Green, P., Higdon, D. und Mengersen, K.** (1995). Bayesian computation and stochastic systems. *Statistical Science*, 10(1), 3–41.
- Drèze, J. und Sen, A.** (1995). *India – Economic Development and Social Opportunity*. Oxford University Press, Oxford.
- Gnedenko, B. W.** (1997). *Lehrbuch der Wahrscheinlichkeitstheorie*. Verlag Harri Deutsch.
- Hastings, W. K.** (1970). Monte Carlo sampling methods using Marcov chains and their application. *Biometrika*, 57, 97–109.
- Jerak, A.** (2000). Markov chain monte carlo-verfahren: Eine kurze einführung. Skriptum, Ludwig-Maximilians-Universität München, Institut für Statistik.
- Klasen, S.** (1999). Does gender inequality reduce growth and development? evidence from cross-country regressions. World Bank Policy Research Report - Working Paper No. 7.
- Klasen, S.** (2000). Malnourished and surviving in South Asia, better nourished and dying in Africa: What can explain this puzzle? SFB 386, Ludwig-Maximilians-Universität München, Discussion Paper 214.
- Little, R. J. A. und Rubin, D. B.** (1986). *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.
- Mankiw, G. N., Romer, D. und Weil, D.** (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economic*, (May), 407–437.
- Rubin, D. B.** (1987). *Multiple Imputation for Nonresponse Surveys*. J. Wilson & Sons, New York.

*Literaturverzeichnis*

- Rubin, D. B.** (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473–489.
- Schafer, J. L.** (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, Londod.
- Schafer, J. L.** (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. Techn. Ber., Pensilvania State University.
- Sen, A.** (1999). *Development as Freedom*. Oxford University Press, Oxford.
- Toutenburg, H.** (1992). *Lineare Modelle*. Physica-Verlag, Heidelberg.
- UNICEF** (1998). *The State of the World's Children 1998: Focus on Nutrition*. Oxford University Press, Oxford.
- WHO** (1983). *Measuring Change in Nutritional Status*. WHO, Geneva.

*Literaturverzeichnis*

Land	<i>Beob.</i>	Land	<i>Beob.</i>	Land	<i>Beob.</i>
Algeria	2	Gabon	1	Niger	2
Argentina	1	Gambia	4	Nigeria	2
Azerbaidjan	1	Ghana	3	Papua New Guinea	2
Bahrein	1	Guatemala	2	Pakistan	4
Bangladesh	5	Guinea	1	Panama	3
Barbados	1	Guyana	3	Paraguay	1
Benin	1	Haiti	3	Peru	4
Bolivia	9	Honduras	3	Philippines	6
Botswana	3	India	4	Romania	1
Brasil	3	Indonesia	2	Russia	2
Burkina	1	Iran	1	Rwanda	4
Burundi	1	Iraq	1	Senegal	4
Central Afr. Rep.	2	Jamaica	5	Seychelles	1
Cameroon	1	Jordan	2	Sierra Leone	3
Cape Verde	2	Kazakstan	1	South Africa	1
Chile	17	Kenya	3	Sri Lanka	3
China	2	Kyrghistan	1	Sudan	8
Colombia	5	Laos	2	Swaziland	1
Comoros	2	Lebanon	1	Syria	1
Congo	1	Lesotho	3	Tanzania	2
Costa Rica	7	Liberia	2	Thailand	1
Croatia	3	Madagascar	4	Togo	4
Czech Rep.	1	Malawi	4	Trinidad	2
Côte d'Ivoir	2	Malaysia	5	Tunisia	3
Dem. Rep. Congo	1	Mali	2	Turkey	2
Djibouti	1	Mauritania	3	Uganda	2
Dominican Rep.	2	Mauritius	2	Uruguay	3
Ecuador	1	Mexico	4	Uzbekistan	1
Egypt	6	Mongolia	1	Venezuela	4
El Salvador	3	Mozambique	2	Vietnam	3
Equatorial Guinea	1	Myanmar	6	Yemen	7
Eritrea	2	Namibia	1	Zambia	3
Ethiopia	2	Nepal	2	Zimbabwe	2
Fiji	1	Nicaragua	3		
Total					273

Tabelle 20.1: Liste der im Datensatz vorhandenen Länder

# Lebenslauf

---

4. August 1968	Geboren in Straubing
7/1983	Hauptschulabschluss, Straubing
8/1983-2/87	Berufsausbildung: Gas- und Wasserinstallateur
9/1987-7/89	Abendkurs: Mittlere Reife
9/1989-7/91	BOS-Landshut: Allgemeine Hochschulreife
8/1991-9/92	Zivildienst bei der Arbeiterwohlfahrt in Straubing
10/1992-12/1997	Universität Regensburg: Studium der Volkswirtschaftslehre
1/1998-7/1999	Wissenschaftlicher Mitarbeiter an der Universität Lüneburg, Wirtschafts- und sozialwissenschaftliche Fakultät
8/1999-1/2003	Wissenschaftlicher Mitarbeiter an der LMU München, Seminar für empirische Wirtschaftsforschung

---