Katja Landau

Messung der Vulnerabilität der Armut

Eine statistische Analyse mit deutschen Paneldaten

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

Messung der Vulnerabilität der Armut

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Messung der Vulnerabilität der Armut

Eine statistische Analyse mit deutschen Paneldaten

Dissertation
zur Erlangung des Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät
der Georg-August-Universität Göttingen

vorgelegt von Katja Landau
aus Hildesheim

Erstbetreuer	Prof. i.R. Dr. Walter Zucchini
Koreferat	Prof. Stephan Klasen, Ph.D.
Tag der mündlichen Prüfung	24.05.2012

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

1. Aufl. - Göttingen: Cuvillier, 2012
Zugl.: Göttingen, Univ., Diss., 2012

978-3-95404-258-6

© CUVILLIER VERLAG, Göttingen 2012
Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.
1. Auflage, 2012
Gedruckt auf säurefreiem Papier

978-3-95404-258-6

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Für meine Eltern

und

für meinen Mann

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin an der Professur für Statistik und Ökonometrie der Georg-August-Universität Göttingen und wurde von der Wirtschaftswissenschaftlichen Fakultät im Mai 2012 als Dissertation angenommen.

Ein besonders herzlicher Dank gilt meinem Doktorvater, Herrn Prof. Dr. Walter Zucchini, für die vertrauensvolle Zusammenarbeit in Lehre und Forschung. Er war stets an neuen Ergebnissen interessiert und unterstütze mich - auch nach seiner Emeritierung - mit wertvollen Hinweisen und regelmäßiger Gesprächsbereitschaft. Ein großer Dank gilt auch meinem Zweitprüfer Herrn Prof. Stephan Klasen Ph.D. für das Interesse an meiner Arbeit sowie die Förderung in den Bereichen der Entwicklungsökonomik und Wirtschaftspolitik. Ebenfalls danke ich Herrn Prof. Dr. Thomas Kneib für die Übernahme der Aufgabe als Drittprüfer in der Disputation.

Dem Deutschen Institut für Wirtschaftsforschung danke ich für die zur Verfügung gestellten quantitativ und qualitativ hochwertigen Daten des Sozio-oekonomischen Panels, die diese Arbeit erst ermöglicht haben. Weiterhin danke ich den Teilnehmern des EUDN PhD Workshops 2011 in Amsterdam und der PEGNet Conference 2011 in Hamburg sowie von Seminaren in Göttingen für hilfreiche Diskussionen und Kommentare. Außerdem möchte ich den Kollegen der Professur für Statistik und Ökonometrie für die gemeinsamen Jahre danken.

Besonders möchte ich meiner Familie und meinen Freunden danken, die mich durch ihre Motivation, ihr Interesse an den Fortschritten meiner Arbeit und ihr Verständnis sehr unterstützt haben. Ganz besonders dankbar bin ich meinen Eltern, die meine Entwicklung und meine Ausbildung über das normale Maß hinaus gefördert haben und mir stets mit Vertrauen und Verständnis zur Seite gestanden haben. Ein ganz besonderer Dank gilt auch meinem Mann Andreas, der mich - auch über die Entfernung hinweg - zum Durchhalten motiviert und mit Aufmerksamkeit und Geduld sowie Korrekturlesen unterstützt hat.

Göttingen im Oktober 2012 Katja Landau

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Zusammenfassung

Zur Bekämpfung der Armut ist es nicht nur wichtig, Haushalte zu identifizieren, die arm sind, sondern auch solche, die dem Risiko unterliegen, arm zu werden. In dieser Arbeit wird die Genauigkeit von Vorhersagen des Risikos gegenüber Einkommensarmut, bezeichnet mit Vulnerabilität der Armut, anhand von Querschnitts- und Paneldaten bewertet. Zur Analyse werden deutsche Paneldaten (das Sozioökonomische Panel, SOEP) herangezogen. Die Vorhersage, ob Haushalte armutsgefährdet sind oder nicht, basiert auf Regressionsmodellen mit unterschiedlichen Kovariablen (Haushaltseigenschaften bzw. Gruppen von Haushalten, stetiges Einkommen oder Einkommen in Klassen, Makrovariablen). Die Genauigkeit der Vorhersagen wird mit der Receiver Operating Characteristic (ROC) gemessen, die nicht nur den Anteil der richtig identifizierten armen Haushalte (True Positive Rate, TPR) sondern auch den Anteil der falsch als armutsgefährdet klassifizierten Haushalte (False Positive Rate, FPR) berücksichtigt. Die Schätzer mit Querschnittsdaten sind weniger genau als mit Paneldaten. Das ist auch der Fall, wenn zur Schätzung nur das Einkommen in zwei Klassen verwendet wird. Für Deutschland ist die Genauigkeit von Vulnerabilitäts-Schätzern selbst dann begrenzt, wenn Paneldaten vorliegen. Gründe dafür sind die niedrige Armutsquote und die hohe Mobilität der Haushalte in die Armut und aus der Armut heraus.

Abstract

In order to reduce poverty it is clearly of interest to identify, not only those households that are poor, but also those that are at risk of becoming poor, i.e. vulnerable to poverty. In this research, the accuracy of the ex ante assessments of vulnerability to income poverty is investigated using cross-sectional and panel data. For this purpose, long-term panel data from Germany (the German Socio-Economic Panel, SOEP) are used and different regression models are applied to classify whether a household is vulnerable or not. These models include various covariates (household covariates or groups according to household characteristics, continuous or discrete previous-year-income, macro covariates). Predictive performance is assessed using the Receiver Operating Characteristic (ROC), which takes account of true positive as well as false positive rates. Estimates based on cross-sectional data are less accurate than those based on panel data. This is true even if only imprecise information about income, i.e. classification of households into two income groups, are known. In the case of Germany, the accuracy of vulnerability predictions is limited even when panel data are used. In part this low accuracy is due to low poverty incidence and high mobility in and out of poverty.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Inhaltsverzeichnis

Inhaltsverzeichnis **IX**

Abbildungsverzeichnis **XIII**

Tabellenverzeichnis **XVII**

Abkürzungsverzeichnis **XXI**

Variablenverzeichnis **XXIII**

1 Einleitung **1**

2 Die Messung der Vulnerabilität der Armut **7**

2.1 Armut 8

2.1.1 Der Begriff Armut in Deutschland 9

2.1.2 Armutsdefinition im Kontext der Schätzung von Vulnerabilität 11

2.2 Vulnerabilität der Armut 13

2.2.1 Konzepte zur Messung 16

2.2.2 Ansätze zur Messung 21

2.3 Bewertung der Genauigkeit von Schätzern mit ROC-Kurven 30

2.4 Schätzmethoden für verschiedene Situationen der Datenverfügbarkeit 34

2.4.1 Verfügbarkeit von Querschnitts- oder Paneldaten 35

2.4.2 Verfügbarkeit von Paneldaten mit Einkommen in Klassen . . 36

2.4.3 Gruppen von Merkmalen 38

3 Datenbeschreibung **41**

3.1 SOEP 41

3.1.1 Stichprobe 42

3.1.2 Grundstruktur 44

3.1.3 Hochrechnung 46

3.1.4 Fehlende Werte 50

3.2 Generierung des Datensatzes 50

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.2.1 Datenauswahl und -beschreibung 51
3.2.2 Datenorganisation 53
3.3 Einkommensmobilität von Haushalten im SOEP 56

4 Empirische Studien zur Vulnerabilität 61
4.1 Identifizierung relevanter Kovariablen 62
4.1.1 Schätzung der Vulnerabilität für 1996 64
4.1.2 Stabilität der Schätzer im Zeitraum 1994-2008 76
4.1.3 Profile von armutsgefährdeten Haushalten 82
4.1.4 Schätzung der Vulnerabilität über n Jahre 86
4.1.5 Zwischenfazit 88
4.2 Schätzung bei ungenauer Einkommensinformation 89
4.2.1 Wahl sinnvoller Einkommensklassen 91
4.2.2 Stabilität im Zeitraum 1994-2008 104
4.2.3 Sensitivität 107
4.2.4 Zwischenfazit 112
4.3 Analyse von Verbesserungsmöglichkeiten zur Schätzung der Vulnerabilität für Deutschland 113
4.3.1 Makroökonomische Ursachen von Armut 115
4.3.2 Die Haushaltsstruktur als Ursache von Armut 129
4.3.3 Möglichkeiten zur Vulnerabilitäts-Messung mit Makrovariablen 139
4.3.4 Messung der Vulnerabilität von Gruppen 142
4.3.5 Zwischenfazit 155

5 Zusammenfassung und Ausblick 157

A Anhang 161
A.1 Daten 161
A.1.1 Datei- und Variablenbezeichnungen im SOEP 161
A.1.2 Auswahl der Daten 162
A.1.3 Darstellung der Einkommensmobilitäten als Markov-Kette . 163
A.2 Experimente zur Modellwahl zur Schätzung der Vulnerabilität . . . 164
A.2.1 Berücksichtigung der Armutshistorie 164
A.2.2 Vulnerabilitäts-Schätzung mit transformiertem Einkommen . 168
A.2.3 Vulnerabilitäts-Schätzung mit heteroskedastischen Methoden 174
A.2.4 Quantilsresiduen weiterer berücksichtigter Verteilungsannahmen 180
A.3 Ergänzungen zu den Modellen mit Klassen 182
A.3.1 Residuenplots für Modelle mit zwei Einkommensklassen . . . 182

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.3.2 Ergebnisse der Regressionen mit drei Einkommensklassen . . 183
A.4 Armut und Vulnerabilität in Gruppen 188
A.4.1 Entwicklung von Anteilen und Armut nach Gewerbe des Haushaltsvorstands 188
A.4.2 Kovariablen zur Vulnerabilitäts-Schätzung in Gruppen für 2005 189
A.4.3 Modellwahl für die Schätzung in Gruppen 191
A.4.4 Schätzung der Vulnerabilität für das Geschlecht für 2005 . . 194
A.4.5 Historische VPLs für die Gruppen Erwerbstätigkeit und Anzahl der Kinder im Haushalt 198
A.4.6 Stabilität der Vulnerabilitäts-Schätzer in den Gruppen Anzahl der Kinder 200

Literaturverzeichnis 203

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Abbildungsverzeichnis

2.1 Relative und absolute Armutslinien (1992-2008) 11
2.2 Entwicklung zugehöriger Armutsquoten (1992-2008) 12
2.3 Entwicklung der Einkommensungleichheit in Deutschland (1992-2008) 13
2.4 Vulnerabilitätskonzept von VEU 19
2.5 Mögliche Modifizierung der Methode von Kamanou/Modurch (2002) 25
2.6 Kontigenztabelle der Diagnoseergebnisse 32
2.7 Qualitätsmaße für die ROC-Kurven 33

3.1 Entwicklung der Fallzahlen privater Haushalte (1992-2009) 43
3.2 Datenstruktur des SOEPs im Querschnitt 44
3.3 Unterstützung durch SOEPinfo 54
3.4 *Derzeitige* Arme zerlegt nach Einkommensklassen des *Vorjahres* (1993-2008) 57
3.5 *Derzeit* nicht-arme Haushalte zerlegt nach den Einkommensklassen des Vorjahres (1993-2008) 58
3.6 Übergangswahrscheinlichkeit zur Armut im nächsten Jahr (1993-2008) 58

4.1 Veranschaulichung von Multikollineariät im Venn-Diagramm 70
4.2 ROC-Kurven für sechs verschiedene Vulnerabilitäts-Schätzungen für 1996 73
4.3 ROC-Kurven erzeugt durch Modell P_1 (1994-2008) 77
4.4 ROC-Kurven erzeugt durch Modell P_6 (1994-2008) 77
4.5 Historische VPLs bei $TPR = 80\%$ und 90% erzeugt durch Modell P_1 (1994-2008) 78
4.6 Historische VPLs bei $TPR = 80\%$ und 90% erzeugt durch Modell P_6 (1994-2008) 78
4.7 Genauigkeit der Vulnerabilität im Zeitraum 1994-2008 (Niveau etwa bei 80% TPR) 80
4.8 Genauigkeit der Vulnerabilität im Zeitraum 1994-2008 (Niveau etwa bei 90% TPR) 81

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.9 Anteil der als armutsgefährdet klassifizierten Haushalte nach Familienstand unter Verwendung von P_1 bei $TPR = 80\%$ in % (1994-2008) 84
4.10 Anteil der als armutsgefährdet klassifizierten Haushalte nach Familienstand unter Verwendung von P_6 bei $TPR = 80\%$ in % (1994-2008) 84
4.11 ROC-Kurven für die n-Jahres-Vulnerabilität für n=1, 2, 3, 4, erzeugt durch Modell P_1 (1996-1999) 87
4.12 ROC-Kurven für die n-Jahres-Vulnerabilität für n=1, 2, 3, 4, erzeugt durch Modell P_6 (1996-1999) 87
4.13 Performance für zwei Einkommensklassen für 2006 98
4.14 Vergleich der Genauigkeit von Schätzern von M_{12}, P_1 und P_6 für 2006 98
4.15 Vergleich der Genauigkeit von Schätzern von M_{22},P_1 und P_6 für 2006 99
4.16 Vergleich der Genauigkeit der Vulnerabilitäts-Schätzer mit zwei und drei Einkommensklassen (Methode M_{1K}) 101
4.17 Vergleich der Genauigkeit der Vulnerabilitäts-Schätzer mit zwei und drei Einkommensklassen (Methode M_{2K}) 101
4.18 False positives in einer Stichprobe von 1 000 Haushalten (Schätzung mit $M_{12,06}$) 103
4.19 False positives in einer Stichprobe von 1 000 Haushalten (Schätzung mit $M_{22,06}$) 103
4.20 VPL von drei Modellen bei 80% TPR (1994-2008) 104
4.21 Stabilität der Vulnerabilitäts-Schätzer von M_{1K} (1994-2008) 105
4.22 Stabilität der Vulnerabilitäts-Schätzer von M_{2K} (1994-2008) 106
4.23 Verteilung des generierten Einkommens bei Annahme einer Gleichverteilung. 109
4.24 Dichte des generierten Einkommens für $y_{h,t}$ bei Annahme einer Normalverteilung. 109
4.25 Auswirkung eines Fehlers auf die Genauigkeit der Vulnerabilitäts-Schätzung in 2001 111
4.26 Arbeitslosenquote (gemessen an den abhängigen zivilen Erwerbspersonen) (1992-2008) 116
4.27 Anzahl der Arbeitslosenhilfe/Hartz IV-Bezieher (1992-2008) 119
4.28 Anzahl der sozialversicherungspflichtig Beschäftigten (1992-2008) . . 120
4.29 Lohnquote (1992-2008) 121
4.30 Veränderung der preisbereinigten Nettolöhne und Gehälter je Stunde zum Vorjahr in % (1992-2008) 122
4.31 Abgabenquote (1992-2008) (Abgrenzung der Volkswirtschaftlichen Gesamtrechnung (VGR)) 124

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.32 Veränderung des Wirtschaftswachstums in Deutschland zum Vorjahr in % (1992-2008) 125
4.33 Renditen Dax-Schlussstände (1992-2008) 125
4.34 Armut im Jahr t in Abhängigkeit von Indikatoren des Arbeitsmarktes des Vorjahres (1992-2008) 127
4.35 Armut im Jahr t in Abhängigkeit von weiteren volkswirtschaftlichen Indikatoren des Vorjahres (1992-2008) 128
4.36 Entwicklung der Haushaltsstruktur nach verschiedenen Merkmalen (1992-2008) 130
4.37 *Fortsetzung* Entwicklung der Haushaltsstruktur nach verschiedenen Merkmalen (1992-2008) 131
4.38 Entwicklung der Armut nach verschiedenen Merkmalen (1992-2008) 133
4.39 *Fortsetzung 1* Entwicklung der Armut nach verschiedenen Merkmalen 1992-2008 134
4.40 *Fortsetzung 2* Entwicklung der Armut nach verschiedenen Merkmalen (1992-2008) 135
4.41 Erklärungsgehalt des Einkommens im Jahr t durch Arbeitsmarktindikatoren des Vorjahres für verschiedene Quantile (1992-2008) . . . 140
4.42 *Fortsetzung* Erklärungsgehalt des Einkommens im Jahr t durch Arbeitsmarktindikatoren des Vorjahres für verschiedene Quantile (1992-2008) 141
4.43 Genauigkeit der Schätzung der Vulnerabilität getrennt nach Merkmalen (1994-2008) 145
4.44 *Fortsetzung 1* Genauigkeit der Schätzung der Vulnerabilität getrennt nach Merkmalen (1994-2008) 146
4.45 *Fortsetzung 2* Genauigkeit der Schätzung der Vulnerabilität getrennt nach Merkmalen (1994-2008) 147
4.46 Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Erwerbstätigkeit für 1994-2008 (Methode 1) 153
4.47 Gesamtbetrachtung der Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Anzahl der Kinder im Haushalt für 1994-2008 (Methode 2) 154

B.1 ROC-Kurven von Vulnerabilitäts-Schätzern mit Armutshistorie (1999-2007) 168
B.2 Residuen der Regressionen mit und ohne log Transformation 173
B.3 ROC-Kurven der Schätzungen mit und ohne log Transformation des Einkommens (1996) 174

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

B.4 ROC für Schätzungen von Modellen mit heteroskedastischer und homoskedastischer Varianz 180
B.5 Quantilsresiduen von Regressionen unter Annahme einer t-Verteilung 180
B.6 Quantilsresiduen von Regressionen mit verschiedenen Verteilungsannahmen 181
C.7 Quantilsresiduen für Regressionen mit zwei Einkommensklassen . . 182
C.8 Quantilsresiduen für Regressionen mit drei Einkommensklassen . . . 187
D.9 Residuen der Regressionen für das Merkmal Geschlecht 192
D.10 Residuen der Regressionen mit transformiertem Einkommen für das Merkmal Geschlecht 193
D.11 Historische VPLs für Schätzungen nach Merkmal Erwerbstätigkeit für den Zeitraum 1994-2008 (Methode 1) 198
D.12 Historische VPLs für Schätzungen der Vulnerabilität nach Merkmal Anzahl der Kinder mit $P_{1,G}$ für den Zeitraum 1994-2008 199
D.13 Historische VPLs für Schätzungen der Vulnerabilität nach Merkmal Anzahl der Kinder mit $P_{6,G}$ für den Zeitraum 1994-2008 200
D.14 Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Anzahl der Kinder im Haushalt mit Methode 2 und $P_{1,G}$ 201
D.15 Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Anzahl der Kinder im Haushalt mit Methode 2 und $P_{6,G}$ 202

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Tabellenverzeichnis

3.1 Stichproben des SOEPs 42
3.2 Auswahl von Variablen 52

4.1 Beschreibung der Variablen in der Regression. 66
4.2 Ergebnisse der Regression für Modell $P_{1,96}$ 67
4.3 Ergebnisse der Regression für Modell $P_{4,96}$ 68
4.4 Ergebnisse der Regression für Modell $P_{6,96}$ 69
4.5 FPR und VPL für TPR = 0.8 und TPR = 0.9 in 1996 74
4.6 True und false positives in einer hypothetischen Population von 1000 Haushalten für die Modelle P_1-P_6 und für drei verschiedene Armutsquoten 75
4.7 Anteil der als armutsgefährdet klassifizierten Haushalte nach Bundesland unter Verwendung von P_1 bei $TPR = 80\%$ in % (1994-2008) 83
4.8 Anteil der als armutsgefährdet klassifizierten Haushalte nach Bundesland unter Verwendung von P_6 bei $TPR = 80\%$ in % (1994-2008) 83
4.9 Anteil der als armutsgefährdet klassifizierten Haushalte nach Gewerbe unter Verwendung von P_1 bei $TPR = 80\%$ in % (1994-2008) 85
4.10 Anteil der als armutsgefährdet klassifizierten Haushalte nach Gewerbe unter Verwendung von P_6 bei $TPR = 80\%$ in % (1994-2008) 85
4.11 FPR und VPL für $TPR = 0.8$ für 1 bis 4 Jahre in der Zukunft . . . 86
4.12 Beschreibung der Einkommensvariablen in den Regressionen $M_{12,06}$-$M_{22,06}$ 91
4.13 Regression für Modell $M_{12,06}$ mit $L = 1.3z$ 92
4.14 Regression für Modell $M_{12,06}$ mit $L = 1.7z$ 93
4.15 Regression für Modell $M_{22,06}$ mit $L = 1.3z$ (untere Einkommensgruppe) 94
4.16 Regression für Modell $M_{22,06}$ mit $L = 1.3z$ (obere Einkommensgruppe) 95
4.17 Einkommensverteilung der Haushalte in 2000 110
4.18 Entwicklung der Steuersätze (1992-2008) 122
4.19 Gruppierungen nach Haushaltseigenschaften 129

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.20 Genauigkeit der Schätzung der Vulnerabilität mit Methode 2 für 2005 in % 150
4.21 *Fortsetzung* Genauigkeit der Schätzung der Vulnerabilität mit Methode 2 für 2005 in % 151
4.22 Vergleich der Genauigkeit der Vulnerabilitäts-Schätzer von Methode 1, Methode 2 (Gesamtbetrachtung) sowie P_1 und P_6 für 2005 in % . 152

A.1 Bezeichnung ausgewählter Variablen und Speicherort im SOEP . . . 162
B.2 Ergebnisse der Regression für Modell P_6 mit einem Panel von 1999-2007 165
B.3 Ergebnisse der Regression für Modell $P_{6,D}$ mit einem Panel von 1999-2007 166
B.4 Ergebnisse der Regression für Modell $P_{6,anz5}$ mit einem Panel von 1999-2007 167
B.5 Ergebnisse der Regression für Modell $\mathrm{P}_{1log,96}$ 170
B.6 Ergebnisse der Regression für Modell $\mathrm{P}_{4log,96}$ 171
B.7 Ergebnisse der Regression für Modell $\mathrm{P}_{6log,96}$ 172
B.8 Regression für Modell $P_{1,heteroskedastisch}$ mit heteroskedastischer Varianz (Erwartungswert) 176
B.9 Regression für Modell $P_{1,heteroskedastisch}$ mit heteroskedastischer Varianz (Varianz) 177
B.10 Regression für Modell $P_{6,heteroskedastisch}$ mit heteroskedastischer Varianz (Erwartungswert) 178
B.11 Regression für Modell $P_{6,heteroskedastisch}$ mit heteroskedastischer Varianz (Varianz) 179
C.12 Regression für Modell $M_{13,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ 183
C.13 Regression für Modell $M_{23,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ (untere Einkommensgruppe) 184
C.14 Regression für Modell $M_{23,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ (mittlere Einkommensgruppe) 185
C.15 Regression für Modell $M_{23,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ (obere Einkommensgruppe) 186
D.16 Anteile der Haushalte nach Beruf des Haushaltsvorstands in % (1992-2008) 188
D.17 Bedingte Armutsquoten der Haushalte nach Berufsgruppe des Haushaltsvorstands in % (1992-2008) 188
D.18 Anteil der Armut von Berufsgruppen in % (1992-2008) 189

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

D.19 Besonderheiten der Kovariablen in den Regressionen für einzelne Merkmale 191
D.20 Ergebnisse der Regression für Modell $P_{1,\text{männlich}}$ 194
D.21 Ergebnisse der Regression für Modell $P_{1,weiblich}$ 195
D.22 Ergebnisse der Regression für Modell $P_{6,\text{männlich}}$ 196
D.23 Ergebnisse der Regression für Modell $P_{6,weiblich}$ 197

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Abkürzungsverzeichnis

AIC	Akaike Information Criterion
AUC	Area under the Curve
BIP	Bruttoinlandsprodukt
BMF	Bundesministerium der Finanzen
BMFSFJ	Bundesministerium für Familie, Senioren, Frauen und Jugend
BRD	Bundesrepublik Deutschland
CNEF	Cross-National Equivalent File
DAX	Deutscher Aktien Index
DDR	Deutsche Demokratische Republik
DIW	Deutsches Institut für Wirtschaftsforschung
DTC	Desktop Companion to the German Socio-Economic Panel
FGLS	Feasible Generalized Least Square Method
FGT	Foster-Greer-Thorbecke-Maß
FHSR	Fachhochschulreife
FPR	False Positive Rate
GRV	Gesetzliche Rentenversicherung
HBS	Household Budget Survey
HV	Haushaltsvorstand
IAB	Institut für Arbeitsmarkt- und Berufsforschung
ifo	Institut für Wirtschaftsforschung
INR	item-non-response
MCB	Monte Carlo Bootstrap Methode
MSE	Mean Squared Error
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
PAUC	Partial Area under the Curve
PRSP	Poverty Reduction Strategy Papers

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ROC	Receiver Operating Characteristic
SOEP	Sozio-oekonomisches Panel
TPR	True Positive Rate
UNR	unit-non-response
VEP	Vulnerability as Expected Poverty
VER	Vulnerability as Exposure to Risk
VEU	Vulnerability as Low Expected Utility
VGR	Volkswirtschaftliche Gesamtrechnung
VPL	Vulnerability to Poverty Line
WMS	Welfare Monitoring Survey

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Variablenverzeichnis

Lateinische Buchstaben

Variable	Bedeutung
A_t	Binäre Variable des erfolgreichen Kontakts eines Haushalts in t
B	Anzahl der Bootstrap-Stichproben
C_h	Downside certainties
$C_{t,w}$	Binäre Variable der erfolgreichen Auswahl eines Haushalts über w Wellen beginnend ab Welle t
D	Binäre Variable der Auswahl eines Haushalts für die Stichprobe in Welle 1
$d_{h,i}$	Deprivation Index von Haushalt h in Zustand i
$E(\cdot)$	Erwartungswert
$e_{h,t}$	Zufälliger Fehlerterm
$F(\cdot)$	Verteilungsfunktion der Wohlfahrt
$f(\cdot)$	Dichtefunktion der Wohlfahrt
$g(\cdot)$	Funktion der Konsumdifferenzen
$g_i(\cdot)$	Linkfunktion
$H_t(\cdot)$	Anzahl der armen Haushalte zum Zeitpunkt t
L_k	Einkommens-Klassengrenze
N	Anzahl der Haushalte in einer Stichprobe
N_{gew}	Hochgerechnete Stichprobe
n	Vorhersagehorizont der Vulnerabilität der Armut
$P(\cdot)$	Wahrscheinlichkeitsfunktion
p_h	Wahrscheinlichkeit der Armut eines Haushalts h
R_t	Binäre Variable des Antwortverhaltens zum Zeitpunkt t
$r_{h,t}$	Zufälliger Fehleranteil
r_U	Arrow-Pratt Maß der absoluten Risikoaversion

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$s_{h,i,j}$		Intensität downside risks
$s_{h,k}$		Intensität downside certainties
$T_{t,t+1}$		Matrix der Übergangswahrscheinlichkeiten von t nach $t+1$
$U_{h,t}$	oder U	Nutzenfunktion eines Haushalts h zum Zeitpunkt t
$Var(\cdot)$		Varianz
$V_{h,t}$	oder V_t	Vulnerabilitätsmaß für h in t
$X_{h,t}$		Haushaltskovariablen für Haushalt h in t
$X^k_{h,t}$		Haushaltskovariablen in der k-ten Einkommensklasse für Haushalt h in t
$\overline{X}_t$		Aggregierte Haushaltskovariablen
$y_{h,t}$	oder y_t	Stetiges Haushalts-Netto-Äquivalenzeinkommen von h in t (kurz *Einkommen* bzw. in anderen Studien Konsum per capita)
$y_{h,t,F}$	oder $y_{t,F}$	Verzerrtes Einkommen von h in t
$y_{h,t,G}$	oder $y_{t,G}$	Stetiges Einkommen von Gruppen von Haushalten in t
$y_{h,t,K}$	oder $y_{t,K}$	(Diskretes) Einkommen in K Einkommensklassen
$Z_{h,t}$		Binäre Variable für die Armut von h in t
z		Armutsgrenze
z_{CE}		Sicherheitsäquivalent

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Griechische Buchstaben

Variable	Bedeutung
α	Tiefe der Armut (FGT-Indizes)
β	Regressionskoeffizienten für die Haushaltskovariablen
β_G	Gruppen-Regressionskoeffizienten für die Haushaltskovariablen
γ	Regressionskoeffizient für das Vorjahreseinkommen
γ_G	Gruppen-Regressionskoeffizient für das Vorjahreseinkommen
δ	Relative Konsumveränderungen von h
η	Parameter der Risikoaversion
θ	Regressionskoeffizienten zur Anpassung der Varianz
$\iota(\cdot)$	Foster-Greer-Thorbecke-Armutsindizes
κ	Parameter der Risikoaversion
λ	Parameter der Risikoaversion
ν	Haushaltsspezifische Effekte
ξ	Zeitspezifische Effekte
Σ	Kovarianzmatrix
σ_e^2	Standardfehler
ς_N	Streuungsparameter der Normalverteilung
ς_U	Streuungsparameter der Gleichverteilung
τ	Quantil
$\Phi(\cdot)$	Verteilungsfunktion der Standardnormalverteilung
ψ	Regressionskoeffizienten für Makrovariablen
Ω	Zustandsraum der Umweltzustände
ω	Umweltzustand

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

1 Einleitung

Die Bekämpfung der Armut stellt eines der wichtigsten Ziele der internationalen Gemeinschaft dar. Erste Schritte zur gemeinsamen Bewältigung dieser Aufgabe wurden auf dem Weltsozialgipfel 1995 in Kopenhagen unternommen. In der Kopenhagener Erklärung (United Nations, 1995) verpflichteten sich die Regierungschefs aller teilnehmenden Länder, jegliche Form der Armut national zu messen und politische Strategien zu ihrer Beseitigung in schriftlicher Form festzuhalten. Weiterhin erklärten sie sich dazu bereit, anderen Ländern durch eine stärkere internationale Zusammenarbeit bei der Bewältigung von Armut behilflich zu sein. Auf dem Milleniumsgipfel 2000 in New York wurden die Maßnahmen durch Aktions- und Zeitvorgaben konkretisiert. Als oberstes Ziel wurde die Halbierung der extremen Armut bis 2015 beschlossen (United Nations General Assembly, 2000).

Der Weltentwicklungsbericht (World Bank, 2000/01) beschreibt drei Strategien zur Bekämpfung der Armut. Diese liegen darin „Möglichkeiten zu fördern“ (z.B. durch Aufstockung des Kapitals der Armen, Beseitigung ungleicher Verteilung in Gruppen, Vermittlung von Wissen), „Mitwirkungsmöglichkeiten“ zu schaffen (z.B. durch Dezentralisierung und Förderung der Einbeziehung der Bürger, Beseitigung sozialer Hemmnisse, Stärkung des Sozialkapitals der Armen) und die „Sicherheit“ gegenüber Risiken zu verbessern (z.B. durch Programme zur Verhinderung und Bewältigung von Krisen auf den Finanzmärkten, Naturkatastrophen, und sozialen Risiken) (World Bank, 2000/01, S. 6-12). Das Risiko, in der Zukunft in Armut zu fallen, wird in dem Bericht mit dem Begriff *Vulnerabilität der Armut* (im Englischen vulnerability to poverty) bezeichnet. Im Folgenden wird der Begriff kurz mit der *Vulnerabilität* bezeichnet.

Zur Verhinderung und Bewältigung von Armutsrisiken ist es von großer Bedeutung, die Haushalte zu identifizieren, die in der Zukunft dem Risiko unterliegen, in Armut zu fallen, d.h. zu berechnen, welche Haushalte *armutsgefährdet* (im Englischen vulnerable to poverty) sind. Seitdem der Begriff „Vulnerabilität“ vom Weltentwicklungsbericht 2000/01 in den Mittelpunkt des Interesses gerückt wurde, haben Forscher verschiedene axiomatische Methoden vorgeschlagen, um das Konzept zu

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

konkretisieren, und Ansätze entwickelt, die Vulnerabilität empirisch zu schätzen. Idealerweise sollten zu ihrer Schätzung Paneldaten, d.h. Daten, die die Lebensläufe von Haushalten über mehrere Jahre aufzeichnen, verwendet werden. Nur solche Art von Daten erfassen die Risiken, denen die Haushalte ausgesetzt sind. Trotzdem verwenden viele Studien für die Schätzung Querschnittsdaten (z.B. Chaudhuri (2002), Günther und Harttgen (2009)). Der Grund dafür ist, dass Studien über die Vulnerabilität der Armut bisher hauptsächlich für Entwicklungsländer durchgeführt wurden und in diesen Paneldaten selten verfügbar sind.

Die Literatur über die Vulnerabilität vernachlässigt bisher, wie genau die Vulnerabilität geschätzt werden kann und welche Definition von Genauigkeit in diesem Kontext überhaupt sinnvoll ist. Die vorliegende Arbeit soll einen Beitrag zur Genauigkeit der Schätzung der Vulnerabilität der Armut mit vorhandenen Methoden in verschiedenen Situationen der Datenverfügbarkeit leisten. Das Vorgehen zur Messung der Vulnerabilität ist wie in Landau et al. (2012).

In dieser Arbeit wird die Vulnerabilität deutscher Haushalte basierend auf dem Wohlfahrtsindikator des Jahreseinkommens untersucht. Da die Einkommen von Ein- und Mehrpersonenhaushalte nicht miteinander vergleichbar sind, genügt es jedoch nicht, das dokumentierte Jahreseinkommen zu verwenden. Stattdessen soll das Äquivalenzeinkommen mit den Äquivalenzgewichten der neuen OECD-Äquivalenzskala (per adult equivalent), das jeder Person eines Haushalts zugeordnet wird, Anwendung finden (Atkinson et al., 2002, S. 84 & S. 98ff.). Da in dieser Arbeit jedoch Haushalte analysiert werden und sich die meisten der betrachteten Eigenschaften auf den Haushalt beziehen, wird die Wohlfahrt eines Haushalts durch sein Äquivalenzeinkommen beschrieben, d.h. jedem Haushalt wird nur *ein* Äquivalenzeinkommen zugeordnet.

Als Datengrundlage wird das deutsche Sozio-oekonomische Panel (SOEP) (genauer „Soepv26“) verwendet. Der Vorteil gegenüber Daten von Entwicklungsländern besteht darin, dass sie von hoher Qualität und über einen langen Zeithorizont, nämlich 26 Jahre, verfügbar sind. Diese Arbeit stützt sich auf die Beobachtungen der Jahre 1992-2009. Das SOEP beinhaltet verständlich aufbereitete und relativ vollständige Informationen zu Haushaltseigenschaften und Einkommen. Die meisten der in dieser Arbeit verwendeten Daten sind im Cross-National Equivalent File (CNEF) harmonisiert und unterliegen damit einem besonderen Gütesiegel. In dieser Arbeit wird angenommen, dass sich die Haupterkenntnisse der deutschen Daten auch auf andere Länder übertragen lassen. Die Daten ermöglichen es, die Vulnerabilität mit Querschnittsdaten und Paneldaten verschiedener Länge für mehrere

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Jahre zu schätzen und retrospektiv zu überprüfen, welche, der als armutsgefährdet geschätzten Haushalte, tatsächlich arm geworden sind.

Ein guter Schätzer sollte natürlich einen hohen Anteil der Haushalte identifizieren, die in einem vorher festgelegten Zeitintervall, z.B. dem nächsten Jahr, arm werden oder arm bleiben, unter der Bedingung, dass keine Maßnahmen der Intervention der Armut getroffen wurden. Dieses Kriterium kann leicht erfüllt werden, indem einfachhalber alle Haushalte als armutsgefährdet deklariert werden mit der Begründung, dass „selbst Millionäre verwundbar für Armut sind“ (Pritchett et al., 2000). Dieses Kriterium allein ist jedoch nicht sinnvoll, da sonst Ressourcen für Haushalte ausgegeben werden, die weit entfernt davon sind, in Armut zu fallen. Aus diesem Grund muss ein guter Schätzer der Vulnerabilität nicht nur die Haushalte schätzen, die tatsächlich arm werden, sondern auch diejenigen, *die tatsächlich nicht-arm werden*. Die Receiver Operating Characteristic (ROC) ist eine etablierte, statistische Methode, um die Güte solcher Diagnose-Probleme zu messen. Sie zieht nicht nur den Anteil der Haushalte in Betracht, die richtig als armutsgefährdet identifiziert wurden (TPR, True Positive Rate) sondern auch den Anteil, der fälschlicherweise als armutsgefährdet deklariert wurde (FPR, False Positive Rate).

Die Schätzer werden mit verschiedenen Regressionsmodellen erzeugt. Im ersten und dritten Teil der empirischen Untersuchung wird das Einkommen mit linearen Regressionsmodellen für die nächste Periode prognostiziert. Aufgrund der Festlegung einer Vulnerabilitätsgrenze (VPL) nach bestimmten Kriterien wird ein Haushalt als armutsgefährdet oder nicht-armutsgefährdet klassifiziert. Es ist *nicht* das Ziel der Prognose das exakte Einkommen in der nächsten Periode vorauszusagen, sondern abzuschätzen, ob Haushalte oberhalb oder unterhalb der festgelegten VPL liegen. Je höher die VPL gewählt wird, desto mehr der armen Haushalte werden richtig klassifiziert, da dann auch arme Haushalte mit größerem Vorhersagefehler des Einkommens unter die VPL fallen. Gleichzeitig rutschen natürlich auch mehr nicht-arme Haushalte (mit kleinerem Vorhersagefehler des Einkommens) unter die VPL, so dass auch ein höherer Anteil der nicht-armen Haushalte falsch klassifiziert wird.

Im zweiten Teil der empirischen Untersuchung werden Haushalte aufgrund einer, durch ein logistisches Regressionsmodell geschätzen, Wahrscheinlichkeit als armutsgefährdet oder nicht-armutsgefährdet klassifiziert. Die Schätzung der Vulnerabilität über das Einkommen und die Wahrscheinlichkeit sind äquivalent und führen zu denselben Ergebnissen. Denn unabhängig davon, ob die Wahrscheinlichkeit oder das Einkommen geschätzt wird, muss eine Grenze gewählt werden, die Haushalte

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

als armutsgefährdet oder nicht armutsgefährdet klassifiziert. Bei beiden Vorgehensweisen wird mit einem Haushalt, dessen Schätzer nah an der Armutsgrenze liegt, auch ein höheres Risiko verbunden, in Armut zu fallen als mit einem Haushalt, der nah an der VPL liegt. Eine kleine Verschiebung der VPL nach unten würde dazu führen, dass dieser Haushalt nicht mehr als armutsgefährdet klassifiziert wird.

Mit diesen Werkzeugen sollen in dieser Arbeit empirisch untersucht werden,

1. welche Kovariablen (Haushaltseigenschaften, Vorjahreseinkommen) die Genauigkeit von Vulnerabilitäts-Schätzern erhöhen,
2. ob und inwieweit Schätzer unter Verwendung der Kovariable Vorjahres-Einkommen in Einkommensklassen Schätzer, basierend auf Querschnittsdaten, verbessern und solche, basierend auf Paneldaten und stetigem Einkommen, verschlechtern,
3. ob sich die Schätzer der Vulnerabilität für Deutschland verbessern lassen, wenn Makrovariablen (z.B. Arbeitsmarkt, Steuern, Wirtschaft, Finanzmarkt) zur Schätzung hinzugezogen werden oder wenn die Vulnerabilität für verschiedene Merkmalsgruppen separat geschätzt wird.

Die Ergebnisse der ersten Studie können dazu genutzt werden, für Querschnittsdaten oder Paneldaten anzugeben, wie genau die armutsgefährdeten Haushalte vorhergesagt werden können, wenn *ideale* Daten vorhanden sind. Es zeigt sich, dass die Genauigkeit von Schätzern, die auf Paneldaten basieren, deutlich höher ist als solche von Querschnittsdaten. Jedoch sind selbst die Schätzer basierend auf Paneldaten nicht besonders genau. Um viele der zukünftigen armen Haushalte zu identifizieren, müssen sehr viele zukünftig nicht-arme Haushalte als armutsgefährdet klassifiziert werden. Ursachen dafür liegen in der niedrigen Armutsquote in Deutschland sowie der Mobilität der Haushalte in die Armut und aus der Armut heraus. Die Messung der Genauigkeit von Vulnerabilität der Armut stellt eine wertvolle Erweiterung der Forschung dar, die sich bisher hauptsächlich auf Ansätze zur Schätzung der Vulnerabilität stützt. Sie sagt nicht nur aus, welche Haushalte armutsgefährdet sind, sondern auch mit welcher Wahrscheinlichkeit diese Haushalte arm bzw. nicht-arm werden, d.h., wie gut die Schätzer die zukünftig armen Haushalte vorhersagen können. Diese Information ist für Politiker bei Entscheidungen über die Verteilung von Ressourcen und Maßnahmen zur Prävention von Armut von großer Bedeutung (Landau et al., 2012).

Die zweite Studie trifft Aussagen über die Genauigkeit, wenn weniger gute Daten in Bezug auf das Einkommen, d.h. Einkommen in wenigen Einkommensklassen und mit Fehlern bei der Zuordnung der Haushalte zu den Klassen, vorliegen. Die-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

se Analyse liefert, insbesondere für die Schätzung der Vulnerabilität in Entwicklungsländern, in denen Daten über Einkommen kaum oder nur ungenau vorhanden sind, wichtige Informationen. Die Analysen zeigen, dass, im Falle von vorhandenen Einkommensdaten, diese die Schätzung basierend auf Haushaltsvariablen auch dann verbessern, wenn die Qualität der Einkommensdaten nicht so hoch ist. Falls keine Einkommensvariablen vorhanden sind, bleibt noch die Möglichkeit, Einkommen durch einen anderen Wohlfahrtsindikator, z.B. den Konsum, zu ersetzen.

Die dritte Studie dient dagegen der Messung der Vulnerabilität konkret für Deutschland. Es ist anzunehmen, dass volkswirtschaftliche Risiken (z.B. Anstieg der Arbeitslosigkeit, Finanzkrisen) die Vulnerabilität der Haushalte beeinflussen. Weiterhin könnten Veränderungen der Haushaltsstruktur (z.B. Anstieg der alleinstehenden und alleinerziehenden Haushalte) in der Stichprobe zu einer Veränderung der Vulnerabilität führen. Es werden Möglichkeiten zur Verbesserung der Schätzer unter Berücksichtigung dieser beiden Aspekte analysiert. Es zeigt sich, dass die Schätzung der Vulnerabilität mit Makrovariablen und den verwendeten Methoden Regressionen und Quantilsregressionen nicht verbessert werden kann. Makrovariablen können bislang nur zur Erklärung der Armut ex post verwendet werden. Durch die Einbeziehung des Strukturwandels der Haushalte kann die Genauigkeit durch Schätzung der Vulnerabilität in Gruppen geringfügig verbessert werden.

Die Arbeit ist im weiteren Verlauf wie folgt aufgebaut. In Kapitel 2 werden die Grundlagen für diese Arbeit beschrieben. Diese beinhalten die verwendete Armutsdefinition, den Forschungsstand zum Thema Vulnerabilität der Armut, die Beschreibung der Bewertung von Schätzern mit ROC sowie die angewandten Schätzmethoden im empirischen Teil in Kapitel 4.

Kapitel 3 zeigt die Grundstruktur des SOEP-Datensatzes auf, die Voraussetzung für die Generierung des Datensatzes dieser Arbeit ist. Detailliert werden die Variablen sowie die Generierung des Datensatzes und die Datenorganisation beschrieben. Weiterhin wird aufgezeigt, dass es über die Zeit nicht ausreicht, lediglich die Armut von Haushalten zu betrachten, da eine Mobilität deutscher Haushalte zwischen den Zuständen arm und nicht-arm zu beobachten ist. Diese Erkenntnisse zeigen, dass es neben der Armut sinnvoll ist, auch die Vulnerabilität deutscher Haushalte zu betrachten.

Kapitel 4 bildet den zentralen Teil dieser Arbeit und umspannt die drei bereits erwähnten Teilanalysen zur Genauigkeit der Vulnerabilitäts-Schätzer. Der erste Teil untersucht, wie genau sich die Vulnerabilität mit Haushaltskovariablen und/oder Vorjahreseinkommen unter Verwendung von Querschnitts- oder Paneldaten mes-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sen lässt. Im zweiten Teil wird analysiert, welche Auswirkung die Verwendung von Einkommen in wenigen Klassen (zwei oder drei) anstelle von stetigen Einkommensdaten auf die Genauigkeit der Schätzer hat. Da die Einkommensvariablen im SOEP stetig vorliegen, wird die Genauigkeit für verschiedene Klassengrenzen betrachtet. Im Anschluss wird für eine festgelegte Klassengrenze bei zwei Einkommensklassen untersucht, wie die Genauigkeit von Schätzern zu bewerten ist, die auf grober Einkommensinformation, d.h. ungefährer Zuordnung der Haushalte auf die Einkommensklassen, basieren. Im dritten Teil werden Möglichkeiten untersucht, die Ergebnisse aus dem ersten Teil zu verbessern. Dazu werden zunächst neben den Haushaltskovariablen weitere Ursachen für die Vulnerabilität in Deutschland auf den Ebenen Strukturwandel der Volkswirtschaft sowie der Haushalte betrachtet. Es wird mit zwei Methoden untersucht, ob sich die Vulnerabilität in Gruppen besser schätzen lässt als für alle Haushalte gemeinsam und ob die Einbeziehung von Makrovariablen zu einer Verbesserung der Schätzer des ersten Teils führt.

Das abschließende Kapitel 5 gibt eine Zusammenfassung der wichtigsten Erkenntnisse dieser Arbeit und ihre Anwendungsmöglichkeiten sowie einen Ausblick auf mögliche Erweiterungen der vorgestellten Analysen und Modelle.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2 Die Messung der Vulnerabilität der Armut

In diesem Kapitel werden die Grundlagen dieser Arbeit vorgestellt. Zum entwicklungsökonomischen Fundament zählen die Definition und die Abgrenzung der Begriffe *Armut* und *Vulnerabilität der Armut* sowie die verschiedenen axiomatischen Ansätze zu ihrer Messung. Besonderen Wert wird auf die Abwägung von Vor- und Nachteilen dieser Ansätze als Basis für das empirische Vorgehen gelegt. Als statistischer Grundstein wird das etablierte Instrument ROC, das die Genauigkeit von Diagnosemethoden misst und in dieser Arbeit zur Bewertung der Vulnerabilität genutzt wird, beschrieben.

Die Grundlagen liefern wichtige Hintergrundinformationen für die Messung der Vulnerabilität. Die Nennung und Beschreibung der unterliegenden Modelle erfolgt ebenfalls in diesem Kapitel, während die Begründung der Modelle im empirischen Teil erfolgt.

Das Kapitel gliedert sich wie folgt. Im Abschnitt 2.1 wird der Begriff der Armut definiert und beschrieben, wie Armut im absoluten und im relativen Kontext gemessen wird. In Bezugnahme auf die verwendeten deutschen Daten in dieser Arbeit wird die Entwicklung und Akzeptanz des Armutsbegriffs in Deutschland beschrieben. Auf Basis der Daten wird eine Armutslinie für diese Arbeit gewählt und die Situation von Armut und Ungleichheit in Deutschland erläutert.

Abschnitt 2.2 erläutert den Begriff Vulnerabilität der Armut und grenzt diesen zum Begriff der Armut ab. Es folgt eine Beschreibung von axiomatischen Ansätzen zur Messung der Vulnerabilität. Besonderes Augenmerk wird auf die Maße Vulnerabilität als erwartete Armut (Vulnerability as Expected Poverty, VEP), Vulnerabilität als erwarteter Nutzen (Vulnerability as Low Expected Utility, VEU) sowie den axiomatischen Ansatz von Calvo und Dercon (2005) gelegt. Wichtige Ansätze, die diese Konzepte anwenden, werden beschrieben und diskutiert.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Das Verfahren ROC wird in Abschnitt 2.3 erläutert und diskutiert, welches der Gütemaße in dem Kontext dieser Arbeit zur Messung der Genauigkeit von Schätzern der Vulnerabilität zu verwenden ist.

In Abschnitt 2.4 werden die vewendeten Schätzmethoden der Vulnerabilität beschrieben. Die Methoden berücksichtigen verschiedene Situationen der Datenverfügbarkeit. In Abschnitt 2.4.1 werden Situationen beschrieben, bei denen einerseits nur Querschnitts- und andererseits Paneldaten und stetiges Einkommen verfügbar sind. Abschnitt 2.4.2 betrachtet unterschiedliche Modelle für die Situation, in der zwar Paneldaten aber Einkommen nur in wenigen Einkommensklassen zur Verfügung steht. Abschnitt 2.4.3 betrachtet die Modellierung für Gruppen von Merkmalen.

2.1 Armut

Armut ist ein Phänomen, das sowohl in Entwicklungsländern als auch in Wohlfahrtsstaaten, wenn auch oft unerkannt, auftritt. Da sich das Ausmaß der Armut wesentlich unterscheidet, werden in der Armutsforschung verschiedene Begriffe gebraucht: die absolute oder extreme Armut und die relative Armut. Unter absoluter Armut versteht man, wenn Individuen die zum Überleben wichtigen Grundbedürfnisse wie Kleidung, Nahrung, Wasser und Obdach fehlen (Ray, 1998, S. 250f.). Als absolut arm werden Menschen per Definition angesehen, wenn ihnen weniger als 1.25$ pro Tag zur Verfügung stehen. Derzeit sind etwa 1.4 Milliarden Menschen in Entwicklungsländern von absoluter Armut betroffen (brand eins und Statista, 2009, S. 42).

Nach dem Konzept der relativen Armut gilt ein Individuum als arm, wenn es weit unter dem durchschnittlichen Lebensstandard seines Landes lebt. Aufgrund dieser Definition werden Individuen als einkommensarm bezeichnet, wenn ihr Haushaltseinkommen gewichtet um die Anzahl und das Alter der Personen, das sogenannte *Äquivalenzeinkommen*, kleiner ist als 50% oder 60% des Medians der Äquivalenzeinkommen aller Personen (Atkinson, 2002, S. 83ff.).[1] Aufgrund dieser Definition steigt die Armutsgrenze, wenn der durchschnittliche Lebensstandard steigt, jedoch

[1] Da Armut als mehrdimensionale Größe gemessen werden soll, werden zusätzlich zur Höhe des Einkommens noch die Lebenslagen wie Berufstätigkeit, Bildung, Gesundheit, Unterkunft hinzugezogen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

führt ein steigender durchschnittlicher Lebensstandard nicht zwangsläufig zu einer sinkenden relativen Armutsquote.

In dieser Arbeit werden deutsche Daten analysiert, was für die Wahl einer relativen Armutsgrenze sprechen würde. Zur Schätzung der Vulnerabilität ist es allerdings notwendig, zukünftige Werte der Armutsgrenze zu kennen. Würde eine relative Armutsgrenze verwendet werden, so wären zukünftige Werte unbekannt und müssten z.B. durch den aktuellen Wert geschätzt werden. Für längere Zeitreihen könnte die Armutslinie anhand von Zeitreihenmodellen vorhergesagt werden. Um Variation durch diese Schätzfehler zu vermeiden, ist es im Kontext dieser Arbeit daher sinnvoll eine absolute Armutsgrenze zu wählen. Das Niveau der absoluten Armutsgrenze wird nahe der relativen Armutsgrenzen bei 50% des Median-Einkommens, deren Schwankungen relativ moderat sind, festgelegt.

Während Abschnitt 2.1.1 auf die Entwicklung und Akzeptanz des Armutsbegriffs in Deutschland eingeht, stellt Abschnitt 2.1.2 die Fakten zu Armut und Ungleichheit in Deutschland über den beobachteten Zeitraum dar.

2.1.1 Der Begriff Armut in Deutschland

Während dem Thema Armut in Deutschland in der Vergangenheit wenig Aufmerksamkeit gezollt wurde, so hat dieses in den letzten Jahren sowohl in der Politik als auch in den Medien stark an Aktualität gewonnen. Obwohl bereits auf dem Weltsozialgipfel 1995 in Kopenhagen von den Regierungschefs beschlossen wurde, Armut national zu dokumentieren sowie Strategien gegen sie auf regionaler, nationaler und internationaler Ebene vorzuschlagen (United Nations, 1995), wurde dies in Deutschland erst 2001 mit dem 1. Armuts- und Reichtumsbericht für die Bundesrepublik unter dem Namen „Lebenslagen in Deutschland“ (Deutscher Bundestag, 2001) umgesetzt. Seitdem sollen weitere Berichte immer zur Mitte einer Legislaturperiode folgen (2. Armuts- und Reichtumsbericht 2005 (Deutscher Bundestag, 2005), 3. Armuts- und Reichtumsbericht 2008 (Deutscher Bundestag, 2008)). Zwar wurde es von Armutsforschern als positiv aufgefasst, dass die Bundesregierung die Existenz von Armut in Deutschland anerkannte; die Berichte werden jedoch stark kritisiert (siehe z.B. Butterwegge, 2009, S. 59ff.).

Die Messung der Armut in den Armuts- und Reichtumsberichten orientieren sich an den EU-Indikatoren, auf die sich die Länder der EU im Jahr 2000 auf einer Sondertagung des Europäischen Rates verständigt haben und die gewährleisten sollen, dass Armut in Europa einheitlich gemessen wird und damit vergleichbar

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ist. Es wurde festgelegt, Armut als *mehrdimensionale* und *relative* Größe zu messen. Dieser Grundsatz wurde in 18 Indikatoren, sogenannten Laeken-Indikatoren, formuliert. Armut wird nach diesem Konzept nicht mehr, wie lange Zeit üblich, allein auf der Grundlage des Einkommens sondern auch basierend auf anderen Lebensgrundlagen wie Gesundheit, Arbeit und Bildung gemessen (Atkinson et al. (2002, S. 33f.), Krause und Ritz (2006)).

Einkommensbasierte Indikatoren sollten zum Vergleich der Einkommenssituation von Individuen oder Haushalten das Äquivalenzeinkommen nutzen. Dies ist definiert als das Haushaltseinkommen, das aufgrund einer Äquivalenzskala unter Berücksichtigung der Haushaltsstruktur (Anzahl und Alter der Haushaltsmitglieder) gewichtet wird. Verwendet werden sollte die neue OECD-Äquivalenzskala. Der Grund für die Gewichtung ist, dass Ein- und Mehrpersonenhaushalte auf der Basis des einfachen Haushaltseinkommens nicht miteinander vergleichbar sind; durch Zusammenleben von Personen in einem Haushalt entstehen Skaleneffekte, z.B. durch Nutzung gemeinsamer Haushaltsgeräte (Atkinson et al. (2002, S. 98ff.), Krause und Ritz (2006)).

Die Indikatoren verwenden Armutsschwellen von 60% sowie 40%, 50% und 70% des Median-Äquivalenzeinkommens. Die zur 60%-Schwelle gehörende Armutsquote wird als *Armutsgefährdungsquote* bezeichnet.[2] Die Armuts- und Reichtumsberichte verwenden dagegen 50% und 60% des Mittelwert- und Median-Äquivalenzeinkommens (Deutscher Bundestag (2001), Deutscher Bundestag (2005), Deutscher Bundestag (2008)).

Obwohl Armut idealerweise im multidimensionalen Kontext definiert werden sollte, verwenden zahlreiche Studien (z.B. Frick et al. (2005a), Frick und Grabka (2005b)) nur die Einkommensdimension, um Individuen oder Haushalte als arm oder nicht-arm zu klassifizieren. In dieser Arbeit werden Haushalte ebenfalls aufgrund des Einkommens, genauer dem äquivalenzgewichtetem Jahresnettoeinkommen, als arm oder nicht-arm klassifiziert. Der Vorteil dieser Definition ist, dass weniger Daten benötigt werden und nach Wahl der Armutsgrenze eine eindeutige Aussage gemacht werden kann, ob ein Individuum oder Haushalt (gemäß dieser Definition) arm ist.

[2] Da die 60%-Schwelle zu einer recht hohen Armutsquote führt, wird ein abschwächender Begriff verwendet, um „keine falschen Signale an die Öffentlichkeit zu senden“. Dieser Begriff könnte leicht zu Missverständnissen führen und als Gefahr der Armut in der Zukunft verstanden werden (Krause und Ritz, 2006).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2.1.2 Armutsdefinition im Kontext der Schätzung von Vulnerabilität

In dieser Arbeit sind Haushalte Gegenstand der Untersuchungen. Jedem Haushalt wird das aufgrund seiner Haushaltsstruktur gewichtete Haushaltseinkommen zugewiesen. Armutsgrenzen werden, anders als in vielen Studien, bzgl. des Medians von je *einem* Äquivalenzeinkommen pro Haushalt berechnet und nicht von den Äquivalenzeinkommen aller Personen (Stauder und Hüning, 2004). Die Unterschiede in den Armutsgrenzen beider Vorgehensweisen sind für Deutschland nicht besonders groß. Die Niveaus der *haushaltsbezogenen* Armutsgrenzen sind etwas höher als die der *personenbezogenen* Armutsgrenzen.

Abbildung 2.1 zeigt die relativen Armutslinien bei 50% und 60% des oben beschriebenen Median-Äquivalenzeinkommens auf Basis der im nächsten Kapitel beschriebenen Daten. Wie dort noch näher erläutert wird, handelt es sich bei der verwendeten Einkommensvariable um Jahreseinkommen, das für ein aktuelles Jahr t nicht im Befragungsjahr sondern erst im darauffolgenden Jahr $t + 1$ bekannt ist. Die x-Achse in Abbildung 2.1 zeigt das Jahr, in dem das Einkommen erhalten wurde und nicht das Beobachtungsjahr. Dies gilt ebenso für die folgenden Abbildungen 2.2 und 2.3.

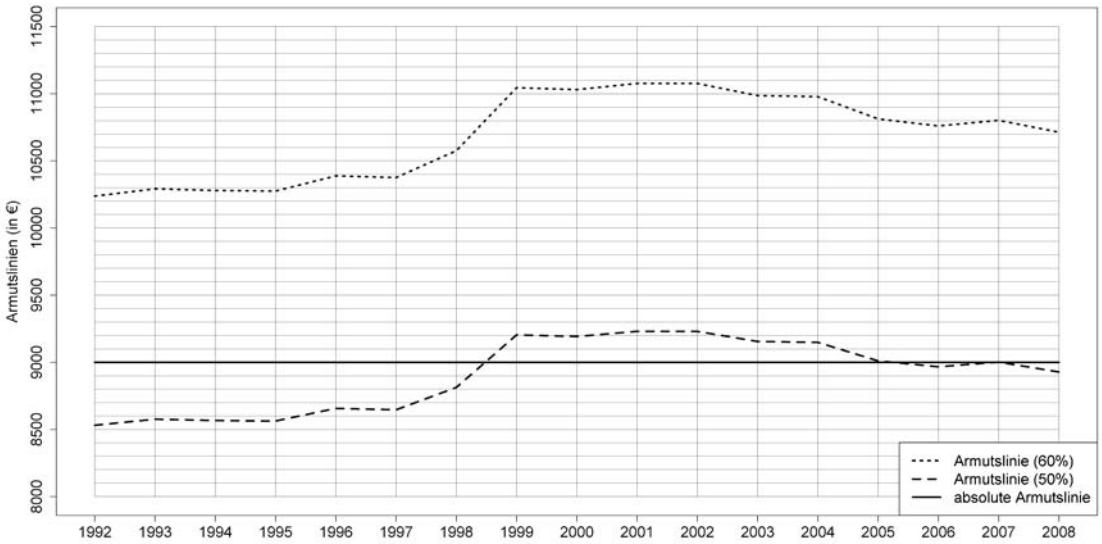


Abbildung 2.1: *Relative und absolute Armutslinien (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die relativen Armutslinien sind von 1992 bis 1997 relativ konstant und steigen dann bis 1999 an. Bis 2004 sind die Linien erneut in etwa konstant und fallen dann bis 2008 leicht ab. Für diese Arbeit wird das reale Niveau der absoluten Armutsgrenze bei 9 000 € gewählt, deren Wert relativ nah an der relativen 50%-Armutsgrenze liegt (siehe Abbildung 2.1).

In Abbildung 2.2 sind die zu den Armutsgrenzen zugehörigen Armutsquoten dargestellt. Anfang der 90er Jahre schwankten alle Armutsquoten recht stark. Ab 1994 lag eine leicht fallende Tendenz aller drei Armutsquoten vor, während diese im Zeitraum von 1999 (bzw. 1998 für die höhere relative Armutsquote) bis 2006 die Armutsquoten waren. Danach stagnierten sie auf diesem Niveau. Die Armutsquoten, die sich bei absoluter Armutsgrenze von 9 000 € und relativer 50%-Armutsgrenze ergaben, sind sich sehr ähnlich. Auffallend ist die besonders niedrige absolute Armutsquote in 1999. Diese lässt sich durch den Anstieg des Medianeinkommens ab 1997 erklären. Bis 1998 liegt die relative 50%-Armutsgrenze unter der absoluten, ab 1999 bis 2005 darüber.

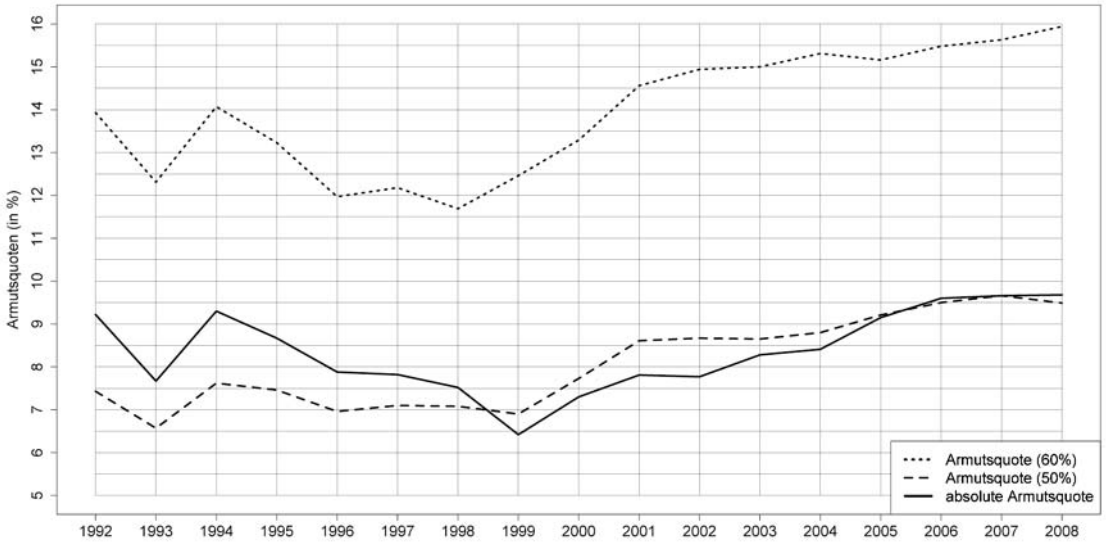


Abbildung 2.2: *Entwicklung zugehöriger Armutsquoten (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Nach der These von Butterwegge (2009, S. 11) ist der Anstieg der Armutsquote auf den Anstieg der sozialen Ungleichheit zurückzuführen. Als Maß für die Ungleichheit der Verteilung des Einkommens wird der Gini-Index (Creedy, 1998, S. 13-17), der in Abbildung 2.3 dargestellt ist, herangezogen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Es zeigt sich, dass der Gini-Koeffizient von 1997 bis 2005 von 0.27 auf 0.31 angestiegen ist und seitdem zwischen 0.3 und 0.31 stagniert.

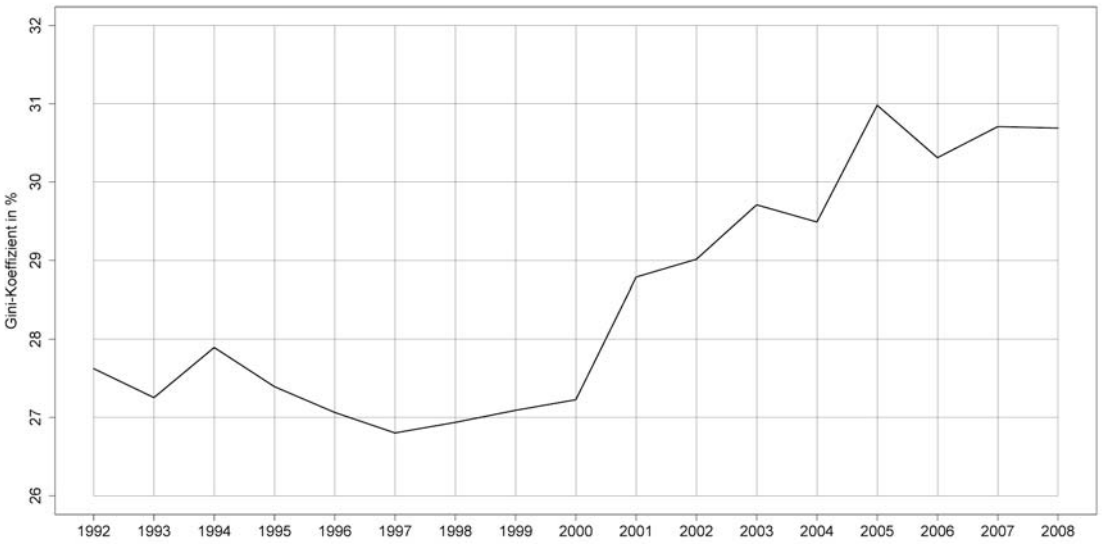


Abbildung 2.3: *Entwicklung der Einkommensungleichheit in Deutschland (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Studien der OECD von 2008 und 2011, die die Einkommensungleichheit und Armut von Deutschen (Individuen) mit dem im vorherigen Abschnitt beschriebenen Methoden untersuchten, kamen zu dem Ergebnis, dass die Ungleichheit und die Armut seit dem Jahr 2000 stark gestiegen sind (sogar stärker als in jedem anderen OECD-Land). Diese Ergebnisse bestätigen sich auch bei der Analyse im Haushaltskontext und mit der vorgeschlagenen absoluten Armutsquote. Auf die Ursachen wird in Kapitel 4 eingegangen.

2.2 Vulnerabilität der Armut

Der Begriff *Vulnerabilität* ist in den letzten zwei Jahrzehnten, insbesondere in der Klima- und Armutsforschung, häufig verwendet worden (Dietz, 2006). Die wissenschaftliche Bedeutung des Begriffs hat sich aus zwei verschiedenen Konzepten, dem „Risk- or Natural-Hazard Approach“ und dem „Social Vulnerability Approach“ entwickelt, die ihren Ursprung in der Katastrophenforschung in den 80er und 90er Jahren haben. Der erste Ansatz misst das Ausmaß von externen Schocks, z.B. auf

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

eine Region (Dietz (2006), Yamin et al. (2005)). Aus dieser Sicht betrachtet, könnte die Vulnerabilität der Armut in einer Region durch Naturkatastrophen, z.B. das Auftreten und die Stärke von Erdbeben, erklärt werden. Ein Nachteil dieses Ansatzes ist, dass Daten über solche extremen Ereignisse selten verfügbar sind und dass das genaue Ausmaß schwierig zu bewerten ist. Der zweite Ansatz konzentriert sich auf die Mikroebene. Vulnerabilität wird als ein Risiko betrachtet, das aufgrund der Struktur und Eigenschaften von Haushalten und Individuen besteht. Pionierarbeit für diesen Ansatz ist der „Entitlement Approach“ von Sen (1981), der im Kontext des „Social Vulnerability Approach“ angewendet wird. Ansprüche und deren Veränderungen messen nach diesem Ansatz die Fähigkeiten von Haushalten, Schocks wie Dürren und Preisanstiege zu überstehen.

Aktuelle Politikforschung zur Bekämpfung der Armut (siehe z.B. World Development Report 2000/01, Poverty Reduction Strategy Papers (PRSP)) zielen darauf ab, die Vulnerabilität sowohl zu externen Schocks (Finanzkrisen, Naturkatastrophen, politischem Aufruhr, Kriminalität usw.) als auch zu Schocks auf der Haushaltsebene (Arbeitslosigkeit, Krankheit, Veränderungen in der Haushaltsstruktur usw.) zu verringern. Auf dem Gebiet der Entwicklungsökonomik basieren die Definitionen und Interpretationen des Begriffs Vulnerabilität im Wesentlichen auf den Ausführungen von Chambers (1989), wie von Dietz (2006) berichtet wird. In seiner Definition hat Vulnerabilität „two sides: an external side of risks, shock and stress to which a household or an individual is subject; and an internal which is defencelessness, meaning a lack of means to cope with damaging loss“. Obwohl Armut und Vulnerabilität der Armut eng miteinander verbunden sind, handelt es sich doch um zwei verschiedene Konzepte. Der Begriff Armut bezieht sich auf einen Zustand in einem *Zeitpunkt*, der gewöhnlich ex post durch das Haushaltseinkommen oder die Konsumausgaben gemessen wird. Dagegen beschreibt der Begriff der Vulnerabilität einen möglichen Zustand in der Zukunft, d.h. ein Ereignis, das eintreten oder nicht eintreten könnte. Damit hat die Vulnerabilität, anders als Armut, den Charakter einer Wahrscheinlichkeitsvorhersage, d.h. einer ex ante Bewertung des Armutsrisikos.

Seit der Veröffentlichung des World Development Reports (World Bank, 2001) haben sich viele Forscher damit beschäftigt, die Vulnerabilität nach der Definition des Berichts zu messen: *“... the risk that a household or individual will experience an episode of income or health poverty over time. But vulnerability also means the probability of being exposed to a number of other risks (violence, crime natural disasters, being pulled out of school”)* (World Bank, 2001). Die Studien basieren hauptsächlich auf Daten von Entwicklungsländern. Im Anbetracht der dynami-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

schen Struktur von Vulnerabilität benötigt ihre Berechnung Paneldaten (World Bank, 2001). Trotzdem stützen viele Autoren ihre Analysen auf Querschnittsdaten (z.B. Chaudhuri (2002), Christiaensen und Subarrao (2005), Günther und Harttgen (2009), Jha und Dang (2010)). Der Grund dafür liegt hauptsächlich darin, dass Paneldaten in Entwicklungsländern selten verfügbar sind. Der Nachteil bei der Verwendung von Querschnittsdaten ist, dass strenge Annahmen getroffen werden müssen. Speziell wird unterstellt, dass die Vulnerabilität konstant über die Zeit bleibt und dass weder die Kovariablen noch der erwartete Response sich verändern. In der Notation der Modelle ist daher keine Zeitkomponente erforderlich. Nur die Schocks ändern sich. In der Tat wird angenommen, dass sich die Varianz über die Zeit durch die Varianz im Querschnitt messen lässt und dass Schocks nicht zeitlich korreliert sind.

Obwohl der Definition und der Berechnung der Vulnerabilität viel Aufmerksamkeit geschenkt wurde, existiert noch keine einheitlich akzeptierte Definition, z.B. basierend auf einer ökonomischen Theorie, und damit keine akzeptierten Indikatoren und Methoden für ihre Messung (Chambers, 1989). Konzepte für die Vulnerabilität der Armut, die entwickelt wurden, sind Vulnerability as Exposure to Risk (VER), Vulnerability as Expected Poverty (VEP), z.B. Chaudhuri (2002), Pritchett et al. (2000), und Vulnerability as Low Expected Utility (VEU), Ligon und Schechter (2003) sowie einen axiomatischen Ansatz, der erwartete Armut mit der Stärke der erwarteten Armut kombiniert, z.B. Calvo und Dercon (2005), Povel (2010). Alle Ansätze haben gemeinsam, dass sie basierend auf einem Modell die Wohlfahrt $y_{h,t+1}$ (z.B. Konsum, Einkommen) vorhersagen. VEP, VEU und der axiomatische Ansatz stimmen außerdem darin überein, dass sie die Wahrscheinlichkeit p_h berechnen, dass Individuen oder Haushalte unter eine Benchmark z (bei VEP Armutslinie genannt) fallen. VER berechnet dagegen keine Wahrscheinlichkeiten sondern misst, ob beobachtete Schocks zu Wohlfahrtsverlusten führen. VER weist Ähnlichkeiten mit dem „Risk- or Natural-Hazard-Approach“ Ansatz auf und erzeugt ein ex post Maß. Die Vulnerabilität wird bei den Konzepten VEP, VEU und dem axiomatischen Ansatz auf der Ebene von Individuen bzw. Haushalten ($V_{h,t}(y_{h,t}, z, p_{h,t})$) gemessen und daraus das aggregierte Maß, d.h. die Vulnerabilität in der Gesamtbevölkerung, berechnet (Hoddinott und Quisumbing (2008), Ligon und Schechter (2004)).

In den nächsten beiden Abschnitten werden die Konzepte mit Ausnahme von VER vorgestellt. Weiterhin werden die führenden Studien beschrieben und die Vor- und Nachteile der Vorgehen gegeneinander abgewägt.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2.2.1 Konzepte zur Messung

Vulnerability as Expected Poverty (VEP)

Die Idee dieses Ansatzes ähnelt der Modellierung des Ausfallsrisikos eines Portfolios von Fishburn (1977), ein sogenanntes „safety-first measure“, wie Christiaensen und Subbarao (2005) zitieren. In diesem Kontext entspricht die Vulnerabilität dem erwarteten Ausfall in der nächsten Periode, d.h. der Erwartung eines Haushalts, in der nächsten Periode in Armut zu fallen. Die Berechnung der Vulnerabilität ergibt sich als Erwartung eines Wertes für den Armutsindex, der von der Armutslinie und der zukünftigen Wohlfahrt abhängt, bedingt durch die kumulative Verteilung $F(\cdot)$ des unterliegenden Wohlfahrtsindikators mit Dichtefunktion $f(\cdot)$, siehe Gleichung (2.1). Die Variable $w_{h,t+1}$ stellt den minimalen Wert des Wohlfahrtsindikators für einen Haushalt dar (Christiaensen und Subbarao, 2005).

$$V_{h,t} = E[\iota(z, y_{h,t+1})|F(y_{h,t+1})] = \int_{w_{h,t+1}}^{z} \iota(z, y_{h,t+1}) f(y_{h,t+1})\, \mathrm{d}y_{h,t+1} \tag{2.1}$$

Als Armutsmaß findet die Familie der Foster-Greer-Thorbecke Armutsindizes (FGT) (Foster et al., 1984) Anwendung, siehe Definition (2.2). Die Armutsindizes nehmen den Wert 0 an, wenn der Wert des Wohlfahrtsindikators oberhalb der Armutslinie liegt und eine nichtfallende Funktion in Abhängigkeit der drei Parameter $z, y_{h,t}, \alpha$, wenn sein Wert unter der Armutslinie liegt.

$$\iota(z, y_{h,t}, \alpha) = \begin{cases} \left(\frac{z-y_{h,t}}{z}\right)^{\alpha}, & \text{wenn } y_{h,t} < z \\ 0, & \text{wenn } y_{h,t} \geq z \end{cases} \tag{2.2}$$

Der Parameter α ist ein Parameter für die Tiefe der Armut; ein größeres α gewichtet die Ärmsten höher. Wird $\alpha = 0$ gewählt, so liefert der Index lediglich die Information, ob ein Haushalt arm ist oder nicht, d.h. es wird vernachlässigt, wie ausgeprägt die Armut ist. Im Falle $\alpha = 1$ misst der Index, „wie arm die Armen sind“, daher wird auch von „Armutslücke“ gesprochen. Im Falle von $\alpha > 2$ erfüllt die FGT Familie die von Sen (1976) formulierten Axiome für Armutsindizes (Foster et al.,1984).

$$V_{h,t,\alpha} = F(z) \int_{w_{h,t+1}}^{z} \left(\frac{y_{h,t+1} - z}{z}\right)^{\alpha} \frac{f(y_{h,t+1})}{F(z)}\, \mathrm{d}y_{h,t+1} \tag{2.3}$$

Durch Einsetzen von Gleichung (2.2) in (2.1) und Erweiterung ergibt sich die Vulnerabilität als Produkt der Wahrscheinlichkeit unter die Armutslinie z zu fallen, und einer wahrscheinlichkeitsgewichteten Funktion des relativen Einkommensausfalls, siehe Gleichung (2.3) (Christiaensen und Subbarao, 2005). Durch die Wahl von α

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

werden verschiedene Aspekte des Ausfalls dargestellt. Im Falle von $\alpha = 0$ entspricht die Vulnerabilität der Wahrscheinlichkeit, in Armut zu fallen. Je höher α gewählt wird, desto stärker sind Haushalte armutsgefährdet, je weiter ihre zukünftige Wohlfahrt von der Armutslinie entfernt liegt (Christiaensen und Boisvert, 2000). Zur Berechnung der aggregierten Vulnerabilitätsmaße aus den individuellen siehe Weikard (2004).

Der Vorteil des Konzeptes VEP ist, dass es leicht verständlich und einfach zu interpretieren ist. Es wird jedoch kritisiert (z.B. Ligon und Schechter (2003), Hoddinott und Quisumbing (2008), Gaiha und Imai (2008)), dass die Familie der FGT-Indizes die Risikopräferenzen nur ungenügend abbildet, wenn sie als Nutzenfunktion interpretiert wird. Ein Kritikpunkt besteht darin, dass nur der Wohlfahrt unter der Armutsgrenze z ein Nutzen zugeordnet wird. Weiterhin erhöht steigende Wohlfahrt unter der Armutsgrenze nicht den Nutzen. Diese Eigenschaft zeigt sich darin, dass die Indizes monoton fallend sind und damit die erste Ableitung $(\frac{\delta\iota(z,y_{h,t},\alpha)}{\delta y_{h,t+1}} = -\alpha\left(\frac{1}{z}\right)^{\alpha}(z - y_{h,t+1})^{(\alpha-1)} < 0)$ negativ für alle $\alpha > 0$ ist. Die Berechnung des Arrow-Pratt Koeffizienten zur absoluten Risikoaversion

$$r_{U(y_{h,t+1})} = \frac{\alpha - 1}{z - y_{h,t+1}}$$

führt in Abhängigkeit von α zu folgender Risikoneigung der Haushalte

$$\begin{cases} \text{risikofreudig} & \text{für } 0 \leq \alpha < 1 \\ \text{risikoneutral} & \text{für } \alpha = 1 \\ \text{risikoavers} & \text{für } \alpha > 1 \end{cases}$$

Für risikoaverse Haushalte steigt demnach die Risikoaversion mit steigender Wohlfahrt unter der Armutslinie, was im Widerspruch zu empirischen Belegen steht (Calvo und Dercon, 2005).

Hoddinott und Quisumbing (2008) beschreiben eine Situation, in der das Konzept VEP zu einer unplausiblen Entscheidung führt. Angenommen werden zwei Szenarien. Im ersten Szenario ist der Haushalt sicher unter der Armutslinie, während im zweiten Szenario dieser Haushalt mit Wahrscheinlichkeit 0.5 unter die Armutslinie fällt und mit Wahrscheinlichkeit 0.5 nicht arm wird. Unter der Annahme, dass der Haushalt risikoavers ist, bedeutet dies im FGT-Kontext, dass sich der Haushalt im zweiten Szenario verschlechtert, obwohl sich die Vulnerabilität von 1 auf 0.5 verringert. Aufgrund der Risikoaversion würde er den sicheren Konsum dem unsicheren vorziehen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Aus den genannten Gründen nutzen einige Studien ein anderes Vulnerabilitätsmaß, VEU, von dem behauptet wird, dass es die Risikopräferenzen besser abbilden kann als die FGT-Indizes (siehe z.B. Ligon und Schechter (2004), Cunningham und Maloney (2000)).

Vulnerability as Low Expected Utility (VEU)

Unter dem Konzept VEU wird die Vulnerabilität zum Zeitpunkt t als Differenz von dem Nutzen $U_h(\cdot)$ eines „Sicherheitsäquivalents“ und dem erwarteten Nutzen des Konsums in möglichen Umweltzuständen ω in $t+1$ (siehe Gleichung (2.4)) (Ligon und Schechter, 2003) definiert.

$$V_{h,t} = U_h(z_{CE}) - E\left(U_h(y_{h,t+1}(\omega))\right) \tag{2.4}$$

Ein Sicherheitsäquivalent ist das sichere Konsumniveau μ, d.h. der erwartete Konsum aller Haushalte, der zu demselben Nutzen führt wie das unsichere Konsumniveau. Wenn ein Haushalt dieses erwartete Konsumniveau erreicht, hat er eine Vulnerabilität von Null.

$U_h(\cdot) : R \rightarrow R$ ist eine monoton steigende, schwach konkave Funktion, die dem Konsum einen Nutzen zuweist. Die Krümmung der Nutzenfunktion stellt die Annahme dar, dass Haushalte risikoavers sind. Als Nutzenfunktion verwenden die Autoren die Funktion in (2.5). Diese hat die Form einer Box-Cox Transformation (Box und Cox, 1964) mit Parameter $1-\lambda$.

$$U_h(y_{h,t}) = \begin{cases} \frac{y_{h,t}^{1-\lambda}}{1-\lambda}, & \text{wenn } \lambda \neq 1 \\ log(y_{h,t}), & \text{wenn } \lambda = 1 \end{cases} \tag{2.5}$$

Die Nutzenfunktion hat eine konstante relative Risikoaversion λ und eine zwischenzeitliche Rate der Substitution von $\frac{1}{\lambda}$, d.h. die Bereitschaft, Konsum zwischen den Perioden zu ersetzen. Je größer λ, umso weniger ist ein Haushalt bereit, Konsum über die Zeit zu substituieren. Als Grenznutzen ergibt sich $U_h'(y_{h,t}) = y_{h,t}^{-\lambda}$. Der Grenznutzen fällt umso stärker, je mehr der Konsum steigt, d.h. umso größer λ ist. Es wird angenommen, dass die Haushalte einen stabilen Konsum relativ zu höherem Konsum in der Zukunft vorziehen.

Die Zerlegung der Vulnerabilität der Armut in einen ex post (Armut) und einen ex ante Teil (Risiko) wird in Gleichung (2.6) dargestellt und in Abbildung 2.4, wobei $\lambda = 0.8$ gewählt wurde, veranschaulicht. Die Komponente der Armut erfüllt für eine spezielle Form der Nutzenfunktion die Axiome von Foster et al. (1984),

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

während die Komponente des Risikos konsistent mit dem ordinalen Risikomaß von Rothschild und Stiglitz (1970) ist, wie Ligon und Schechter (2003) zitieren.

$$V_{h,t} = \overbrace{[U_h(z) - U_h(Ey_{h,t+1}(\omega))]}^{Armut} + \overbrace{[U_h(Ey_{h,t+1}(\omega)) - EU_h(y_{h,t+1}(\omega))]}^{Risiko} \tag{2.6}$$

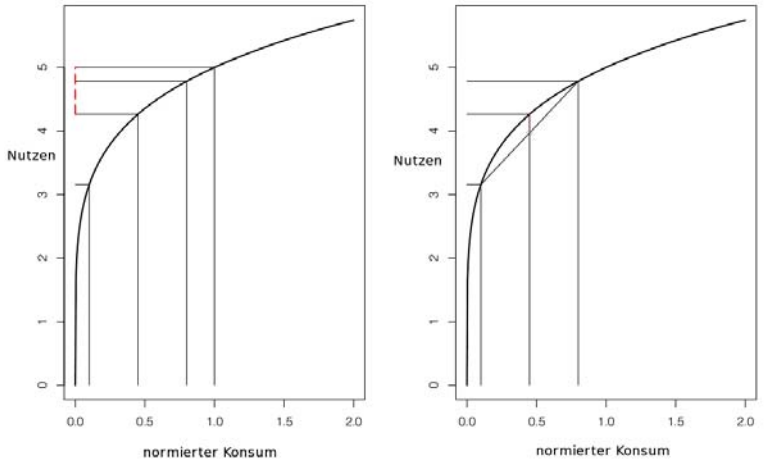


Abbildung 2.4: *Vulnerabilitätskonzept von VEU [Eigene Darstellung mit künstlich erzeugten Daten].*

Die Abbildung auf der linken Seite zeigt den Teil der Armut, der keine Zufalfsvariablen enthält. $U_h(z)$ ist der konstante Nutzen, bewertet am Niveau des Sicherheitsäquivalents, während $U_h(Ey_{h,t+1}(\omega))$ der Nutzen des erwarteten Konsums in möglichen Umweltzuständen ist. Wenn $Ey_{h,t+1}(\omega)$ unverändert bleibt, ist es derselbe Punkt auf der Nutzenkurve, unabhängig wie sich die Zustände verhalten. Der rechte Teil veranschaulicht die Risikokomponente in der Definition. $EU_h(y_{h,t+1}(\omega))$ hängt von den einzelnen Umweltzuständen ab, auch wenn $Ey_{h,t+1}(\omega)$ unverändert bleibt. Je weiter die eingetretenen Konsumniveaus der Umweltzustände voneinander entfernt liegen, desto größer ist die Differenz $[U_h(Ey_{h,t+1}(\omega)) - EU_h(y_{h,t+1}(\omega))]$ und umso höher ist das Risiko, dem der Haushalt ausgesetzt ist.

Gaiha und Imai (2008) und Christiaensen und Subbarao (2005) kritisieren, dass die Ergebnisse des VEU Konzeptes abhängig von der Wahl der Nutzenfunktion sind. Weiterhin argumentieren Christiaensen und Subbarao (2005), dass der Ansatz VEU nur Sinn mache, wenn die individuellen Risikopräferenzen tatsächlich geschätzt werden können. Selbst wenn diese empirisch geschätzt werden könnten, würde dies nicht bedeuten, dass das Vorgehen zur Schätzung der Vulnerabilität geeignet sei, da die Haushalte schlecht über ihre Präferenzen informiert seien (Griffin, 1986)

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

und Shackle (1965) und Kanbur (1987) zeigen auf, dass „it is hard to imagine that human knowledge can be so perfect that tomorrows hunger or pain can be felt today.“(Christiaensen und Subbarao, 2005). Just und Pope (2003) schreiben, dass es schwierig ist, die Ursachen von Risiko „richtig“ zu identifizieren und abzubilden, und die Ergebnisse als Grundlage für politische Entscheidungen zu nutzen. „We conclude that understanding of why risk response occurs is very limited. As a result, after decades of research, the profession remains in a weak position to offer definitive policy analyses in matters related to risk.“.

Die Vulnerabilität nach dem Konzept VEU kann ebenfalls im Konzept von VEP dargestellt werden, wie Gleichung (2.7) (für das aggregierte Vulnerabilitätsmaß) zeigt (Calvo und Dercon, 2005).

$$V_t = \sum_{h=1}^{N} p_h(U_h(z_{CE}) - U_h(y_{h,t+1})) \tag{2.7}$$

Z.B. Calvo und Dercon (2005) sowie Christiaensen und Subbarao (2005) kritisieren, dass das sogenannte Fokus Axiom verletzt ist. Das Fokus Axiom (Foster et al., 1984) in diesem Kontext besagt, dass nur Haushalte, die dem Risiko unterliegen unter die Armutslinie zu fallen, das Vulnerabilitätsmaß verändern. Die Verletzung des Fokus-Axioms führt also dazu, dass die Vulnerabilität unterschätzt wird („... negative states of the world would be compensated by positive ones“ (Christiaensen und Subbarao, 2005).

Axiomatischer Ansatz

Calvo und Dercon (2005) schlagen zur Schätzung der Vulnerabilität einen axiomatischen Ansatz vor. Dieser Ansatz ist ähnlich zu Ansätzen in der Armutsforschung, bei denen zunächst Axiome definiert werden, die ein geeignetes Maß erfüllen soll und anhand von deren Vorgaben Maße konstruiert werden. Calvo und Dercon (2005) verstehen Vulnerabilität der Armut als „the magnitude of threat of future poverty“. Dabei beziehen sie die Gefahr der Armut sowohl auf das Risiko, d.h. die Wahrscheinlichkeit des Eintritts, als auch auf die Intensität der Armut und notieren ein Vulnerabilitätsmaß als $V_h(z, \omega, p_\omega)$. Dabei ist z wieder die Armutslinie, ω ein Vektor mit Werten für das Einkommen in verschiedenen Umweltzuständen und p_h ein Vektor mit Eintrittswahrscheinlichkeiten der Umweltzustände. Die Autoren definieren sechs Axiome[3], die ein geeignetes Vulnerabilitätsmaß erfüllen sollte und

[3] Diese Axiome sind *Axiom 1 - Symmetrie*, *Axiom 2 - Fokus*, *Axiom 3 - Wahrscheinlichkeitsabhängiger Effekt des Einkommens*, *Axiom 4 - Wahrscheinlichkeitstransfer*, *Axiom 5 - Risikosensibilität*, *Axiom 6 - Skaleninvarianz* und *Axiom 7 - Normalisierung*. Für Details, siehe Calvo und Dercon (2005).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

die von (2.8) erfüllt werden. Eine empirische Studie liegt z.B. von Povel (2010) für Thailand und Vietnam vor. Wird zusätzlich eine Normalisierung des Vulnerabilitätsmaßes auf das Intervall $[0, 1]$ vorgenommen und weitere Forderungen an die Risikosensibilität gestellt, so nennen die Autoren zwei Familien, die diese Forderungen erfüllen. Das Maß (2.9) erfüllt zusätzlich zu den bereits genannten Axiomen noch das Axiom der konstanten relativen Risikoaversion, während das Maß (2.10) das Axiom der konstanten absoluten Risikoaversion erfüllt.

$$V(z, \omega, p_\omega) = \sum_{h=1}^{|\Omega|} p_h \frac{y_h}{z} \tag{2.8}$$

$$V(z, \omega, p_\omega)^\epsilon = 1 - E[x_h^\epsilon] \quad 0 < \epsilon < 1 \tag{2.9}$$

$$V(z, \omega, p_\omega)^\kappa = E\left[\frac{e^{\kappa(1-x_h)} - 1}{e^\kappa - 1}\right] \quad \kappa > 0 \tag{2.10}$$

In Bezug auf die oben diskutierten Konzepte ergibt sich, dass das Maß des Konzeptes VEU für $\alpha > 1$ die Axiome 1-6 erfüllt, während das Fokus-Axiom verletzt wird und das Skaleninvarianzaxiom nur für eine spezielle Nutzenfunktion in logarithmischer Form erfüllt ist, siehe dazu Calvo und Dercon (2005).

Kritisch zu bemerken ist, dass die axiomatische Formulierung eines Maßes keine Auskunft darüber gibt, wie die Vulnerabilität bzw. die Einkommen in den verschiedenen Umweltzuständen und deren Eintrittswahrscheinlichkeiten zu schätzen sind. Negativ ist weiterhin zu bemerken, dass bei den Maßen keine Zeitkomponente berücksichtigt wurde.

2.2.2 Ansätze zur Messung

VEP

Der am häufigsten angewendete Ansatz (z.B. Suryahadi und Sumatro (2003), Günther und Harttgen (2009), Jha und Dang (2010)) ist die Methode von Chaudhuri et al. (2002) zur Messung der Vulnerabilität anhand von Querschnittsdaten. Chaudhuri et al. (2002) definieren die Vulnerabilität eines Haushalts auf Basis des Konsums y_h als Wahrscheinlichkeit, in der nächsten Periode arm zu werden bzw. arm zu bleiben.

$$V_h = P(y_h < z) \tag{2.11}$$

Zur Berechnung der Vulnerabilität schätzen die Autoren das Modell in (2.12). Modelliert wird der logarithmische Konsum $\ln(y_h)$ als eine Funktion von Haushaltseigenschaften X_h und haushaltsspezifischen Schocks e_h, die normalverteilt mit Er-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

wartungswert 0 und Varianz $\sigma^2_{e_h}$ sind. Es wird angenommen, dass die Schocks nicht autokorreliert sind. Die Struktur der Volkswirtschaft wird als relativ stabil über die Zeit angenommen, so dass Schocks nur auf der Haushaltsebene auftreten. Die Varianz des Fehlers wird in Abhängigkeit der Haushaltseigenschaften dargestellt. Die Variablen β und γ stellen Regressionskoeffizienten dar.

$$\begin{aligned} \ln y_h &= X_h\beta + e_h \\ \sigma^2_{e_h} &= X_h\theta \\ e &\sim N(\underline{0}, \Sigma) \end{aligned} \tag{2.12}$$

Die bedingten Erwartungswerte und die Varianzen werden mit einer 3-Schritt-FGLS-Methode (Feasible Generalized Least Square) (Amemiya, 1977) wie in (2.13) und (2.14) geschätzt.

$$E[\ln(y_h)|X_h] = X_h\hat{\beta} \tag{2.13}$$

$$Var[\ln(y_h)|X_h] = X_h\hat{\theta} \tag{2.14}$$

Mit der kumulativen Normalverteilung wird die Vulnerabilität anhand des geschätzten Erwartungswertes und der Varianz wie in (2.15) dargestellt.

$$\hat{V}_h = \Phi\left(\frac{\ln z - X_h\hat{\beta}}{\sqrt{X_h\hat{\theta}}}\right) \tag{2.15}$$

Der Ansatz basiert auf Querschnittsdaten und unterliegt den bereits genannten strengen Annahmen und Schwächen.

Günther und Harttgen (2009) schätzen mit einer Erweiterung des Ansatzes von Chaudhuri et al. (2002) die Vulnerabilität in Madagaskar. Die Autoren verwenden zur Schätzung des Konsums mehrstufige Daten, d.h. Eigenschaften von Gemeinden und von Haushalten in diesen Gemeinden. Die Begründung dafür ist, dass Haushalte einer Gemeinde ähnlichen Bedingungen ausgesetzt sind und damit nicht unabhängig voneinander sind. Da die Erweiterung ebenfalls auf Querschnittsdaten basiert, birgt sie dieselben Schwächen wie der Ansatz von Chaudhuri et al. (2002).

Christiaensen und Subbarao (2005) schätzen die Vulnerabilität für Kenia für das Jahr 1997 mit dem Welfare Monitoring Survey (WMS) von 1994 (t) und 1997 ($t+1$). Sie verwenden ebenfalls hierarchische Daten (Haushalte h und Gebiete j) und ergänzen das Modell von Chaudhuri et al. (2002) um historische Schocks sowie den Interaktionstermen aus Schocks und Haushaltskovariablen. Konkret verwenden die Autoren Schocks für Dürre („covariant shock“) und Malaria („idiosynratic

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

shock“), gemessen durch die Niederschlagsmenge (als prozentuale Abweichung des Regenfalls von 1996 vom historischen Durchschnitt) und den Anteil der Erwachsenen pro Haushalt mit Malaria in den letzten 2 Wochen in 1994. Es ist positiv zu bewerten, dass die Autoren externe Schockdaten in ihre Studie einbeziehen. So kann in *trockenen Gebieten* die Variable für die Abweichung des Niederschlags einen negativen Wetterschock abbilden, während in *feuchten Gebieten* die Variable Malaria einen negativen Gesundheitsschock abbilden kann. Die Hinzunahme der Interaktionsterme bewirkt, dass nicht alle Haushalte in gleichem Maße von den Schocks betroffen sind. Als Beispiel nennen die Autoren, dass das Ausmaß einer Dürre auf die Wohlfahrt eines Haushalts davon abhängt, in welchem Umfang seine Felder bewässert werden und welches Vermögen ihm zur Verfügung steht. Auf diese Weise könnte die Schätzgenauigkeit der armutsgefährdeten Haushalte verbessert werden. Andererseits bietet diese Form bessere Möglichkeiten, die Wirkung von Maßnahmen zu bewerten, etwa um wieviel die Armutsgefährdung durch Bekämpfung von Malaria reduziert werden kann. Es ist allerdings zu bemerken, dass die empirische Spezifikation von Wetter- und Gesundheitsschocks mit Schwierigkeiten verbunden ist, da diese von Jahr zu Jahr variieren können. Extreme Schocks treten nur sehr selten auf, so dass solche Daten bestenfalls als worst case Szenario verwendet werden können. Zusätzlich zur Vulnerabilität als Wahrscheinlichkeit in $t+1$ in Armut zu fallen ($\alpha = 0$) berechnen die Autoren auch die Tiefgründigkeit der möglichen Armut ($\alpha = 1$ und $\alpha = 2$).

Kamanou und Modurch (2002) schätzen die Vulnerabilität als Differenz von erwarteter Armut und aktueller Armut (siehe Gleichung (2.16)).

$$V_t(\alpha) = E\iota(\alpha, y_{h,t+1}, z) - \iota(\alpha, y_{h,t+1}, z) \tag{2.16}$$

Für den Parameter $\alpha = 0$ ergibt sich zur Berechnung der Vulnerabilität die Definition (2.17), wobei ω ein Zustand der Volkswirtschaft, Ω der Zustandsraum, d.h. die Menge aller möglichen Zustände und H_t die Anzahl der Armen zum Zeitpunkt $t = 1$ ist.

$$V_1(0) = \frac{1}{N}\sum_{h=1}^{N}\sum_{\Omega} P(y^{\omega}_{h,t+1} < z|X_h)\iota(0, z, y_{0,t+1}) - \frac{H_t}{N} \tag{2.17}$$

In diesem Kontext verstehen die Autoren die Vulnerabilität als die erwartete Armut in $t+1$ subtrahiert um die aktuell Armen („isolate those households that did not become poor (or that did not become as poor) but that might have“).

Der Unterschied zu bisher betrachteten Studien ist, mit Ausnahme der Vulnerabilitätsdefiniton, dass verschiedene Zustände (*states of the world*) betrachtet wer-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

den. Die Verteilungen des Konsums in den jeweiligen Zuständen wird mit einem nichtparametrischen Verfahren, der Monte Carlo Bootstrap Methode (Efron und Tibshirani, 1994), bestimmt.

Die Autoren setzen zwei Panelwellen, benannt mit $t = 1$ und $t+1 = 2$ mit Informationen zu Konsum und Haushaltseigenschaften als gegeben voraus. Für das Jahr 2 nehmen sie an, dass die Verteilung des Konsums $P(y_{h,2})$ bekannt ist, jedoch nicht die gemeinsame Verteilungsfunktion $P(y_{h,2}, \omega)$, d.h. die Verteilung des Konsums in anderen möglichen Umweltzuständen. Diese wird mittels der Monte Carlo Bootstrap Methode (MCB), wie im folgenden Algorithmus dargestellt, geschätzt. Die Idee des Algorithmus besteht darin, eine Verteilung von möglichen Konsumausgaben für jeden Haushalt zu ziehen, die charakteristisch für Haushalte mit gegebenen Haushaltseigenschaften und der Konsumentwicklung von „ähnlichen Haushalten“ ist.

Algorithmus

i. Ziehe B Bootstrap Stichprobe aus den Originaldaten; diese bestehen aus zwei Panelwellen mit jeweils n Haushalten.

ii. Für jede Bootstrap Stichprobe wird die folgende Regression geschätzt:

$$\delta_h = X_h\beta + e, \quad \text{wobei } \delta_h = \frac{y^b_{h,2} - y^b_{h,1}}{y^b_{h,1}}$$

iii. Schätzung des Konsums in $t+1$ mit dem Einkommen aus t, der geschätzten Veränderung $\hat{\delta}$ und einem simulierten Fehler e_{mc}.

$$\hat{y}^{mcb}_{h,2} = y^b_{h,1}(1 + \hat{\delta}_h + e_{mc})$$

iv. Berechne den Armutsindex für jede Bootstrap-Stichprobe.

$$\hat{\iota}^{mcb}_{t+1} = \frac{H_b}{N}$$

v. Berechne den erwarteten Index aller Bootstrap-Stichproben.

$$\hat{E}\iota_{t+1} = \frac{1}{B}\sum_{b=1}^{B} \hat{\iota}^{mcb}_{t+1}$$

Positiv ist bei dem Ansatz von Kamanou und Modurch (2002), dass verschiedene *states of the world* betrachtet werden. Allerdings kommen als Schocks nur solche auf der Haushaltsebene in Frage, da für Makroschocks längere Zeitreihen benötigt würden. Als negativ ist zu bewerten, dass die gemeinsamen Verteilungen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

für $P(y_{h,2}, \omega)$ bestimmt werden, obwohl die Realisationen des Konsums für Welle 2 bereits bekannt sind. Eine mögliche Verbesserung wäre, die Ergebnisse in Welle 3 zu projezieren (siehe Abbildung 2.5). In dieser Hinsicht könnte die Betrachtung von relativen Konsumveränderungen, wie sie die Autoren verwenden, von Vorteil sein. Wird angenommen, dass die relativen Konsumänderungen δ_h konstant über die Zeit sind, so könnte mit dieser, sowie den gezogenen Konsumniveaus aus Welle 2 $y^b_{h,2}$, ein Bootstrap-Konsumniveau für Welle 3 $y^b_{h,3}$ konstruiert werden. Wird das Vorgehen sehr häufig durchgeführt, ließe sich die Verteilung $P(y_{h,3}, \omega)$ für Welle 3 konstruieren und die Vulnerabilität berechnen. Chaudhuri (2003) kritisiert weiterhin, dass die Schocks von verschiedenen Haushalten aus derselben Verteilung gezogen werden.

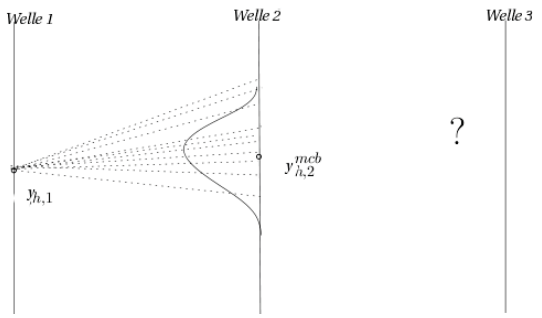


Abbildung 2.5: *Mögliche Modifizierung der Methode von Kamanou/Modurch (2002)*

Die einzige Studie, die die Vulnerabilität über mehrere Perioden schätzt, ist die von Pritchett et al. (2000). Über einen Zeithorizont von n Jahren wird ein Haushalt als armutsgefährdet definiert, wenn die Wahrscheinlichkeit, in mindestens einem der n Jahre, unter die (Konsum-)Armutslinie zu fallen eine festgelegte Grenzwahrscheinlichkeit übersteigt. Vorausgesetzt, dass die Schocks (im Konsum) unabhängig sind, lässt sich die Vulnerabilität wie in Gleichung (2.18) darstellen:

$$V_{h,n} \;=\; 1 \;-\; \Pr(y_{h,t+1} > z)\,\Pr(y_{h,t+2} > z)\,\cdot\cdots\cdot\,\Pr(y_{t+n} > z) \qquad (2.18)$$

Die Schätzung der Vulnerabilität in der nächsten Periode ist bereits mit vielen Schwierigkeiten verbunden, für die es bisher keine zufriedenstellende Lösung gibt. Die Schätzung der Vulnerabilität für n Perioden ist natürlich mit noch größeren Schwierigkeiten verbunden. Statt den Konsum für die nächste Periode zu schätzen, muss dieser nun für n Perioden in der Zukunft geschätzt werden. Wenn Kovariablen (z.B. Haushaltskovariablen, Marovariablen) genutzt werden sollen, müssen diese

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ebenso für n Perioden geschätzt werden. Dazu sind lange Paneldaten notwendig. Die Autoren schätzen jedoch im empirischen Teil ihrer Studie Vulnerabilität über drei Jahre mit lediglich zwei Panelwellen.

Für den Ansatz von Pritchett et al. (2000) müssen strikte Annahmen getroffen werden. Zunächst muss angenommen werden, dass die Ereignisse Armut bzw. keine Armut über die n Perioden unabhängig und identisch verteilt sind. Außerdem wird angenommen, dass die Differenzen im Konsum unabhängig und identisch normalverteilt sind. Unter diesen Annahmen ergibt sich die Vulnerabilität wie in Gleichung (2.19), wobei $g(\cdot)$ eine Funktion der Differenzen des Konsums darstellt.

$$V_{h,n}(n, z, e, \sigma) = 1 - \left(1 - \Phi\left(\frac{z - g(y_h)}{\sigma}\right)\right)^n \tag{2.19}$$

In einer Vielzahl der Studien (z.B. Chaudhuri et al. (2002), Pritchett et al. (2000)) wird ein Haushalt als armutsgefährdet definiert, wenn $\hat{V}_h > 0.5$, d.h., wenn die Wahrscheinlichkeit in Armut zu fallen größer ist als nicht arm zu bleiben. Alternativ definiert Chaudhuri et al. (2002) Haushalte als armutsgefährdet, wenn die Vulnerabilität die beobachtete Armutsquote übersteigt. Beide Vorgehensweisen sind willkürlich gewählt.

VEU

Ligon und Schechter (2003) schätzen mit dem Ansatz VEU die Vulnerabilität auf Basis des Household Budget Survey (HBS) für Bulgarien mit monatlichen (Panel-) Daten von 1994. Die Schätzung basiert nicht auf der Schätzung des Konsums in verschiedenen Umweltzuständen im Kontext der Definition (2.20) sondern mit monatlichen Realisationen des Konsums. Die Autoren zerlegen zur Schätzung der Vulnerabilität die Risikokomponente in Gleichung (2.4) weiter in aggregierte Risiken, haushaltspezifische Risiken und unerklärtes Risiko bzw. Messfehler, siehe Gleichung (2.20). Für diese Darstellung werden haushaltspezifische und aggregierte Haushaltskovariablen $X_{h,t}$ bzw. $\overline{X}_t$ eingeführt. Die Schätzung der Terme in (2.20) erfolgt unter gewissen Annahmen einzeln.

$$\begin{aligned} V_{h,t} = & \overbrace{[U_h(z_{CE}) - U_h(Ey_{h,t})]}^{Armut} \\ & + \overbrace{[U_h(Ey_{h,t}) - EU_h[E(y_{h,t}|\overline{X}_t)]]}^{aggregiertes\ Risiko} \\ & + \overbrace{[EU_h[E(y_{h,t}|\overline{x}_t)] - EU_h[E(y_{h,t}|\overline{X}_t, X_{h,t})]]}^{haushaltsspezifisches\ Risiko} \\ & + \overbrace{[EU_h[E(y_{h,t}|\overline{X}_t, X_{h,t})] - EU_h(y_{h,t})]}^{\text{unerklärtes}\ Risiko\ und\ Messfehler} \end{aligned} \tag{2.20}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Wie bereits erwähnt, wird z_{CE} als Mittelwert des Konsums aller Haushalte gewählt und zu 1 normiert, d.h. $z_{CE} = \frac{1}{N}\frac{1}{T}\sum_{h=1}^{N}\sum_{t=1}^{T} y_{h,t} = 1$. Unter Annahme von Stationarität des Konsums über die Zeit ergibt sich $\hat{E}y_{h,t} = \frac{1}{T}\sum_{t=1}^{T} y_{h,t}$. Weiterhin wird angenommen, dass der Konsum $y_{h,t}$ mit einem Messfehler $\{e_{h,t}\}$ behaftet ist. Der Prozess mit Messfehler wird mit $\tilde{y}_{h,t}$ bezeichnet und ist definiert als $y_{h,t} + e_{h,t}$. Es wird angenommen, dass der Prozess $e_{h,t}$ die Eigenschaft $E(e_{h,t}|\overline{X}_t, X_{h,t}) = E(e_{h,t}y_{h,t}) = 0$ hat, so dass $E(\tilde{y}_{h,t}|\overline{X}_t, X_{h,t}) = E(y_{h,t}|\overline{X}_t, X_{h,t})$ gilt. Diese Annahme führt dazu, dass der Messfehler nur das unerklärte Risiko beeinflusst. Ligon und Schechter (2003) schlagen als Parametrisierung von $E(y_{h,t}|\overline{X}_t, X_{h,t})$ den Ausdruck in Gleichung (2.21) vor, wobei ν_h haushaltsspezifische Effekte auf das vorhergesagte Einkommen sind, die sich zu 0 summieren, ξ_t sind zeitspezifische Effekte, die die Veränderung des Konsums der Population beschreiben und β ist ein Vektor von Regressionskoeffizienten, die die Veränderungen von Haushaltseigenschaften auf den Konsum eines Haushalts beschreibt.

$$\hat{E}(y_{h,t}|\overline{X}_t, X_{h,t}) = \nu_h + \xi_t + X'_{h,t}\beta \; mit \; \theta = (\nu_h, \xi_t, \beta) \tag{2.21}$$

Ligon und Schechter (2003) wählen als Parameter λ für die in Gleichung (2.5) definierte Nutzenfunktion $\lambda = 2$, was vergleichbar zur quadrierten Armutslücke im Konzept VEP ist (Cunningham und Maloney, 2000). Für Bulgarien finden die Autoren in 1994 ein Vulnerabilitätsmaß von 20%. Aufgrund der Normierung ist dies wie folgt zu interpretieren: Der Nutzen eines durchschnittlichen Haushalts fällt um 20% niedriger aus als unter einer Konsumverteilung ohne Ungleichheit und Risiko. Für die Zerlegung des Vulnerabilitätsmaßes ergibt sich, dass Armut den größten Anteil (über die Hälfte) ausmacht, gefolgt von dem unerklärten Risiko inklusive des Messfehlers. Weiterhin hat das aggregierte Risiko eine größere Bedeutung als das idiosynkratische Risiko. Haushalte, deren Haushaltsvorstand männlich, gebildet und erwerbstätig ist, unterliegen weniger Risiken als andere Haushalte.

Das Vorgehen von Ligon und Schechter (2003) ist insofern kritisch zu sehen, dass sie zwar Paneldaten verwenden, aber trotzdem kein ex ante Maß für die Vulnerabilität berechnen. Statt den Konsum in verschiedenen Umweltzuständen zu schätzen, wie sie es in der Definition (2.20) vorgeschlagen haben, zerlegen sie ein Maß, das aus den Realisationen ihrer Daten berechnet ist, in eine ex post Komponente Armut und mehrere Risikokomponenten. Dieses Vorgehen setzt voraus, dass ihr Maß über die Zeit konstant ist.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Perceived Vulnerability to Downside Risk

Povel (2010) folgt dem Ansatz von Calvo und Dercon (2005) und definiert die Vulnerabilität als „probability-weighted average of state specific deprivation index“ (siehe Definition (2.22)).

$$V_h = \sum_{i=1}^{|\Omega|} p_{hi} d_{h,i}^{\alpha}$$
$$0 \leq d_{h,i} \leq 1 \quad \sum_{i=1}^{|\Omega|} p_{h,i} = 1 \tag{2.22}$$

In dieser Definition ist p_{hi} die Eintrittswahrscheinlichkeit eines möglichen Umweltzustands in der Zukunft und d_{hi} der Verlust („deprivation index“), den dieser Zustand bei dem Haushalt hervorgerufen hat. Die Variable α ist wieder ein Parameter für die Risikoeinstellung des Haushalts und $|\Omega|$ die Anzahl der möglichen Zustände in der Zukunft.

Im Gegensatz zu anderen Studien verwendet der Autor als Armutsgrenze die aktuelle Wohlfahrt eines Haushalts. Als Begründung führt der Autor an, dass eine feste Armutslinie für alle Haushalte dem Sachverhalt nicht gerecht wird („... households without a probability of falling below it are labeled *non vulnerable* regardless of how much their level of wellbeing may decrease in future. Consequently, the question arises why a household which is likely to fall from high above the poverty line to very close above it should be entirely non vulnerable whereas a household which is likely to fall much less, namely from right above the poverty line to right below it, is vulnerable?“). Die neue Definition der Armutsgrenze führt auch zu einer neuen Interpretation des Fokus Axioms (siehe Studie Povel (2010)).

Anders als bei den existierenden Studien über die Vulnerabilität ist auch, dass der Autor die Analysen nicht auf vergangenen Daten sondern auf eigens erhobenen Daten basiert, die Haushalte zu den subjektiv erwarteten Eintritten von elf ausgewählten Risiken und deren Intensität in den nächsten fünf Jahren befragt. Zur Bewertung dieses Vorgehens argumentiert der Autor wie folgt: „It is not crucial whether subjective risk perception or information about past shocks is better in predicting the future, but the question which one of both has a better impact on household's wellbeing and economic behavior“. Als Antwortmöglichkeiten bestehen für die Eintrittshäufigkeiten sechs Ausgänge und für die Intensität vier. Den Antworten wurden Werte (0%,20%,40%,60%, 80%,100%) und (0,0.33,0.66,1) zugewiesen. Ein Risiko wird als „downside risk“ X_h bezeichnet, wenn die subjektiv erwartete Eintrittswahrscheinlichkeit zwischen 0 und 100% liegt und als “downside

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

certainty“ C_h, wenn diese sicher erwartet oder nicht erwartet wird. Damit wird ein Haushalt mit $X_h + C_h$ möglichen Ereignissen in der Zukunft konfrontiert. Die Anzahl der möglichen Zustände in der Zukunft $|\Omega|$ wird vom Autor als Summe der Binomialkoeffizienten $\binom{X_h}{j}$ berechnet (siehe Gleichung (2.23)), wobei j die Kombination von Risiken ist.

$$|\Omega| = \sum_{j=0}^{X_h} \frac{X_h!}{(X_h - j)!j!} = 2^{X_h} \tag{2.23}$$

Der „deprivation index“ ergibt sich aus der Summe der Intensitäten aller „downside certainties“ $s_{h,k}$ und aller “downside risks“ $s_{h,i,j}$ wie in Gleichung (2.24).

$$d_{h,i} = \sum_{k=1}^{c_h} s_{h,k} + \sum_{j=0}^{X_{h,i}} s_{h,i,j} \tag{2.24}$$

Zur Berechnung der Eintrittswahrscheinlichkeit eines Zustands i ist anzunehmen, dass die Eintrittswahrscheinlichkeiten voneinander abhängig sind, so dass eine bedingte Wahrscheinlichkeit verwendet werden müsste. Trotzdem nimmt Povel (2010) vereinfachend an, dass die Risiken voneinander unabängig sind.

Die meisten Studien sagen auf Basis von Haushaltseigenschaften einen Indikator (etwa die Armutswahrscheinlichkeit oder das Einkommen) voraus und definieren einen Haushalt auf dieser Vorhersage als armutsgefährdet oder nicht-armutsgefährdet. Povel (2010) dagegen berechnet mit seinen subjektiv erhobenen Daten einen Vulnerabilitätsindex und regressiert diese auf eine Reihe von Haushaltskovariablen, um deren Erklärungsgrad der Vulnerabilität zu ermitteln. Da der Vulnerabilitätsindex linksseitig beschränkt ist, wird das Tobit Modell verwendet. Der Autor findet in Studien mit Daten aus Vietnam und Thailand, dass Vietnam anfälliger gegenüber „perceived vulnerability to downside risk“ ist als Thailand (für detailliertere Ergebnisse siehe Studie).

Der Autor verwendet für seine Studie Querschnittsdaten. Eine interessante Frage ist es, wie das Vorgehen ist, wenn dieses auf Paneldaten übertragen würde. Ändert sich die Armutsgrenze in jeder Zeitperiode oder gibt es einen Ausgangszustand? Das Vorgehen von Povel kann eindeutig als ex ante bezeichnet werden, da die Haushalte zu ihrer Risikoeinschätzung in den nächsten fünf Jahren befragt werden. Aber andererseits basiert die Studie auf subjektiven Daten. Argumente gegen die Verwendung von Risikowahrnehmungen, wie Povel (2010) selbst aufführt, sind die „heuristics of probabilities“ (Botteril und Mazur (2004), Kahnemann, Slovic and Tversky(1982)).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2.3 Bewertung der Genauigkeit von Schätzern mit ROC-Kurven

In der Literatur über die Vulnerabilität wurde der Beurteilung der Genauigkeit von Vulnerabilitäts-Schätzern bisher relativ wenig Aufmerksamkeit gezollt. Ausnahmen stellen die Autoren Ligon und Schechter (2004), Jha und Dang (2010) und Zhang und Wan (2009) dar.

Um die Genauigkeit von verschiedenen Vulnerabilitäts-Schätzern zu vergleichen, führen Ligon und Schechter (2004) Monte Carlo Experimente von künstlich erzeugten Datensätzen („data generating process by Jalan and Ravallion (1999)“), in denen Werte für die zukünftige Wohlfahrt bekannt sind, unter verschiedenen Bedingungen (Stationarität, Nichtstationarität, Vorhandensein oder Abwesenheit von Messfehlern) durch. Das Vorgehen zur Erzeugung ihrer Vulnerabilitäts-Schätzer erfolgt in zwei Schritten: Im ersten Schritt schätzen die Autoren den Konsum mit verschiedenen Schätzern (Pritchett et al. (2000), Schätzer mit Querschnittsdaten, Chaudhuri (2001), Ligon und Schechter (2003)). Im zweiten Schritt wenden sie auf jeden Schätzer des Konsums verschiedene Vulnerabilitätsmaße (Foster-Greer-Thorbecke Maße (Foster et al.,1984) für $\alpha \in \{0, 1, 2\}$, Ligon und Schechter Maß (Ligon und Schechter, 2003) und Calvo und Dercon Maß (Calvo und Dercon, 2003)) an. Die Genauigkeit dieser Vulnerabilitäts-Schätzer messen die Autoren durch den (logarithmischen) mittleren quadratischen Fehler (Mean Squared Error, MSE) und den Spearman Rangkorrelationskoeffizienten als eine Funktion der Größe der Panelstichprobe und der Anzahl der Panelwellen. Sie kommen zu dem Ergebnis, dass Schätzer von Querschnittsdaten unter allen Bedingungen eine geringe Genauigkeit haben. Unter der Annahme der Stationarität ist unabhängig vom Vulnerabilitätsmaß der Schätzer von Chaudhuri (2001) am besten, wenn der Konsum nicht mit Fehler gemessen wird. Ist das Vulnerabilitätsmaß dagegen „risk-sensitive“ (steigendes Risiko führt zu einem Anstieg des Vulnerabilitätsmaßes) und der Konsum mit Fehler behaftet, so hat der Schätzer von Ligon uns Schechter (2003) die besten Eigenschaften. Die Autoren ermitteln weiterhin, dass unter Annahme von Nicht-Stationarität der Schätzer von Pritchett et al. (2000) am besten ist. Allerdings stützen sich ihre Ergebnisse auf simulierte Daten, was, wie sie zugeben „of course assume away the complexity of the real-world data“. Ihre Analysen basierend auf der Vietnamesischen Haushaltsbefragung (2 Perioden) und der Bulgarischen Haushaltsbefragung (12 Perioden) vergleichen die Beziehungen zwischen verschiedenen Vulnerabilitätsmaßen, d.h. die Korrelation zwischen diesen. Der Schätzer, der als bester in ihren Experimenten identifiziert wurde, wurde auf die Befragungen an-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

gewendet und das erhaltene Vulnerabilitätsmaß wurde in Armut, aggregiertes und spezifisches Risiko und Messfehler zerlegt. Diese Zerlegung misst allerdings nicht die Genauigkeit der Vulnerabilität, d.h.welche Haushalte arm geworden sind und welche nicht.

Jha und Dang (2010) messen die Genauigkeit als den Anteil von Haushalten in der Stichprobe, die richtig klassifiziert wurden. Auf dieser Grundlage stellen sie fest, dass die Vulnerabilitäts-Schätzer „einigermaßen gute Arbeit leisten“. Allerdings ist der Anteil der richtigen Klassifizierungen ein irreführendes Maß der Qualität einer Diagnosemethode, da es stark von den Anteilen der armen und nicht-armen Haushalte in der Bevölkerung abhängt. Die Receiver Operating Characteristic (ROC), die unten erläutert wird, wurde speziell dafür konstruiert, um die Genauigkeit solcher Diagnosen zu quantifizieren.

Zhang und Wan (2009) untersuchen den Einfluss von verschiedenen Einkommensmaßen, Armuts- und Vulnerabilitätslinien auf die Genauigkeit von Vulnerabilitäts-Schätzern. Sie definieren Genauigkeit als den Anteil der richtig als armutsgefährdet klassifizierten Haushalte an allen als armutsgefährdet klassifizierten Haushalten. Dieses Maß berücksichtigt, im Gegensatz zur ROC, nicht die Haushalte, die arm werden aber als nicht-armutsgefährdet klassifiziert wurden.

Die ROC ist ein sehr etabliertes Werkzeug, das die Genauigkeit einer Diagnosemethode quantitativ misst und damit einen Vergleich verschiedener Verfahren ermöglicht. Es findet in vielen Fachrichtungen Anwendung (siehe z.B. Egan (1975), Spackman (1989), Thompson and Zucchini (1989), Swets et al. (2000), Fawcett (2006)). Im medizinischen Kontext wird es verwendet, um den Zielkonflikt zwischen der *Sensitivität* oder True Positive Rate (TPR, d.h. die Wahrscheinlichkeit, einen kranken Patienten richtig zu identifizieren) und der *Spezifität* (die Wahrscheinlichkeit, einen nicht-kranken Patienten richtig zu identifizieren); letzteres ist gleich zu eins subtrahiert um die „Fehlalarmhäufigkeit“, d.h. 1 - False Positive Rate (FPR). In dieser Arbeit soll bewertet werden, wie gut verschiedene Methoden die zukünftige Armut vorhersagen können, d.h. wie genau sie vorhersagen können, welche Haushalte arm werden und welche nicht.

Eine Diagnose läuft wie folgt ab: Zunächst wird das zukünftige Einkommen mit Modellen, die im nächsten Abschnitt vorgestellt werden, vorhergesagt. Dann wird eine Vulnerabilitätslinie (VPL) festgelegt, so dass Haushalte, die unter diese Grenze fallen, als armutsgefährdet klassifiziert werden und alle anderen als nicht-armutsgefährdet. Folglich sind vier verschiedene Kombinationen von Diagnoseergebnissen möglich, die in der Kontingenztabelle in Abbildung 2.6 dargestellt werden.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Tatsächliche Klasse: arm	Tatsächliche Klasse: nicht-arm
Vermutete Klasse: armutsgefährdet	TRUE POSITIVES	FALSE POSITIVES
Vermutete Klasse: nicht-armutsgefährdet	FALSE NEGATIVES	TRUE NEGATIVES
	A	N

Abbildung 2.6: *Kontigenztabelle der Diagnoseergebnisse [Darstellung nach Fawcett (2006)].*

Indem die VPL verändert wird, können die Anteile der Fehler (false negatives und false positives) abgestimmt werden. Die ROC ist ein Graph, in dem die FPR auf der x-Achse gegen die TPR auf der y-Achse aufgetragen wird. Sie ist eine nichtfallende Funktion, die im Punkt (0,0) beginnt und im Punkt (1,1) endet. Grob lässt sich sagen, dass ein Verfahren umso besser ist, je schneller die Kurve $y = 1$ erreicht. Also ist die Methode optimal, die sofort $y = 1$ erreicht, d.h. durch den Punkt (0,1) verläuft. Die ROC ermöglicht es, die im nächsten Abschnitt vorgestellten Modelle zu vergleichen, wobei als VPLs die Einkommen (in aufsteigender Sortierung) der Reihe nach verwendet werden. Die VPLs für verschiedene Modelle müssen nicht dieselben sein.

Als Qualitätsmaße für die ROC gibt es verschiedene Möglichkeiten, von denen drei in Abbildung 2.7 dargestellt sind. Das am häufigsten verwendete ist die Fläche unter der Kurve (Area under the Curve, AUC), d.h. die Performance der Kurve wird auf einen einzelnen Wert reduziert, der zwischen 0 und 1 liegt. Eine Methode wird als besser gegenüber einer andere angesehen, wenn seine AUC größer ist; optimal also wenn sie den Flächeninhalt 1 hat. Der Nachteil, AUC zur Bewertung der Modelle von Vulnerabilitäts-Schätzern zu betrachten, ist, dass mit dem Qualitätsmaß AUC die gesamte Spanne von TPRs verglichen wird. Um die Vulnerabilität zu schätzen, ist es irrelevant, ob eine Methode besser als eine andere für kleine Werte der TPR ist, z.B. ist eine Methode, die nur 10% der zukünftig armen Haushalte identifizieren kann ($TPR = 10\%$), nicht von Nutzen.

Nur große Werte der TPR sind von praktischem Interesse. Das bedeutet, dass AUC als Qualitätsmaß in diesem Kontext nicht genügt. Dasselbe gilt für Seltens Maß

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

des „voraussagenden Erfolges“ (Selten und Krischker (1983), Selten (1991)). Eine Möglichkeit zur Verbesserung des Maßes AUC für diesen Zusammenhang besteht darin, die Spanne der TPR bis zu einem Wert, z.B. $FPR = 25\%$, zu beschränken und als Qualitätsmaß nur eine Teilfläche unter der ROC zu betrachten (Partial Area under the Curve, PAUC) (Thompson und Zucchini, 1989).

In dieser Arbeit werden Methoden aufgrund der FPR, die bei einer vorher festgelegten TPR eintritt, bewertet (Fienberg und Stern, 2005). Meistens wird als TPR 80% und in einigen Fällen 90% verwendet, d.h. es sollen mindestens 80% bzw. 90% der Haushalte identifiziert werden, die im kommenden Jahr arm werden. Der Wert $TPR = 80\%$ ergibt rückwirkend die VPL, die wiederum die FPR bestimmt. Die FPR stellt die Kosten dar, die in Kauf genommen werden müssen, um die TPR von 80% zu erreichen. Eine Methode wird gegenüber einer anderen als besser angesehen, wenn die FPR dieser Methode (für die festgelegte TPR) kleiner ist als die FPR der anderen Methode. Wenn die FPR einer Methode bei TPR=80% sehr hoch ist, z.B. 40%, dann ist die absolute Anzahl der falsch als armutsgefährdeten Haushalte möglicherweise zu groß, so dass die Methode keinen praktischen Nutzen zur Schätzung der Vulnerabilität erbringt.

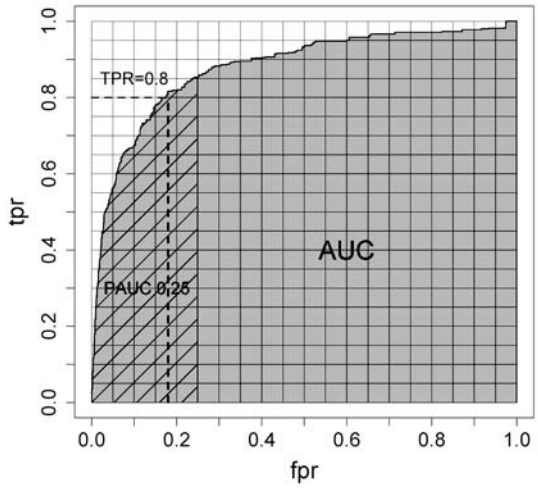


Abbildung 2.7: *Qualitätsmaße für die ROC-Kurven: AUC (graue Fläche), PAUC (schraffierte Fläche), FPR (=0.17) korresponierend zu TPR=0.8 [Darstellung (Thompson und Zucchini, 1989) auf Basis der in Kapitel 3 beschriebenen Daten].*

Bisher wurde davon ausgegangen, dass die VPL ex post so festgelegt wird, dass eine TPR von 80% erzielt wird. Um die Vulnerabilität zu schätzen, muss die VPL

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

natürlich ex ante fixiert werden. Idealerweise sollte dies so geschehen, dass die TPR möglichst konstant bei 80% liegt. Das Vorgehen besteht darin, dass für rollierende Panels die Vulnerabilität geschätzt wird, und auf Basis der historischen VPLs bei 80% TPR eine Festlegung getroffen wird.

Das Vorgehen lässt sich dem Konzept VEP zuordnen.

2.4 Schätzmethoden für verschiedene Situationen der Datenverfügbarkeit

Im vorhergehenden Abschnitt wurde erläutert, was in dieser Arbeit unter der Genauigkeit von Vulnerabilitäts-Schätzern zu verstehen ist und wie das Vorgehen zur Schätzung der Vulnerabilität ist. Der erste Schritt besteht in der Prognose eines Indikators, in dieser Arbeit das Einkommen bzw. die Armutswahrscheinlichkeit der Haushalte in der nächsten Periode. Dieser Abschnitt widmet sich der Vorstellung und Beschreibung der Prognosemodelle, die in Kapitel 4 angewendet werden und die zusammen mit der VPL die Grundlage zur Schätzung der Vulnerabilität darstellen.

Die verwendeten Modelle sind korrelationsbasierte Modelle, deren Ziel sich darauf beschränkt, das aktuelle Einkommen bzw. die Armutswahrscheinlichkeit vorherzusagen und keine „kausalen“ Modelle, die die Entwicklung des Einkommens erklären. Die Modellwahl in dieser Arbeit ist, u.a. wegen der Betrachtung von Teildatensätzen (näheres dazu in Kapitel 4), mit Schwierigkeiten verbunden. Insbesondere für den ersten Teil der Analyse wurden weitere Regressionsmodelle u.a. mit logarithmischen Response und Heteroskedastizität, wie sie in den Studien in der Literatur Anwendung finden, untersucht. Da diese aber weniger geeignet bzw. zu keinen besseren Ergebnissen führten, wurde auf die einfacheren Modelle aus Abschnitt 2.4.1 zurückgegriffen. Eine Darstellung der Ergebnisse der anderen Modelle wird in Kapitel 4 gegeben.

Die Anpassung von Modellen hängt stark von der Verfügbarkeit der Daten ab. In dieser Arbeit soll eine möglichst umfassende Beschreibung verschiedener Situationen dargestellt werden. Abschnitt 2.4.1 stellt verschiedene lineare Regressionsmodelle zur Schätzung des aktuellen Einkommens mit Querschnitts- und Paneldaten unter der Annahme bereit, dass stetiges Einkommen (als Response in Querschnittsdatensätzen bzw. als Response und Kovariable in Paneldatensätzen) vorhanden ist. In Abschnitt 2.4.2 wird die Situation modelliert, dass zwar Panel-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

daten verfügbar sind, aber dass das Einkommen nur in Klassen vorhanden ist. Dazu werden verschiedene logistische Regressionsmodelle verwendet, die die bedingte Wahrscheinlichkeit, dass Haushalte in der aktuellen Periode in Armut fallen, schätzen. Abschnitt 2.4.3 betrachtet ein Querschnitts- und ein Panelmodell für Merkmalsgruppen getrennt.

2.4.1 Verfügbarkeit von Querschnitts- oder Paneldaten

Es werden sechs verschiedene Modelle betrachtet, um die Vulnerabilität der Armut zu schätzen. Die Modelle unterscheiden sich nach Verfügbarkeit der Art der Datensätze: Querschnittsdaten (P_1-P_3) oder Paneldaten (P_4-P_6) und den Kovariablen, die verfügbar sind: nur Haushaltskovariablen (P_1-P_2), nur Einkommen (P_3), Haushaltskovariablen und Vorjahreseinkommen (P_4-P_6). Einige der Modelle basieren auf den aktuellen Haushaltseigenschaften (P_2, P_4, P_5) und einige auf denen des Vorjahres (P_1, P_6). Die Modelle P_4 und P_5 unterscheiden sich im Regressionskoeffizienten $\hat{\gamma}$ der Einkommenskovariable; in P_5 wird $\hat{\gamma} = 1$ gewählt, so dass die Haushaltskovariablen die Veränderungen des Einkommens in aufeinanderfolgenden Jahren erklären.

Die Schätzer für Einkommen im Jahr t, d.h. $\hat{y}_{h,t}$, gemäß dieser sechs Modelle lauten:

$$\hat{y}_{h,t} = X_{h,t-1}\hat{\beta} \tag{P_1}$$

$$\hat{y}_{h,t} = X_{h,t}\hat{\beta} \tag{P_2}$$

$$\hat{y}_{h,t} = y_{h,t-1} \tag{P_3}$$

$$\hat{y}_{h,t} = X_{h,t}\hat{\beta} + y_{h,t-1}\hat{\gamma} \tag{P_4}$$

$$\hat{y}_{h,t} = X_{h,t}\hat{\beta} + y_{h,t-1} \tag{P_5}$$

$$\hat{y}_{h,t} = X_{h,t-1}\hat{\beta} + y_{h,t-1}\hat{\gamma} \tag{P_6}$$

wobei $X_{h,t}$ die Kovariable für Haushalt h zum Zeitpunkt t, $\hat{\beta}$ und $\hat{\gamma}$ die geschätzten Regressionskoeffizienten für die Haushaltskovariablen bzw. das Vorjahres-Einkommen und $e_{h,t}$ die Residuen für Haushalt h zum Zeitpunkt t repräsentieren. Die Regressionskoeffizienten wurden mit ordinary least squares (OLS) geschätzt. Für alle sechs Modelle wird angenommen, dass die Beobachtungen unabhängig und identisch verteilt und dass die Residuen normalverteilt mit Erwartungswert Null und konstanter Varianz sind (Fahrmeir et al., 2007, S. 159ff). Es liegen keine Infor-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

mationen vor, die die Annahme der identischen Verteilung widerspricht. Die Beobachtungen können als unabhängig angenommen werden, da, wie später erläutert, nur Haushalte betrachtet werden, die in allen Wellen des analysierten, kurzen Panels anwesend sind. D.h., dass von Haushalten, die sich im Beobachtungszeitraum teilen, nur ein Teil (mit der ursprünglichen Haushaltsnummer und aktuellen Haushaltsnummer), in der Analyse betrachtet wird. Die Annahme der Normalverteilung wird im Abschnitt 4.1.1 bzw. Anhang A.2.2 untersucht.

Ein Haushalt wird für eines der Modelle P_1-P_6 als armutsgefährdet bezeichnet, wenn das geschätzte Einkommen $\hat{y}_{h,t}$ unter einer vorher festgelegten VPL liegt, die sich für jedes Modell unterscheiden kann. Um diese Modelle anzuwenden, muss zunächst (mit Ausnahme von P_3) berechnet werden, welche Kovariablen am besten geeignet sind, um das Einkommen vorherzusagen. Die Variablenauswahl wird in Kapitel 4 betrachtet.

2.4.2 Verfügbarkeit von Paneldaten mit Einkommen in Klassen

In der vorliegenden Situation mit Einkommensklassen wird nicht, wie im vorhergehenden Abschnitt, das aktuelle bedingte Einkommen sondern die Wahrscheinlichkeit, in der aktuellen Periode in Armut zu fallen, bedingt durch die Zugehörigkeit zu einer gewissen Einkommensklasse und die Eigenschaften des Haushalts, vorhergesagt. Wie bereits erwähnt, sind diese beiden Vorgehensweisen äquivalent und führen zu denselben Ergebnissen. Zur Übersichtlichkeit wird im Folgenden in der Notation der Variablen im Index auf die Kennzeichnung des Haushalts durch h verzichtet.

Der Response wird also durch eine binäre Variable dargestellt, die durch Z_t in (2.25) definiert wird.

$$Z_t = \begin{cases} 1, & \text{wenn} \quad y_t < z \\ 0, & \text{wenn} \quad y_t \geq z. \end{cases} \tag{2.25}$$

Die Einkommensklassen werden durch das minimal und maximal mögliche Einkommen[4] sowie durch $K-1$ Einkommensgrenzen beschrieben, die mit $L_1, \ldots, L_{K-1}$ (im Fall von zwei Einkommensklassen mit L) notiert werden. Die Bewertung von

[4] In den empirischen Analysen wird das maximal mögliche Einkommen auf 30 000 € beschränkt. Näheres dazu in Kapitel 4.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Schätzern unter Verwendung von Einkommensklassen wird in dieser Arbeit für zwei und drei Einkommensklassen betrachtet.

$$y_{t-1,K} = \begin{cases} K-1 & \text{wenn} \quad y_{t-1} < L_1 \\ K-2 & \text{wenn} \quad L_1 \leq y_{t-1} < L_2 \\ \dots & \\ 1 & \text{wenn} \quad L_{K-2} \leq y_{t-1} < L_{K-1} \\ 0 & \text{wenn} \quad y_{t-1} \geq L_{K-1}. \end{cases} \tag{2.26}$$

Zur Schätzung der bedingten Armut bieten sich logistische Regressionsmodelle (McCullagh und Nelder, 1989, S. 107-114) an. Es werden zwei verschiedene Modelle betrachtet, um die Vulnerabilität der Armut mit Paneldaten und Einkommensklassen zu schätzen. In Modell M_{1K} wird die Variable *Einkommensklassen* als Faktor in die logistische Regression aufgenommen. Modell M_{2K} besteht aus K Teilmodellen (ein Modell für jede Einkommensklasse), so dass sich für jede Einkommensklasse i verschiedene Regressionskoeffizienten $\widehat{\beta}_i$, $i = 1, \dots, K$ ergeben.

$$\widehat{P}(Z_t = 1 | X_{t-1}, y_{t-1,K}) = \frac{\exp(X_{t-1}\widehat{\beta} + y_{t-1,K}\widehat{\gamma})}{1 + \exp(X_{t-1}\widehat{\beta} + y_{t-1,K}\widehat{\gamma})} \tag{M_{1K}}$$

Zur Schätzung M_{2K} wird der Datensatz in K Teildatensätze mit Haushalten derselben Einkommensgruppe aufgespalten. Nachdem die bedingten Wahrscheinlichkeiten durch die Teilmodelle getrennt geschätzt wurden, werden die Teildatensätze wieder zu einem Datensatz vereint, d.h. es wird *eine* ROC für die Beobachtungen aller Haushalte konstruiert.

$$\hat{P}(Z_t = 1 | X_{t-1}^{(1)}, y_{t-1,K} = K-1) = \frac{\exp(X_{t-1}^{(1)}\widehat{\beta}_1)}{1 + \exp(X_{t-1}^{(1)}\widehat{\beta}_1)}$$

$$\dots$$

$$\hat{P}(Z_t = 1 | X_{t-1}^{(K-1)}, y_{t-1,K} = 1) = \frac{\exp(X_{t-1}^{(K-1)}\widehat{\beta_{K-1}})}{1 + \exp(X_{t-1}^{(K-1)}\widehat{\beta_{K-1}})} \tag{M_{2K}}$$

$$\hat{P}(Z_t = 1 | X_{t-1}^{(K)}, y_{t-1,K} = 0) = \frac{\exp(X_{t-1}^{(K)}\widehat{\beta_K})}{1 + \exp(X_{t-1}^{(K)}\widehat{\beta_K})}$$

wobei

$$X_{t-1} = \begin{cases} X_{t-1}^{(1)} & \text{wenn} \quad y_{t-1,K} = K-1 \\ \dots & \\ X_{t-1}^{(K-1)} & \text{wenn} \quad y_{t-1,K} = 1 \\ X_{t-1}^{(K)} & \text{wenn} \quad y_{t-1,K} = 0 \end{cases}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Ein Haushalt wird für die Modelle M_{1K} und M_{2K} als armutsgefährdet bezeichnet, wenn die geschätzte Wahrscheinlichkeit $\widehat{P}(Z_t = 1|\dots)$ unter einer vorher festgelegten VPL, die sich für jedes Modell unterscheiden kann, liegt.

Da im SOEP stetiges Einkommen vorliegt, müssen für diese Analyse neben der VPL zusätzlich Einkommensklassen definiert werden. Dies ermöglicht es, die Genauigkeit der Vulnerabilitäts-Schätzer für verschiedene Wahlen der Einkommensklassen zu bestimmen und Kriterien für eine sinnvolle Wahl der Einkommensklassen zu finden. Das Vorgehen wird in Abschnitt 4.2 erläutert. Die Haushaltskovariablen werden aus dem ersten Teil übernommen.

2.4.3 Gruppen von Merkmalen

Es werden aus Abschnitt 2.4.1 zwei Modelle ausgewählt, um die Vulnerabilität der Armut in Gruppen von Merkmalen zu bestimmen: P_1 basierend auf Querschnittsdaten und P_6 basierend auf Paneldaten. Anders als in Abschnitt 2.4.1 wird in den Modellen ein logarithmisch transformierter Response (und das logarithmisch transformierte Vorjahreseinkommen als Kovariable) verwendet. Die Begründung dafür wird im analytischen Teil gegeben. Die Haushaltskovariablen sind ähnlich wie im ersten Teil, werden aber den Gegebenheiten der Untergruppen angepasst. Die Notation der Modelle aus 2.4.1 wird modifiziert zu:

$$E\left(\ln(y_{t,G})|X_{t-1,G}\right) = X_{t-1,G}\hat{\beta}_G \qquad (P_{1,G})$$

$$E\left(\ln(y_{t,G})|X_{t-1,G}, y_{t-1,G}\right) = X_{t-1,G}\hat{\beta}_G + \ln(y_{t-1,G})\hat{\gamma}_G \qquad (P_{6,G})$$

Bei der Schätzung der Vulnerabilität werden zwei verschiedene Vorgehensweisen betrachtet. Im ersten Ansatz werden die aktuellen Einkommen für Gruppen getrennt geschätzt, während die Vulnerabilität von allen Gruppen gemeinsam gemessen wird, d.h. bei 80% TPR aller geschätzten Einkommen der Haushalte (ähnlich wie bzgl. M_{2K} in Abschnitt 2.4.2). In diesem Fall wird angenommen, dass die Haushaltseigenschaften und die Einkommensverläufe in den Gruppen homogener sind und damit zu besseren Schätzungen führen, als wenn die Einkommen aller Haushalte gemeinsam geschätzt werden. Dieser Ansatz stellt eine Gesamtbetrachtung dar, der direkt die Frage beantwortet, inwiefern die Schätzung des Einkommens in Gruppen zu einer Verbesserung in der Messung der Vulnerabilität führt.

Der zweite Ansatz dagegen schätzt sowohl das Einkommen als auch die Vulnerabilität für die Gruppen einzeln, d.h. bei 80% TPR von den geschätzten Einkommen jeder Gruppe eines Merkmals. Das bedeutet, dass in Gruppen mit vielen armen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Haushalten (in der nächsten Periode) auch mehr Haushalte als armutsgefährdet identifiziert werden müssen als in Gruppen mit wenig armen Haushalten. Die Vorgehensweise beruht auf der Annahme, dass sich die Einkommen in getrennten Modellen besser schätzen lassen und auch die Verteilung der geschätzten Einkommen, und damit die Vulnerabilität zwischen den Gruppen unterschiedlich ist. Die Methode liefert das Ergebnis, wie gut die Vulnerabilität in den Gruppen gemessen wird und ggfs. für welche Gruppen die Schätzung verbessert werden sollte. Eine Gesamtbetrachtung erhält man durch Hochrechnung der FPR und TPR der Gruppen auf die Gesamtpopulation und Addition der Anteile über alle Gruppen eines Merkmals.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3 Datenbeschreibung

Dieses Kapitel dient der Beschreibung des Sozio-oekonomischen Panels (SOEP) sowie der Skizzierung der Auswahl von Variablen und der Aufbereitung des Datensatzes für die in Kapitel 4 folgenden Analysen. Für diese Arbeit liegt die Version „Soepv26“ des SOEP-Datensatzes mit den Daten für die Jahre von 1984-2009 vor.

Im ersten Teil des Kapitels wird zunächst das für die Generierung des Datensatzes und die Analysen nötige Grundwissen über das SOEP vorgestellt. Anschließend werden im zweiten Teil des Kapitels die Datenauswahl und die Generierung des Datensatzes beschrieben. Der letzte Abschnitt beschäftigt sich mit der Analyse von Einkommensmobilitäten der Haushalte im Datensatz, die erste Erkenntnisse zur Messbarkeit der Vulnerabilität mit den deutschen Daten liefert.

3.1 SOEP

Das SOEP ist eine Panel-Erhebung in Deutschland, die seit 1984 durch das Umfrageinstitut TNS Infratest Sozialforschung für das Deutsche Institut für Wirtschaftsforschung Berlin (DIW) durchgeführt wird. Das SOEP stellt für die sozial-, verhaltens- und wirtschaftswissenschaftliche Grundlagenforschung Mikrodaten bereit, mit dem Ziel soziale, gesellschaftliche und politische Entwicklungen in Deutschland *langfristig* zu untersuchen. Darüber hinaus werden die SOEP-Daten für die Sozialberichterstattung und die Politikberatung genutzt (Wagner et al., 2008).

Der Begriff Panel-Erhebung bezeichnet eine periodisch wiederholte Befragung derselben Haushalte über einen gewissen Zeitraum. Die Befragung findet jährlich, zum größten Teil im ersten Drittel des Jahres[5], zu den Themen Haushaltszusammensetzung, Erwerbs- und Familienbiographie, Erwerbsbeteiligung und berufliche Mobilität, Einkommensverläufe, Gesundheit und Lebenszufriedenheit statt (Wagner et al., 2008).

[5] Bis April sind etwa 80% der Befragungen abgeschlossen. Bis zum Jahresende wird versucht, die nicht-erreichten Haushalte/Personen nachzuerheben.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Erhebung erfolgt durch das Ziehen von Haushaltsstichproben, bei denen alle Personen in diesen Haushalten Teil der Erhebung werden. Interviewt werden sowohl der Haushaltsvorstand stellvertretend für den gesamten Haushalt (Haushaltsinterviews) als auch Personen (Personeninterviews). Kinder werden erst ab dem 17. Lebensjahr befragt. Verlassen Haushaltsmitglieder den Haushalt (z.B. durch Scheidung), so werden diese innerhalb Deutschlands weiterverfolgt. Personen, die mit diesen zusammenziehen, werden ebenfalls auf Dauer in das SOEP einbezogen. Das Konzept der Weiterverfolgung, sogenannter „Schneeballeffekt“, wird durch eine Gewichtung berücksichtigt (Wagner et al., 2008).

3.1.1 Stichprobe

Die SOEP-Stichprobe besteht aus (derzeit) neun Teilstichproben, die zu unterschiedlichen Zeitpunkten gestartet worden sind und mit Großbuchstaben A-I (in Reihenfolge der Startzeitpunkte) bezeichnet werden, siehe Tabelle 3.1 (in Anlehnung an Goebel et al. (2008)).

Stichprobe	Art	Start
Stichprobe A	Haushalte in der BRD[6]	1984
Stichprobe B	Ausländische Haushalte in der BRD[7]	
Stichprobe C	Privathaushalte in der DDR	1990
Stichprobe D	Zuwanderer-Privathaushalte in Deutschland	1994/95
Stichprobe E	Auffrischungsstichprobe	1998
Stichprobe F		2000
Stichprobe H		2006
Stichprobe I		2009
Stichprobe G	Hocheinkommensstichprobe	2002

Tabelle 3.1: *Stichproben des SOEPs [Wagner et al., 2008]*

Das Ziehungsdesign der Stichproben ist unterschiedlich. Grundsätzlich wird zur Ziehung der Stichproben ein zweistufiges Auswahlverfahren angewendet. Es werden

[6] Haushalte, deren Haushaltsvorstand zum Zeitpunkt der Ziehung *nicht* türkischer, italienischer, jugoslawischer, griechischer oder spanischer Nationalität war

[7] Haushalte, deren Haushaltsvorstand zum Zeitpunkt der Ziehung türkischer, italienischer, jugoslawischer, griechischer oder spanischer Nationalität war

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

zunächst Primäreinheiten (etwa Wahlkreise) und aus diesen im Anschluss die Sekundäreinheiten (Haushalte) gezogen. Die Primäreinheiten werden aus einer mehrfach stratifizierten Zufallsstichprobe gezogen, wobei die Schichtung regional erfolgt (etwa Bundesland, Regierungsbezirk, Gemeindetyp) (Rendtel, 1995, S. 31ff.). Eine Ausnahme bildet die Stichprobe B, bei der die Schichtung nach Nationalität des Haushaltsvorstands erfolgt (Haisken-DeNew und Frick, 2005, S. 155f.). Die Ziehung erfolgt zumeist über systematisches und größenproportionales Ziehen mit festem Intervall und zufälliger Startzahl (Särndal, Swenson und Wretmann, 2003, S. 87-99).

Zu Beginn der Panel-Erhebung bezog sich die Zielpopulation auf die in Westdeutschland lebenden (deutschen und ausländischen) Haushalte inklusive West-Berlin. Nach der Wende wurde diese ergänzt durch die in Ostdeutschland lebenden Haushalte. Im Laufe der Zeit wurden außerdem Auffrischungsstichproben gezogen, um der Panelsterblichkeit entgegenzuwirken. Gründe für das Ausscheiden von Haushalten/Personen aus dem Panel sind z.B. Tod, Erkrankung, Verzug. Zusätzlich wurde unabhängig von allen anderen Stichproben eine Hocheinkommensstichprobe gezogen (Rendtel (1995, S. 29-45), Haisken-DeNew, und Frick (2005, S. 19-21)).

Die Entwicklung der Fallzahlen privater Haushalte ist in der Abbildung 3.1 dargestellt.

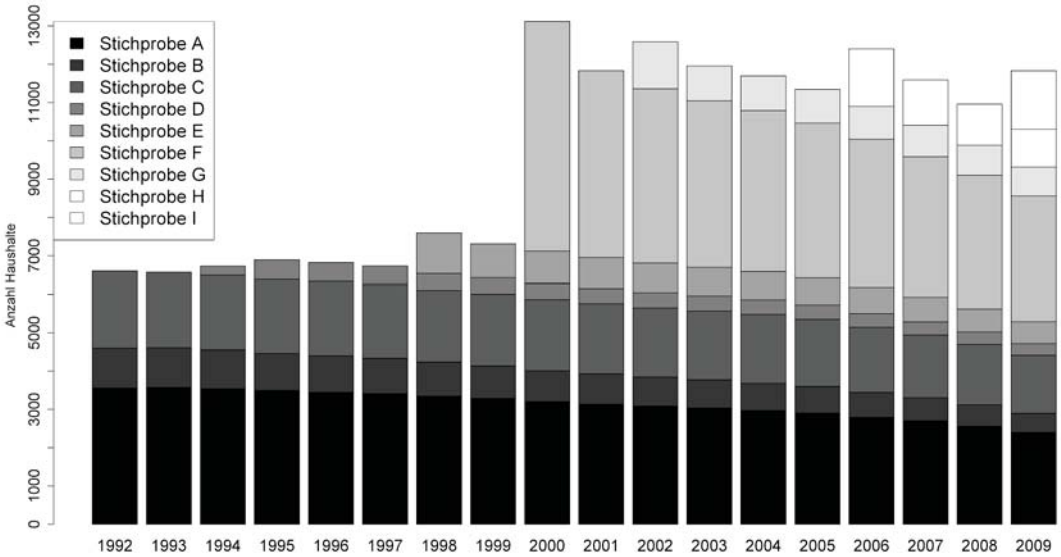


Abbildung 3.1: *Entwicklung der Fallzahlen privater Haushalte (1992-2009) [Darstellung auf Basis der in Kapitel 3 beschriebenen Daten in Anlehnung an Goebel et al. (2008)].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.1.2 Grundstruktur

Das SOEP ist ein sehr komplexer Datensatz. Das Ziel dieses Teilabschnittes ist es, einen Einblick in die Datenstruktur zu geben.[8] Für detailliertere Informationen sowie der Benennung von Dateien und Variablen sei auf Haisken-DeNew und Frick (2005) oder das webbasierte Informationssystem SOEPinfo von Ingo Sieber und John Haisken-DeNew verwiesen.

Die SOEP-Befragungen finden jährlich statt, so dass die Antworten zu Einkommen, Beruf, Familienstand usw. *stichtagsbezogen* sind. Pro Welle werden diese in separaten Dateien aufbereitet und liegen somit in einer *Querschnittsstruktur* vor, siehe dazu Abbildung 3.2.

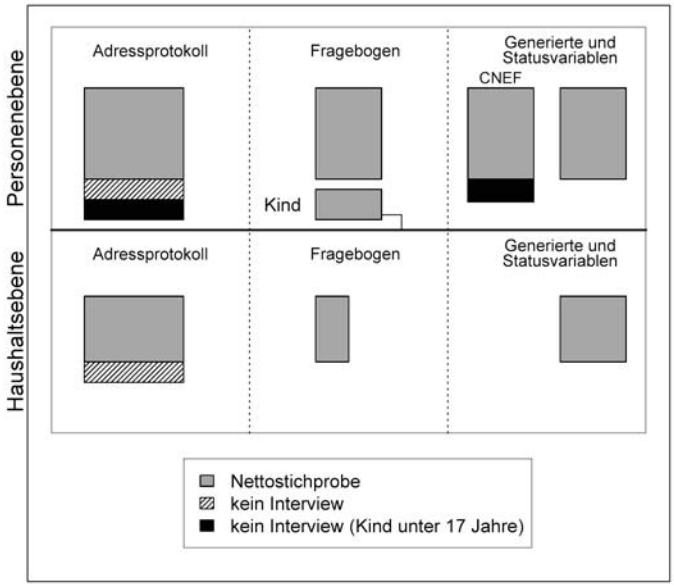


Abbildung 3.2: *Datenstruktur des SOEPs im Querschnitt [in Anlehnung an Goebel et al. (2008)]*

[8] Seit der Datenversion für 2009, „Soepv26“, werden die Teildatensätze nicht nur im „weiten Tabellenformat“, d.h. Variablen unterschiedlicher Zeitpunkte werden nebeneinander angeordnet, sondern auch „im langen Tabellenformat (SOEPlong09)“, d.h. die Querschnitte werden untereinander geschrieben, bereitgestellt. Da in dieser Arbeit die Generierung der Daten auf Basis des „weiten Tabellenformats“ erfolgt, bezieht sich dieser Abschnitt auf die Datenstruktur in dieser Darstellung.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Dateien werden getrennt nach Personen- und Haushaltsebene angelegt und enthalten die Nettostichprobe, d.h. die Personen (älter als 17 Jahre) bzw. Haushalte, die tatsächlich an der Welle teilgenommen haben. Durch sogenannte Schlüsselvariablen (eindeutige Personennummer auf der Personenebene sowie die Ursprungshaushaltsnummer und die aktuelle Haushaltsnummer auf der Haushaltsebene), die jede Datei enthält, ist es möglich Personen und Haushaltsdateien sowie verschiedene Querschnitte zu verbinden.

Neben Dateien mit direkten Antworten auf die Fragen des Fragebogens werden sowohl auf der Haushalts- als auch auf der Personenebene generierte und Statusvariablen bereitgestellt. Ein Teil dieser Variablen wird aus den direkten Antworten generiert, z.B. der höchste Bildungsstand einer Person. Einige Fragen werden nicht im jährlichen Turnus sondern nur im ersten Interview oder bei Veränderungen gestellt, z.B. die Wohnfläche eines Haushalts. In den Statusvariablen wird der aktuelle Stand dieser Informationen dem Nutzer *jährlich* verfügbar gemacht (Goebel et al., 2008).

Bei dem anderen Teil der generierten Variablen handelt es sich um *harmonisierte Variablen* aus dem Cross-National Equivalent File (CNEF). Das CNEF ist eine 1991 geschaffene Datenbasis von derzeit sieben nationalen Panelstudien[9], in die das SOEP integriert ist. Im CNEF werden ausgewählte Variablen harmonisiert, die mindestens in zwei (anerkannten) nationalen Panelstudien enthalten sind. Das Ziel besteht darin, Analysen verschiedener Länder vergleichbar zu machen und die Qualität der nationalen Panelstudien zu erhöhen. Die Daten des CNEFs sind auf der Personenebene angelegt und beinhalten Informationen zur Haushaltsstruktur, der Erwerbstätigkeit, dem Gesundheitszustand und demographischen Merkmalen der Individuen. Eine Besonderheit der Daten liegt darin, dass Brutto- und Nettojahreseinkommen der Haushalte sowie eine Reihe von Einkommenskomponenten zur Verfügung gestellt werden (Frick et al., 2008).

Weiterhin werden auf beiden Ebenen für jede Welle Bruttostichproben, sogenannte Adresslogs, verfügbar gemacht, die allgemeine Informationen über das Geschlecht, das Geburtsjahr, Informationen zum Interview (auf der Personenebene) sowie Informationen zum Wohnsitz und Anzahl der Mitglieder im Haushalt (auf der Haus-

[9] Die weiteren nationalen Panelstudien sind die US-amerikanische Panel Study of Income Dynamics (PSID), der British Household Panel Survey (BHPS) aus Großbritannien, der Survey of Labour and Income Dynamics (SLID) aus Kanada, der Household Income and Labour Dynamics Survey (HILDA) aus Australien, das Schweizerische Haushalt-Panel (SHP) sowie die Korean Labor & Income Panel Study (KLIPS)

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

haltsebene) über alle in die Stichprobe gezogenen Personen/Haushalte geben (Haisken-DeNew und Frick, 2005, S. 29).

Neben den Dateien im *Querschnittsformat* werden auch Dateien im *Längsschnittformat* angelegt (siehe z.B. Abbildung in Haisken-DeNew und Frick (2005, S. 32)). Der Hauptteil der Längsschnittdaten besteht aus sogenannten *Spelldaten* bzw. Kalendarien. Ein *Spell* bezeichnet einen Lebensabschnitt bzw. eine Zeitdauer, in der eine Untersuchungseinheit in einem bestimmten Zustand verweilt, z.B. Zeit in Schulbildung. Während Paneldaten eine Information für eine Person/einen Haushalt pro Jahr (bzw. je eines äquidistanten Zeitintervalls) erfassen, geben Spelldaten Auskunft über eine Information für eine Person/einen Haushalt pro Spell. Auf der Personenebene sind Einkommens- und Beschäftigungskalender angelegt, bei denen die Verweildauern in der Anzahl der Monate gemessen wird. Der Einkommenskalender stellt die Grundlage zur Generierung der jährlichen Einkommen im CNEF dar. Weiterhin gibt es Spelldaten für die Beschreibung der aktuellen Tätigkeit sowie den Familienstand. Auf der Haushaltsebene werden Spelldaten bereitgestellt, in denen Haushalte erfasst werden, die Sozialleistungen empfangen haben. Ein weiterer Block der Längsschnittdaten stellen die Biographiedaten zu unterschiedlichen Themen dar.

Einen übergeordneten Charakter in der Panelstudie haben die sogenannten Masterdateien auf der Personen- und der Haushaltsebene. Sie enthalten alle Einheiten jeder Welle, die vom SOEP kontaktiert wurden, d.h. die Bruttostichprobe der gesamten Panelstudie. In diesen Dateien werden u.a. Informationen zur Demographie, der Zugehörigkeit der Stichprobe und der Teilnahmestatus dokumentiert (Haisken-DeNew und Frick, 2005, S. 31).

Um das Ziehungsdesign der Stichproben sowie den Abgang und den Neuzugang von Einheiten zu berücksichtigen, werden für die Bruttostichprobe sowohl auf der Personen- als auch auf der Haushaltsebene Gewichtungsfaktoren bereitgestellt. Querschnitts- und Längsschnittsgewichte ermöglichen es, sowohl repräsentative Studien von Querschnitts- als auch Längsschnittdaten durchzuführen (Goebel et al., 2008).

3.1.3 Hochrechnung

Das Erhebungsdesign (unterschiedliche Ziehungswahrscheinlichkeiten der Teilstichproben) und die Weiterverfolgungsregeln des Panels (Fusionen/Abspaltung von Haushaltsteilen, neue Erhebungseinheiten, Ausfälle) machen eine Gewichtung der

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Erhebungseinheiten notwendig, um aus der Stichprobe verlässliche Rückschlüsse auf die Grundgesamtheit ziehen zu können. In diesem Teilabschnitt werden die Grundzüge der Gewichtung im SOEP beschrieben, die notwendig sind, die Gewichtung in Kapitel 4 nachzuvollziehen. Für detailliertere Informationen siehe Galler (1987), Rendtel (1995), Haisken-DeNew und Frick (2005, S. 153-186) und Goebel et al. (2008).

Die Generierung der Gewichte im SOEP basiert auf dem Konzept von Horvitz und Thompson (1952), bei dem jede Einheit mit seiner inversen Ziehungswahrscheinlichkeit gewichtet wird. Voraussetzung für die Anwendung dieser Gewichtung ist allerdings, dass alle gezogenen Einheiten auch tatsächlich beobachtet wurden. In Haushaltspanels trifft dies selten zu, so dass die Auswahlwahrscheinlichkeit neben der Ziehungswahrscheinlichkeit noch durch die Teilnahmebereitschaft der Einheiten beeinflusst wird. Weiterhin führen Ausfälle im Verlauf eines Panels dazu, dass Einheiten, die in früheren Wellen noch beobachtet wurden, nun die Teilnahme verweigern. „Insbesondere Galler (1987) und Rendtel (1995) haben das Gewichtungsschema von Horvitz und Thompson für das Design eines komplexen Haushaltspanels erweitert“ (zitiert durch Goebel et al. (2008)). Dieser Ansatz ermöglicht die Gewichtung von Erhebungseinheiten auch im Längsschnitt (und damit ebenfalls im Querschnitt einer beliebigen Welle).

Die Erhebungseinheiten im SOEP sind sowohl Haushalte als auch die Personen dieser Haushalte. In der Startwelle stimmen die Gewichte überein. Im Verlauf des Panels ändern sie sich jedoch, z.B. bei Teilnahme eines Haushaltes aber Verweigerung einer einzelnen Person dieses Haushaltes. Aus diesem Grund werden im SOEP Quer- und Längsschnittsgewichte für Haushalte und Personen generiert. Da in dieser Arbeit Analysen auf der Haushaltsebene durchgeführt werden, beschränken sich die folgenden Betrachtungen, soweit möglich, auf Haushaltsgewichte (Goebel et al., 2008).

Die Gewichte des SOEPs werden auf drei Ebenen entwickelt, die im folgenden grob beschrieben werden sollen. Generierung von

1. Querschnittsgewichten der Startwelle (einer Stichprobe)
2. Längsschnittsgewichten
3. Querschnittsgewichten von Folgewellen.

Dazu seien D, A_t, R_t und $C_{t,w}$ binäre Variablen mit den folgenden Bedeutungen (siehe Definitionen (3.1)-(3.4)):

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$$D = \begin{cases} 1 & \text{wenn h in Welle eins in die Stichprobe gezogen wurde,} \\ 0 & \text{sonst.} \end{cases} \tag{3.1}$$

$$A_t = \begin{cases} 1 & \text{wenn h in Welle } t \text{ kontaktiert wurde,} \\ 0 & \text{sonst.} \end{cases} \tag{3.2}$$

$$R_t = \begin{cases} 1 & \text{wenn h in Welle } t \text{ geantwortet hat,} \\ 0 & \text{sonst.} \end{cases} \tag{3.3}$$

$$C_{t,w} = \begin{cases} 1 & \text{wenn h über } w \text{ Wellen (ab } t\text{) erfolgreich ausgewählt wurde,} \\ 0 & \text{sonst.} \end{cases} \tag{3.4}$$

Die Wahrscheinlichkeit, dass ein Haushalt in der ersten Welle in die Stichprobe gezogen wurde und geantwortet hat $P(C_{1,1} = 1)$, ergibt sich aus dem Produkt der Ziehungswahrscheinlichkeit eines Haushalts $P(D = 1)$ und seiner Antwortwahrscheinlichkeit, gegeben, dass er in die Stichprobe gelangt ist $P(R_1 = 1 \mid D)$. Die Ziehungswahrscheinlichkeit ergibt sich aus dem Ziehungsdesign. Im Ziehungsdesign der Startwelle werden die Einheiten durch Haushalte beschrieben (Rendtel, 1995). $P(R_1 = 1 \mid D)$ ist unbekannt und muss geschätzt werden. Dies wird über die relativen Beobachtungshäufigkeiten in den Regionalzellen gemacht. Die Beobachtungswahrscheinlichkeiten werden auf das 10-fache des teilprobenspezifischen Medians begrenzt, und es wird eine Randanpassung an den Mikrozensus (zumeist) nach Region, Alter, Geschlecht, Haushaltsgröße, Familienstand, Job, Schulbildung und Nationalität (Haisken-DeNew und Frick, 2005, S.176) vorgenommen. Gründe dafür liegen in möglichen Mängeln beim Auswahlverfahren und der Schätzung der Antwortwahrscheinlichkeiten. Die geschätzte Auswahlwahrscheinlichkeit $\widehat{P}(C_{1,1} = 1)$ ergibt sich dann wie in Gleichung (3.5).

$$\widehat{P}(C_{1,1} = 1) = P(D = 1) \cdot \widehat{P}(R_1 = 1 \mid D) \tag{3.5}$$

Die inversen (geschätzten) Auswahlwahrscheinlichkeiten sind die Querschnittsgewichte für die Startwelle im SOEP, die jedem Haushalten zugewiesen wurden.

Die Konstruktion von Längsschnittsgewichten (ausgehend vom Startjahr) im SOEP ist ein mehrstufiger Prozess. Im ersten Schritt werden die Auswahlwahrscheinlichkeiten in der ersten Welle geschätzt. Die nächsten beiden Schritte beziehen sich auf die Schätzung der Kontakt- und Antwortwahrscheinlichkeiten in

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

der zweiten Welle. Die Kontaktwahrscheinlichkeit ist diejenige (bedingte) Wahrscheinlichkeit, dass Haushalte, die in der ersten Welle in die Stichprobe gelangt sind und geantwortet haben, auch in der zweiten Welle kontaktiert werden können $P(A_2 = 1 \mid D, R_1)$. Die Antwortwahrscheinlichkeit ist diejenige Wahrscheinlichkeit, dass diese erneut erfolgreich kontaktierten Haushalte auch geantwortet haben $P(R_2 = 1 \mid A_2, R_1, D)$. Für Details zu den Schätzverfahren siehe oben genannte Literaturhinweise. Die geschätzte Auswahlwahrscheinlichkeit $P(C_{1,2} = 1)$ für die ersten zwei Wellen ergibt sich als Produkt von vier Wahrscheinlichkeiten wie in Gleichung (3.6).

$$\widehat{P}(C_{1,2} = 1) = \overbrace{P(D = 1) \cdot \widehat{P}(R_1 = 1 \mid D)}^{Auswahlwahrscheinlichkeit\ Welle\ 1} \cdot \overbrace{\widehat{P}(A_1 = 1 \mid D, R_1) \cdot \widehat{P}(R_2 = 1 \mid D, R_1, A_2)}^{Bleibewahrscheinlichkeit\ Welle\ 2} \tag{3.6}$$

Der Kehrwert der Bleibewahrscheinlichkeit wird als Bleibefaktor bezeichnet und bildet das Gewicht für die Längsschnittgewichtung eines Panels mit zwei Wellen. Dies ergibt sich einfach als Produkt von Querschnittsgewicht und Bleibefaktor. Das Prozedere kann für beliebig viele Wellen erweitert werden. Das Längsschnittgewicht ergibt sich als Produkt des Querschnittsgewichts der Startwelle und der Bleibefaktoren der darauffolgenden Wellen (Haisken-DeNew und Frick, 2005, S. 40f. & S.179ff.). Da nicht jeder Panel mit der Startwelle beginnen soll, wird im nächsten Absatz beschrieben, wie das Gewicht eines Panels mit dem Beginn in einer beliebigen Welle berechnet wird.

Die Schätzung der Querschnittsgewichte in den folgenden Wellen ist komplexer, da sich Haushalte teilen und verbinden. Die Rohgewichte in der zweiten Welle wurden berechnet, indem die Querschnittsgewichte mit den Bleibefaktoren aus der zweiten Welle multipliziert wurden. Es erfolgte eine Randanpassung an den Mikrozensus (Goebel et al., 2008).

Im SOEP werden beispielsweise die Längsschnittsgewichte eines 4-Jahres-Panels ab Welle vier, d.h. ein Panel von Welle vier bis Welle sieben (entspricht den Jahren 1987-1990), berechnet, indem das Querschnittsgewicht in Welle vier multipliziert wird mit den Bleibefaktoren aus den Wellen fünf, sechs und sieben. Das Gewicht, das einem Haushalt zugewiesen wird, gibt die Anzahl der Haushalte an, die er repräsentiert.

Die Gewichtung eignet sich sowohl im designbasierten Kontext, z.B. bei der Berechnung von Anteilen, als auch im modellbasierten Kontext, z.B. in Regressionsmodellen (Robins et al. (1994) und (1995), Wooldridge (2002a), (2002b) und (2004) zitiert durch Goebels et al. (2008)).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.1.4 Fehlende Werte

Obwohl das SOEP eine relativ vollständige Panelstudie ist, besteht dennoch das Problem fehlender Daten aufgrund von Nichtbeantwortung einer Frage („item-non-response“, INR), Nichtteilnahme an der Welle („unit-non-response “, UNR) und unglaubwürdigen Antworten von Einzelpersonen oder gesamten Haushalten.

Das Problem des INR besteht lediglich für kategoriale Daten, da stetige Variablen, etwa das Haushaltseinkommen, imputiert werden. Die Imputation erfolgt im SOEP mit zwei Methoden.

Sind für den Haushalt mit INR bzw. die Person sowohl Querschnitts- als auch Längsschnittsinformationen bekannt, wird die Methode „row-and-column-imputation“ von Little und Su (1989) angewendet. Die Imputationsmethode ist eine multiplikative Verknüpfung von drei Elementen: dem Zeileneffekt, dem Spalteneffekt und einer stochastischen Komponente. Der Spalteneffekt ergibt sich aus dem durchschnittlichen Trend einer Periode, während der Zeileneffekt den Trend der Einheit mit INR angibt. Die stochastische Komponente ergibt sich durch Bestimmung und Ordnung der Zeileneffekte aller Individuen sowie Matching des Haushalts mit INR mit dem nächsten Nachbarn mit vollständiger Information an dieser Stelle.

Sind für den Haushalt mit INR keine Längsschnittdaten vorhanden, d.h., wenn Daten zum ersten Mal verfügbar sind, so wird eine Querschnitts-Imputationsmethode angewendet. Als Imputationsmethode kommt die „hot deck regression“ zum Einsatz. Hier werden die fehlenden Werte auf Basis der bestehenden Werte im Querschnitt vorhergesagt (Frick und Grabka (2004), Frick und Grabka (2007)).

3.2 Generierung des Datensatzes

In dieser Arbeit werden Haushalte als Analyseeinheiten verwendet, um die Ergebnisse mit Studien zur Vulnerabilität der Armut vergleichbar zu machen, die ihre Analysen ebenfalls auf Haushalten basieren. Es werden Daten für die Jahre 1992-2009[10] verwendet, so dass die Analysen Ergebnisse für Deutschland insgesamt liefern. Die Daten werden auf jährlicher Basis (ein Eintrag pro Jahr) betrachtet, d.h., es werden Daten aus SOEP-Querschnittsdateien zu Panels verbunden.

[10] Die Daten für Ostdeutschland stehen ab 1991 bereit; in dieser Arbeit wird erst eine Welle später begonnen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.2.1 Datenauswahl und -beschreibung

In dieser Arbeit werden in hohem Maße harmonisierte Variablen aus dem CNEF ausgewählt. Diese sind verlässlich und haben sich in Anwendungen bewährt. Darüber hinaus sind die Ergebnisse zu anderen Erhebungen (aus dem CNEF) vergleichbar. Weiterhin werden einige Statusvariablen verwendet (Anger (2011), SOEP (2009)). Tabelle 3.2 zeigt die ausgewählten Variablen. Die hervorgehobenen Variablen wurden zwar untersucht, aber letztendlich nicht zur Schätzung oder der Analyse der Vulnerabilität verwendet. Informationen zur Bezeichnung und Speicherungsort im SOEP werden im Anhang in Tabelle A.1 gegeben.

Einige dieser Variablen liegen bereits auf der Haushaltsebene vor (z.B. Haushaltseinkommen, Anzahl der Personen im Haushalt, Wohnfläche)[11], während andere noch auf die Haushaltsebene aggregiert werden müssen (später werden z.B. die Alterstruktur im Haushalt oder die Gesamtarbeitsstunden des Haushalts[12] betrachtet). Andere der Eigenschaften sind individuelle Eigenschaften, die nur für den Haushaltsvorstand (HV) betrachtet werden (z.B. Geschlecht, Bildung, Industrie). Es wird hier vereinfacht unterstellt, dass der Haushaltsvorstand repräsentativ für den Haushalt steht, d.h., dass die Eigenschaften des Haushaltsvorstands als charakteristisch für den Haushalt angesehen werden.

Das Netto-Haushaltseinkommen ist eine vom CNEF bereitgestellte, aggregierte Variable. Sie wurde aus der Summe der Arbeitseinkommen, den Kapitalgewinnen aus Wertpapiergeschäften, privaten (z.B. Schenkungen) und öffentlichen Transferleistungen (z.B. Arbeitslosengeld, Kindergeld, Elterngeld), Renten der Sozialversicherung (z.B. Alters-, Invaliden-, Hinterbliebenenrente) und privaten Renten aller Mitglieder des Haushalts in der Befragungswelle für das *Vorjahr* abzüglich der Steuern des Haushalts berechnet. Die Komponenten des Haushaltseinkommens wurden durch direkte Befragung der Haushalts ermittelt, mit Ausnahme des Kapitaleinkommens und der Steuern. Zur Ermittlung des Kapitaleinkommens wurden die Haushalte zunächst gebeten sich in eine von sechs Klassen (unter 250 €, 250 €-1 000 €, 1 000 €-2 500 €, 2 500 €-5 000 €, 5 000 €-10 000 €, über 10 000 €) einzuordnen. Daraufhin wurde ihnen ein Kapitaleinkommen durch Ziehen von gleichverteil-

[11] Einige der Variablen sind in Dateien auf der Personenebene eingeordnet, wie z.B. das Haushaltseinkommen oder die Anzahl der Personen im Haushalt

[12] Die Arbeitsstunden der Individuen ist die einzige, in den Analysen verwendete, retrospektive Haushaltsvariable. Diese Variable Haushaltsarbeitsstunden wird aus den individuellen Arbeitsstunden, die im Jahr t dokumentiert sind, aufgrund der Haushaltszusammensetzung in t aggregiert, auch wenn diese für $t-1$ erhoben sind.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ten Zufallszahlen in der genannten Klasse zugeordnet. Die Steuern des Haushalts wurden mit der Methode von Schwarze (1995) berechnet (Grabka, 2005a).

Themen	Variablen
Wohlfahrt	`Brutto-Haushaltseinkommen (des Vorjahres)`
	Netto-Haushaltseinkommen (des Vorjahres)
	Kalkulatorische Miete (des Vorjahres)
Demographie	Geschlecht
	Alter
	Familienstand
	Anzahl der Personen im Haushalt
	Anzahl der Kinder im Haushalt
	Haushaltsvorstand
	Bundesland
Bildung	Höchster Bildungsstand (bzgl. Schulbildung)
Arbeitsmarkt	Industrie
	Labour Force
	Arbeitsmarktstatus
	`Öffentlicher Dienst`
	Arbeitsstunden (des Vorjahres)
Wohnsitz	Status des Eigentums
	Wohnfläche
	`Miete`
Haushalt	Art
	Migrationshintergrund
Gesundheit	`Selbsteinschätzung Gesundheit`
	`Krankenhausaufenthalte (im Vorjahr)`
	`Arztbesuche (im Vorjahr)`
	`Arbeitsunfähigkeit`
Makrovariable	Verbraucherpreisindex (für das Vorjahreseinkommen)
Gewichte	Haushalts-Querschnittsgewichte
	Haushalts-Bleibefaktoren

Tabelle 3.2: *Auswahl von Variablen (Schätzung ohne hervorgehobene Variablen)*

Um das Netto-Haushaltseinkommen zwischen den Haushalten mit unterschiedlicher Anzahl an Haushaltsmitgliedern vergleichen zu können, ist es in der EU festgelegt, anstelle des Haushaltseinkommens das Äquivalenzeinkommen als Wohlfahrt-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sindikator zu verwenden. Dieses ist definiert als das Haushaltseinkommen inklusive der kalkulatorischen Mietkosten (Canberra Group, 2001, S.62), das aufgrund einer Äquivalenzskala gewichtet wird. Als Skala findet die modifizierte OECD-Skala Anwendung, die sich nach der Anzahl der Personen im Haushalt und ihrem Alter richtet. Bei der Gewichtung nach dieser Skala erhält der Haushaltsvorstand das Gewicht 1, alle übrigen Haushaltsmitglieder, die älter als 14 Jahre sind, das Gewicht 0.5 und alle Kinder bis 14 Jahre das Gewicht 0.3 (siehe z.B. Atkinson et al. (2002, S. 98ff), Krause und Ritz (2006), Stauder und Hüning (2004)). Um das Äquivalenzeinkommen über die Zeit vergleichbar zu machen, wird es um die Inflation bereinigt (Basisjahr 2005). In Zukunft wird das Äquivalenzeinkommen verkürzt mit *Einkommen* bezeichnet.

Bei der Berechnung des Einkommens muss berücksichtigt werden, dass das Haushaltseinkommen *retrospektiv* ist. Das bedeutet, dass das Haushaltseinkommen im Befragungsjahr sich auf das Haushaltseinkommen im vorhergehenden Jahr bezieht, allerdings aggregiert auf Basis der Haushaltsmitglieder im Befragungsjahr. Dementsprechend bezieht sich auch der im SOEP dokumentierte Verbraucherpreisindex in einem Jahr auf das Einkommen des vorhergehenden Jahres. Auf dieses Problem trifft man häufig, wenn Jahreseinkommen betrachtet werden (Debels und Vandecasteele (2008), Haisken-DeNew und Frick (2005), Grabka (2005)).

3.2.2 Datenorganisation

Die SOEP Daten liegen als Dateien in drei verschiedenen Formaten vor, dem Stata-, SPSS-, und SAS-Format. In dieser Arbeit wurden die Daten im Stata-Format verwendet. Der Datensatz wurde mit Hilfe der webbasierten Informationsplattform SOEPinfo und der Statistiksoftware Stata SE 9.2. generiert. SOEPinfo bietet eine komplette Variablenübersicht, sowohl im Querschnitt als auch im Längsschnitt. Die Variablensuche ist über den direkten Namen sowie über Stichwort- und Themensuche möglich. Weiter werden Fallzahlen zu jeder Variable bereitgestellt sowie direkten Zugriff von einer Variable einer bestimmten Welle auf die der übrigen Wellen. Bereitgestellt werden ebenso die vergangenen und aktuellen Fragebögen der Erhebung in deutscher und englischer Sprache. SOEPinfo bietet zudem eine Programmierunterstützung für die Statistikprogramme Stata, SPSS und SAS, siehe Abbildung 3.3. Im „Basket“ werden die Variablen abgelegt, die in den Datensatz aufgenommen werden sollen, zusätzlich werden automatisch Variablen aus den Masterfiles zugespielt. Hierbei handelt es sich um die Schlüsselvariablen (Personennummer und Haushaltsnummer), Variablen wie Geschlecht und Geburtsjahr

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

(nur auf der Personenebene), Zugehörigkeit der Teilstichprobe sowie Teilnahme in der jeweiligen Welle. Wie die Abbildung zeigt, kann der Nutzer Einstellungen zur Ebene der Analyseeinheiten, dem Design, den Haushalten, dem Geschlecht, den Stichproben sowie der Region vornehmen. Zusätzlich können die Pfade gesetzt werden, unter dem die SOEP-Daten liegen und unter dem die temporären Dateien abgelegt werden sollen. Durch den Button „Generate Stata Code“ wird der Code generiert, der in Stata den Datensatz im breiten Format aufbereitet. Für diese Arbeit wurde folgende Auswahl getroffen:

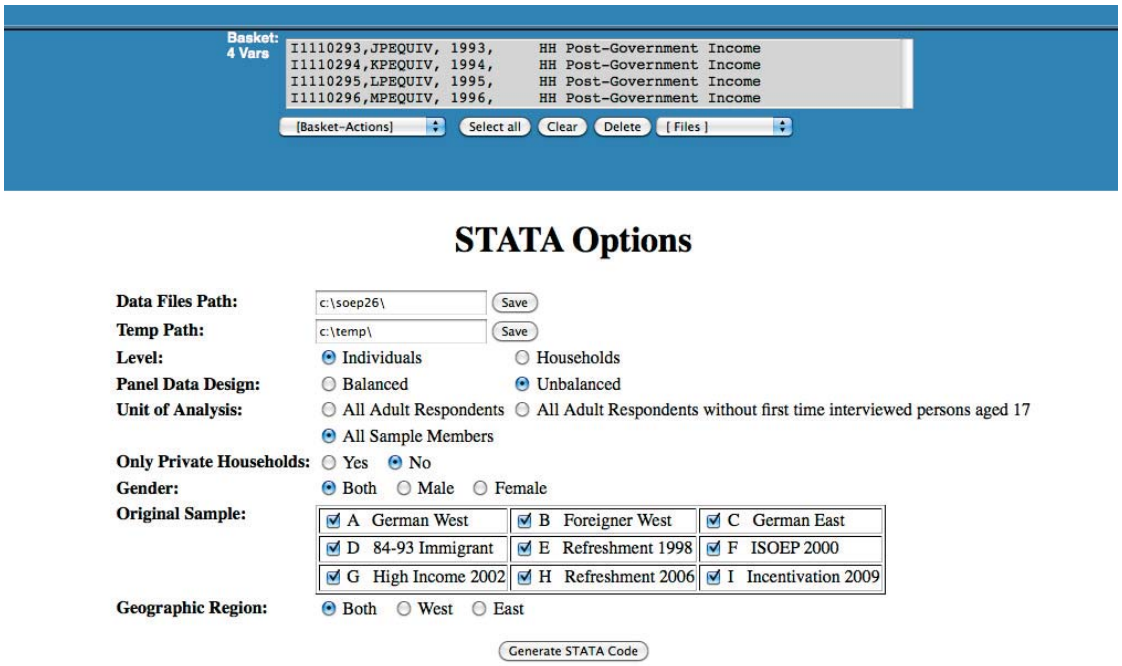


Abbildung 3.3: *Unterstützung durch SOEPinfo*

Der Datensatz wurde zusätzlich zur Überprüfung mit der Statistik-Software `R` generiert. Um die Daten im Stata-Format in `R` lesen zu können, muss die Bibliothek `foreign` eingebunden werden. Die weitere Organisation der Daten erfolgte ebenfalls mit `R`, wie die Beschränkung auf die Privathaushalte, die Berechnung einiger Variablen (Einkommen), die Aggregation einiger Personenvariablen auf die Haushaltsebene (z.B. Altersstruktur) sowie die Beschränkung des Datensatzes auf einen Eintrag pro Haushalt (des Haushaltsvorstands). Die Variablen wurden für jedes Befragungsjahr einzeln organisiert und dann, abhängig von der Analyse, wieder zu Panels verbunden. Die Komponenten Einkommen, Haushalts- und Individualeigen-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

schaften, zu aggregierende Eigenschaften und Haushaltsgewichte wurden ebenfalls getrennt aufbereitet und dann zu einem Jahresdatensatz verbunden.

Zur Berechnung des Einkommens werden folgende zwei Schritte unternommen: Zunächst wird das Haushaltseinkommen aus dem Jahr t den Daten von $t+1$ entnommen (ebenso die kalkulatorischen Mietkosten). Die Summe der beiden Komponenten wird auf Basis der modifizierten OECD-Skala gewichtet. Anders als in vielen Studien wird nicht die Haushaltszusammensetzung des Befragungsjahres verwendet sondern des vorhergehenden Jahres, in dem das Einkommen bezogen wurde. Das führt zu Verzerrungen, wenn Personen, die im letzten Jahr Einkommen bezogen haben, den Haushalt zum Befragungsjahr verlassen haben oder neu hinzugekommen sind (Debels und Vandecasteele, 2008). Ebenso wird der Verbraucherpreisindex für t dem Jahr $t+1$ entnommen, und die Einkommen werden auf das Preisniveau des Jahres 2005 umgerechnet. Da nur *ein Eintrag pro Privathaushalt* betrachtet wird (Variablen wie Einkommen oder Anzahl der Mitglieder im Haushalt sind für alle Haushaltsmitglieder identisch), werden die Daten über die Variablen Beziehung zum Haushaltsvorstand und Art des Haushalts extrahiert. Es werden nur die Informationen von den Haushaltsvorständen (deutscher oder ausländischer Herkunft) von Privathaushalten im Datensatz belassen.

Haushaltseigenschaften, die bereits auf der Haushaltsebene oder als Individualeigenschaften des Haushaltsvorstands vorliegen, werden nur bzgl. der Faktorstufen geändert (siehe dazu Tabelle 4.1). Die Daten werden wieder so extrahiert wie zuvor beschrieben. Einige Eigenschaften müssen dagegen noch aggregiert werden, wie die Altersstruktur (mit vier Altersklassen), Summe der Jahresarbeitsstunden sowie die Anzahl der Voll- und Teilzeitbeschäftigten eines Haushalts. Die Aggregation wird jeweils in getrennten Funktionen durchgeführt, und die Daten werden, wie zuvor beschrieben, extrahiert.

Eigentlich würde für die Bewertung von Modellen mit Paneldaten ein 3-Jahres-Panel ausreichen: zwei Jahre sind notwenig, um die Vulnerabilität zu schätzen und ein weiteres Jahr, um die ROC zu berechnen und damit die Genauigkeit der Schätzer zu bestimmen. Da die Einkommensvariable jedoch um ein Jahr verschoben ist, wird eine SOEP-Welle mehr benötigt als wenn das Einkommen bereits am Ende des relevanten Jahres vorliegt. Daher basieren die Analysen in dieser Arbeit auf symmetrischen 4-Jahres-Panels zwischen 1992 und 2009, d.h., Haushalte werden ausgeschlossen, wenn sie nicht in allen vier Jahren anwesend waren.

Es soll noch einmal darauf hingewiesen werden, dass die Vulnerabilität für das Befragungsjahr und nicht für das Folgejahr geschätzt wird. Das liegt daran, dass

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

das Einkommen des Befragungsjahres erst im folgenden Jahr verfügbar ist. Mit anderen Worten wird die Vulnerabilität für die Gegenwart und nicht für die Zukunft geschätzt. Dieses Problem tritt immer auf, wenn das Einkommen des vergangenen Jahres als Wohlfahrtsindikator verwendet wird. Ein Vorteil besteht jedoch darin, dass die meisten der *aktuellen* Haushaltsvariablen dem Forscher oder Politiker im relevanten Jahr verfügbar sind; sie müssen also nicht vorhergesagt werden.

Die Haushaltsgewichte werden berechnet, indem das Haushalts-Querschnittsgewicht des ersten Jahres mit den Bleibefaktoren auf der Haushaltsebene im zweiten, dritten und vierten Jahr multipliziert wird.

Nach dem Ausschluss von nicht-teilgenommenen Haushalten in einer Welle liegen in dieser Arbeit noch zwei Arten von fehlenden Werten vor: fehlende (z.B. Bildung oder Industrie des Haushaltsvorstands) und verzerrte Variablen (z.B. Altersstruktur des Haushalts bei fehlendem Wert von einem Geburtsjahr eines Haushaltsmitglieds oder Imputation vom Haushaltseinkommen). Im ersten Fall wurden die Haushalte des 4-Jahres-Panels ausgeschlossen, auch wenn Informationen nur in einem der vier Jahre gefehlt haben. Im zweiten Fall werden die (verzerrten) Werte genutzt, die vorhanden sind. Alternativ wurden im ersten Fall nur die Informationen in dem Jahr ausgeschlossen, in dem sie gefehlt haben. In diesem Fall erhält man aus den Modellen P_1-P_6 Vorhersagen unterschiedlicher Länge für dasselbe Jahr, so dass die ROCs nicht direkt miteinander vergleichbar sind. Beide Vorgehensweisen führen zu sehr ähnlichen Regressionsergebnissen, daher wird in dieser Arbeit wie in der ersten Vorgehensweise verfahren.[13]

3.3 Einkommensmobilität von Haushalten im SOEP

Für die Schätzung der Vulnerabilität mit Kovariablen sind natürlich die beiden Aspekte relevant, wie hoch die Wahrscheinlichkeit ist, dass Haushalte in Armut fallen oder ihr entkommen. Dieser Abschnitt befasst sich mit der Untersuchung der Einkommensmobilität im vorliegenden Datensatz.

In Situationen, in denen die Armen arm bleiben und die Nicht-Armen selten arm werden, sind die Begriffe Armut und Vulnerabilität gewissermaßen gleich-

[13] Wenn jedoch mehr Daten fehlen, können sich die Ergebnisse durchaus stark unterscheiden. Fehlende Haushaltsvariablen sind vor allem dann ein Problem, wenn die Vorhersage des aktuellen Einkommens nur auf diesen basiert. Die fehlenden Werte sind dagegen nicht so problematisch, wenn bei der Vorhersage das Einkommen des vergangenen Jahres verwendet wird. Später wird gezeigt, dass das vergangene Einkommen den größten Einfluss auf die Vorhersage hat.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

bedeutend. Das ist nicht der Fall, wenn ein beachtlicher Anteil von Haushalten häufig zwischen den Zuständen arm und nicht-arm wechselt. Um die Situation im vorliegenden Datensatz zu untersuchen, werden die Haushalte aufgrund ihres Äquivalenzeinkommens einer von fünf Einkommensgruppen: [0,9000), [9000,12000), [12000,18000), [18000,22000) und [22000,30000) zugeordnet. Die Gruppen werden mit E, D, C, B und A bezeichnet, wobei E die Gruppe ist, die in dieser Arbeit als arm bezeichnet wird.

Abbildung 3.4 zeigt den Anteil der *derzeitig* armen Haushalte zerlegt nach Zugehörigkeit zu den Einkommensklassen im *Vorjahr*. Z.B. waren 58% der Armen in 1993 ebenfalls arm in 1992, 29% der Armen in 1993 waren in Gruppe D, 9% waren in Gruppe C, 3% in B und 1% in A. Die Zerlegung ändert sich kaum über den beobachteten Zeitraum von 16 Jahren. Es zeigt sich, dass tatsächlich ein großer Teil der Haushalte in Armut verharrt, aber dass auch ein beträchtlicher Anteil im Vorjahr (etwa 40%) nicht arm war.

Für spätere Ausführungen ist es ebenfalls notwendig, den Anteil der *derzeit* nicht-armen Haushalte, zerlegt nach der Zugehörigkeit zu den Einkommensklassen im *Vorjahr*, zu betrachten. Diese Gruppe wird mit NP_t bezeichnet. Abbildung 3.5 zeigt, dass von den Nicht-Armen in 1993 etwa 5% in 1992 arm gewesen waren, 15% kamen aus Gruppe D, 40% aus Gruppe C, 22% aus B und 18% aus A.

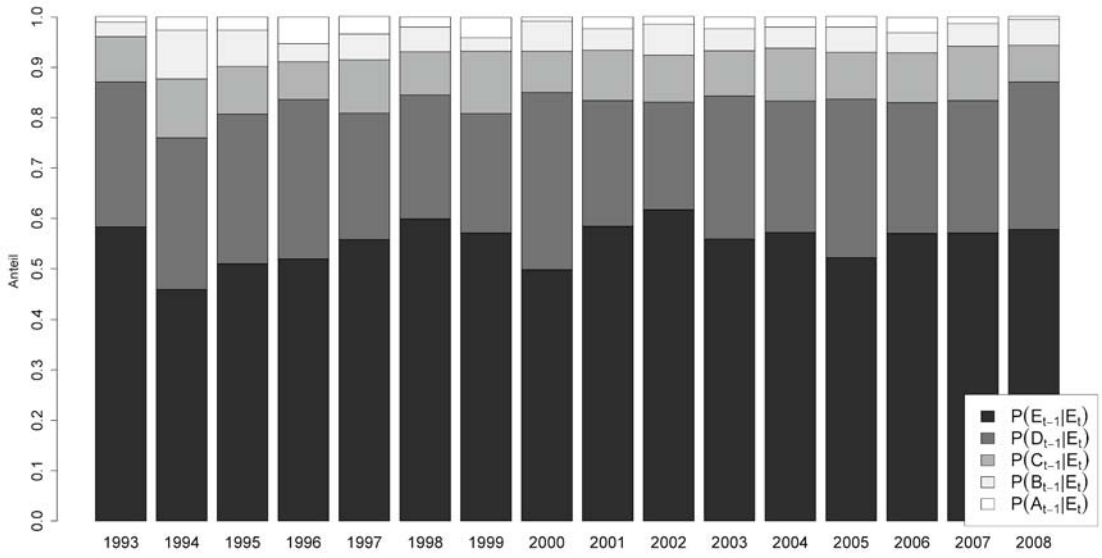


Abbildung 3.4: *Derzeitige Arme zerlegt nach Einkommensklassen des Vorjahres (1993-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

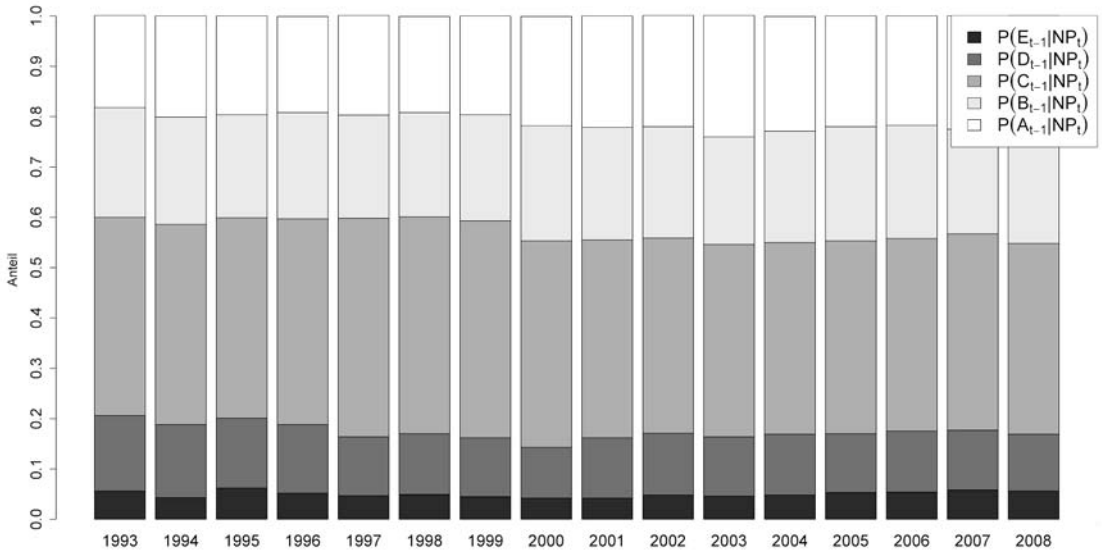


Abbildung 3.5: *Derzeit nicht-arme Haushalte zerlegt nach den Einkommensklassen des Vorjahres (1993-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

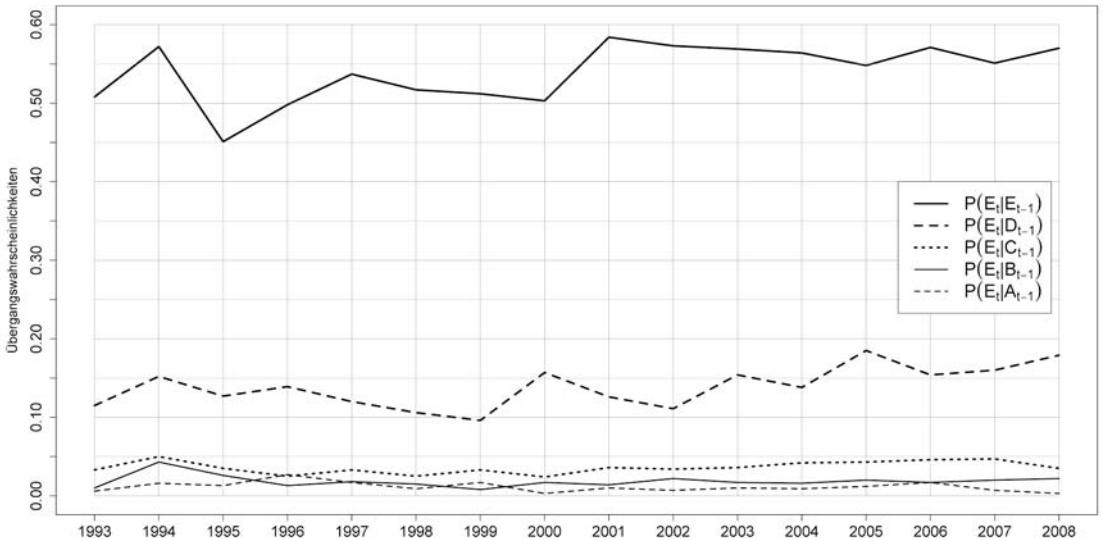


Abbildung 3.6: *Übergangswahrscheinlichkeit zur Armut im nächsten Jahr (1993-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Abbildung 3.6 zeigt die häufiger gebrauchte Darstellung der Übergangswahrscheinlichkeit von einer der fünf Einkommensklassen in $t-1$ zu der (armen) Einkommensklasse E in t.

Die Übergangswahrscheinlichkeiten und die in Abbildung 3.4 dargestellten Anteile unterscheiden sich hinsichtlich ihrer Interpretation. Während sich die Wahrscheinlichkeiten der Übergänge von Haushalten aus den fünf Einkommensklassen in die arme Klasse E_t nicht zu eins addieren müssen, ist dies für die Zusammensetzung der armen Haushalte in t aus den Anteilen der fünf Einkommensklassen des Vorjahres der Fall. Zusammenfassend ergibt sich, dass zwischen 45 und 60% der armen Haushalte im folgenden Jahr arm blieben, 10%-20% aus Gruppe D arm wurden, und weniger als 5% von jeder Gruppe C, B und A arm wurden.

Weiterhin wurde untersucht, ob sich die Übergangswahrscheinlichkeiten wie eine homogene Markov-Kette (Stirzaker, 2005, S. 107-147) verhalten. Wenn dies der Fall wäre, so könnten die Wahrscheinlichkeiten für jeden Haushalt, sich in einem der fünf Zustände zu befinden, für einen beliebigen Zeitpunkt vorausgesagt werden. Es ließe sich ein Zeitpunkt berechnen, in dem sich diese Wahrscheinlichkeiten nicht mehr verändern, d.h. Stabilität eintritt.

Für diese Analyse wurde für die Einkommensklassen aus zwei Jahren die Matrix der Übergangswahrscheinlichkeiten $\hat{T}_{t,t+1}$ berechnet. Die Übergangswahrscheinlichkeit von Jahr t zu Jahr $t+2$ wurde einerseits explizit aus den Daten geschätzt ($\hat{T}_{t,t+2}$), und andererseits (unter der Annahme, dass es sich um eine homogene Markov-Kette handelt) durch das Quadrat der Übergangsmatrix $(\hat{T}_{t,t+1})^2$ berechnet. Da $\hat{T}_{t,t+2} \neq (\hat{T}_{t,t+1})^2$, wird vermutet, dass sich die Übergangswahrscheinlichkeiten nicht wie eine homogene Markov-Ketten verhalten. Die Übergangsmatrizen sind exemplarisch für das Jahr $t = 1992$ in Anhang A.1.3 dargestellt. Insbesondere die Bleibewahrscheinlichkeiten von $(\hat{T}_{92,93})^2$ und $\hat{T}_{92,94}$ deuten darauf hin, dass keine Homogenität vorliegt, da diese sich relativ stark unterscheiden.

Im Allgemeinen kann gefolgert werden, dass, obwohl viele Haushalte in Armut verharren, auch ein beträchtlicher Anteil der Armen im Vorjahr nicht arm war. Zwischen 40% und 55% der armen Haushalte haben es geschafft, der Armut im folgenden Jahr zu entkommen. Angesichts der hohen Mobilität der Haushalte in Armut und aus der Armut hinaus (auch aus höheren Einkommensgruppen), kann nicht erwartet werden, dass die armutsgefährdeten Haushalte vollständig vorhergesagt werden können.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4 Empirische Studien zur Vulnerabilität

In diesem Kapitel wird die empirische Schätzung der Vulnerabilität basierend auf den Methoden in den Abschnitten 2.4.1-2.4.3 mit den Daten des SOEPs und der Statistik Software `R` Version 2.10.0 durchgeführt. Die Genauigkeit, mit der sich die Vulnerabilität schätzen lässt, hängt in hohem Maße von den verfügbaren Daten ab. Diese unterscheiden sich in der Art der Daten (Querschnitts- oder Paneldaten), der Responsevariable (stetig oder binär) sowie der vorhandenen Kovariablen (Vorjahreseinkommen (stetig, in Klassen), Haushaltskovariablen, Makrovariablen) und deren Qualität. Mit den Daten des „SOEPv26“ liegen für diese Arbeit lange Paneldaten hoher Qualität, d.h. relativ vollständig und verlässlich, vor. Als Wohlfahrtsindikator stehen stetige (harmonisierte) Einkommensvariablen zur Verfügung, für die in dieser Arbeit angenommen wird, dass sie dem tatsächlichem Einkommen entsprechen.

Das Ziel des Kapitels ist es, eine umfassende Aussage über die Genauigkeit der Schätzung der Vulnerabilität in oben beschriebenen Situationen der Datenverfügbarkeit zu treffen. Aufgrund der Datenlage können diese Situationen ideal abgebildet werden. In Abschnitt 4.1 werden relevante Kovariablen identifiziert und die Vulnerabilität in Datensätzen unterschiedlicher Länge (Querschnittsdaten, kurze Panels, längere Panels) geschätzt und retrospektiv bewertet.

Abschnitt 4.2 untersucht die Genauigkeit der Vulnerabilitäts-Schätzung, wenn die Kovariable Vorjahreseinkommen lediglich in wenigen Einkommensklassen zur Verfügung steht. Da die Ermittlung des jährlichen, stetigen Haushaltseinkommens mit einem hohen Aufwand verbunden ist und die Haushalte oft keine exakte Angabe über die einzelnen Komponenten ihres Einkommens machen können, wird in vielen Panels das Einkommen in Einkommensklassen (z.B. Mikrozensus) gemessen. In Entwicklungsländern sind Informationen über das Einkommen häufig gar nicht oder nur sehr ungenau verfügbar. Die Datenlage des SOEPs ermöglicht eine Untersuchung bei verschiedener Anzahl und unterschiedlichen Abgrenzungen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

von Einkommensklassen. Weiterhin wird untersucht, welche Auswirkung Fehler bei der Klassifizierung der Haushalte auf die Einkommensklassen für die Genauigkeit der Schätzung haben. Abschnitt 4.2 beantwortet die Frage, ob es von Vorteil ist, solche ungenauen Einkommensinformationen für die Schätzung der Vulnerabilität zu verwenden und ob sich der Aufwand lohnt, das Einkommen trotz der Unzulänglichkeiten zu ermitteln.

Es wird angenommen, dass sich die Ergebnisse, die auf deutschen, *idealen* Daten basieren, auch auf andere Länder mit höheren Armutsquoten als in Deutschland übertragen lassen. Bei hoher Armutsquote müssen *absolut betrachtet* mehr armutsgefährdete Haushalte und weniger nicht-armutsgefährdete Haushalte identifiziert werden als bei niedriger Armutsquote. In den Abschnitten 4.1 und 4.2 wird die Genauigkeit der Vulnerabilitäts-Schätzer für verschiedene Armutsquoten (niedrig, mittel, hoch) interpretiert.

Abschnitt 4.3 untersucht dagegen die Messung der Genauigkeit der Vulnerabilität spezifisch für Deutschland. Es werden Möglichkeiten untersucht, die Genauigkeit bisher erzielter Schätzer zu verbessern. Dazu werden zum einen Makrovariablen herangezogen und zum anderen die Vulnerabilität in Merkmalsgruppen separat geschätzt.

4.1 Identifizierung relevanter Kovariablen

In diesem Abschnitt wird mit den in Abschnitt 2.4.1 beschriebenen Modellen empirisch untersucht, wie genau sich die Vulnerabilität mit Kovariablen in Querschnitts- und Paneldatensätzen schätzen lässt. Dazu muss zunächst bestimmt werden, welche Kovariablen für die Genauigkeit der Vulnerabilitäts-Schätzung relevant sind. Vorausgesetzt wird, dass die in Kapitel 3 vorgestellten Haushaltseigenschaften sowie idealerweise stetiges, unverzerrtes Einkommen über den Zeitraum von 1992 bis 2009 verfügbar sind. Regressionstabellen werden für das Jahr 1996 im nächsten Abschnitt vorgestellt. Das Jahr 1996 ist rein zufällig gewählt. Die Ergebnisse sind auch für die anderen Jahre typisch und variieren nur leicht. Neben den Modellen aus Abschnitt 2.4.1 wurden weitere Modelle (logarithmische Transformation des Einkommens und heteroskedastische Modelle mit anderen Verteilungsannahmen) untersucht, deren Ergebnisse im Anhang abgebildet sind.

Für die Schätzung der Vulnerabilität werden rollierende 4-Jahres-Panels im Zeitraum von 1992 bis 2009 verwendet. Für die Modelle mit Paneldaten (P_4-P_6) werden die Informationen aus den ersten beiden Wellen benötigt, um das Einkommen in

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

der dritten Welle zu prognostizieren. Für die Modelle, die auf Querschnittsdaten basieren (P_1-P_3), wird lediglich die Information aus der zweiten Welle genutzt, um das Einkommen in der dritten Welle vorherzusagen. Für die Modelle P_2, P_4 und P_5 wurde angenommen, dass auch die Haushaltseigenschaften (aber nicht das Einkommen) für die dritte Welle verfügbar sind und auch zur Vorhersage des Einkommens in dieser Welle hinzugezogen werden können. Die beobachteten Einkommen im dritten Jahr wurden dann dazu verwendet, zu ermitteln, welche Haushalte arm geworden sind und welche nicht. Der Grund, dass Paneldatensätze mit vier Wellen zur Schätzung benötigt werden, ist, dass das Einkommen in der dritten Welle erst in der folgenden, d.h. vierten Welle, bekannt ist.

Um die ROC zu konstruieren, wurden Paare (FPR, TPR) für jede mögliche VPL auf Basis des vorhergesagten Einkommens für die dritte Welle ermittelt. Wenn die VPL höher gewählt wird, so steigen sowohl die TPR als auch die FPR.

In der Absicht, ein konkretes Maß für die Genauigkeit der Schätzung anzugeben, wird die Betrachtung der ROC auf einen Punkt reduziert, nämlich (FPR, $TPR = 80\%$) ($TPR = 90\%$ wird ebenfalls berücksichtigt). D.h., dass die VPL so gewählt wird, dass 80% der Haushalte, die am Ende des zweiten Jahres als armutsgefährdet geschätzt werden, im dritten Jahr tatsächlich arm geworden sind. Obwohl die Modelle P_1-P_6 alle dasselbe, nämlich das Einkommen in der dritten Welle, schätzen, können sich die VPLs von Modell zu Modell unterscheiden. Folglich verändert sich auch die zugehörige FPR und wie in diesem Abschnitt gezeigt wird, können die Unterschiede beträchtlich sein. Das beste Modell erreicht den vorgegebenen Anteil der richtigen Vorhersage der zukünftigen Armen und führt zum kleinsten Anteil der falsch als armutsgefährdet klassifizierten Haushalte.

Die Berechnungen zur Bestimmung der ROC werden für jedes der sechs Modelle P_1-P_6 für das Jahr 1996 sowie für die Modelle P_1 und P_6 für jedes Jahr im Zeitraum 1994-2008 durchgeführt. In den Abschnitten 4.1.1-4.1.3 wurde neben der Festlegung der Kovariablen und der Diskussion der verwendeten Modelle bestimmt,

- welche VPL zu einer TPR von 80% bzw. 90% führen,
- wie die Beziehung zwischen der VPL und dem Anteil der armutsgefährdeten Haushalte ist,
- und wie die FPR bei $TPR = 80\%$ und 90% ist.

Diese Analysen beziehen sich auf die Vulnerabilität im nächsten Jahr. In Abschnitt 4.1.4 wird auch noch die Vulnerabilität für n Jahre in der Zukunft betrachtet, wobei $n = 2, 3, 4$ gewählt wird. Für diese Analysen wird ein längerer Panel benötigt.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Analysen beruhen auf Teildatensätzen des 4-Jahres-Panels. Es werden nur die Stichproben der Haushalte betrachtet, die in allen vier Jahren an der Befragung teilgenommen haben, bzw. wenn ein vorhandener Haushalt sich aufgespalten hat, so wird der „Ursprungshaushalt“ betrachtet. Außerdem werden nur Haushalte betrachtet, die in allen vier Jahren ein Äquivalenzeinkommen unter 30 000 € haben. Zwar wird in der Literatur davon ausgegangen, dass die gesamte Population armutsgefährdet ist („Even millionaires are vulnerable to poverty“, z.B. Pritchett et al. (2000)), allerdings sind im vorhandenen Datensatz der Anteil der Haushalte, deren Einkommen bei mindestens 30 000 € liegt *und* die arm geworden sind, klein (0 bis 10 Haushalte von 6 000 bis 10 000 Haushalten pro Welle). Der Grund für den Ausschluss der wohlhabenden Haushalte ist, das Augenmerk in den Regressionen auf die Haushalte mit niedrigem Einkommen zu fokussieren und für die Schätzung der Vulnerabilität irrelevante Einträge, die die Regression beeinflussen, zu beseitigen. Weiterhin werden Haushalte ausgeschlossen, wenn Informationen über die verwendeten Variablen nicht verfügbar sind (zwischen 4% und 16% im verbleibenden 4-Jahres-Panel).

In Abschnitt 4.1.5 wird eine Zwischenbilanz zur Messung der Genauigkeit der Vulnerabilität mit den Methoden aus dem Abschnitt 2.4.1 gezogen.

4.1.1 Schätzung der Vulnerabilität für 1996

In diesem Abschnitt wird die Vulnerabilität mit den Modellen aus dem Abschnitt 2.4.1 exemplarisch für t =1996, das, wie bereits erwähnt, rein zufällig gewählt wird, geschätzt. Um die Modelle aus Abschnitt 2.4.1 anzuwenden, muss, mit Ausnahme des Modells P_3, zunächst bestimmt werden, welche Haushaltseigenschaften nützlich sind, um das Haushaltseinkommen vorherzusagen. Die ausgewählten Eigenschaften sind in Tabelle 4.1 dargestellt. Die Variablen arm1,...,arm5 und n(arm) werden in den folgenden Regressionen nicht verwendet. Später werden diese für die Untersuchung benötigt, ob die Armutshistorie über mehrere Jahre eine relevante Kovariable zur Schätzung der Vulnerabilität darstellt (siehe Anhang A.2.1 oder im Text auf Seite 76). Weitere zusätzliche Kovariablen wurden geprüft, z.B. Beschäftigung im öffentlichen/privaten Sektor, aber werden nicht in die Regression aufgenommen, da sie entweder zu Multikollinearität führen oder ihre Aufnahme nicht zu signifikanten Ergebnissen führt.

Im Folgenden werden die Regressionstabellen der Modelle $\mathrm{P}_{1,96}$, $\mathrm{P}_{4,96}$ und $\mathrm{P}_{6,96}$ dargestellt. Diese korrespondieren mit den Modellen P_1, P_4 und P_6 aus Abschnitt 2.4.1 auf Seite 35 für das Jahr 1996. Die Regressionstabellen der übrigen Modelle

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

enthalten kaum zusätzliche Informationen und werden daher nicht abgebildet. Die Schätzung der Regressionskoeffizienten von P_2 stimmt mit denen von Modell P_1 überein und die von P_5 ähneln den Schätzungen von P_4 bzw. P_6. Modell P_3 basiert nicht auf Haushaltskovariablen.

$$\begin{aligned} \text{Schätze } \beta \text{ mit} \quad & y_{h,95} = X_{h,95}\beta + e_{h,95} \\ \text{Vorhersage mit} \quad & \hat{y}_{h,96} = X_{h,95}\hat{\beta} \end{aligned} \tag{P$_{1,96}$}$$

$$\begin{aligned} \text{Schätze } \beta \text{ und } \gamma \text{ mit} \quad & y_{h,95} = \gamma y_{h,94} + X_{h,95}\beta + e_{h,95} \\ \text{Vorhersage mit} \quad & \hat{y}_{h,96} = \hat{\gamma} y_{h,95} + X_{h,96}\hat{\beta} \end{aligned} \tag{P$_{4,96}$}$$

$$\begin{aligned} \text{Schätze } \beta \text{ und } \gamma \text{ mit} \quad & y_{h,95} = \gamma y_{h,94} + X_{h,94}\beta + e_{h,95} \\ \text{Vorhersage mit} \quad & \hat{y}_{h,96} = \hat{\gamma} y_{h,95} + X_{h,95}\hat{\beta} \end{aligned} \tag{P$_{6,96}$}$$

Die Kovariablen in den Modellen $\mathrm{P}_{1,96}$, $\mathrm{P}_{4,96}$ und $\mathrm{P}_{6,96}$ korrelieren z.T. stark (z.B. Anzahl der Vollzeitbeschäftigten und Anzahl der Personen in den Altersgruppen n$[34, 59)$), d.h., es liegt Multikollinearität vor. Bei Multikollinearität ist es schwierig den Einfluss der Kovariablen auf die abhängige Variable isoliert anzugeben, auch wenn die Korrelation zwischen den beiden Kovariablen bekannt ist.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Variablen	Beschreibung
Einkommen	Äquivalenzeinkommen des Vorjahres
Geschlecht	Geschlecht des HVs (**männlich**, weiblich)
$n[0,18)$	Anzahl der Haushaltsmitglieder im Alter unter 18
$n[18,34)$	Anzahl der Haushaltsmitglieder im Alter zwischen 18 und unter 34
$n[34,59)$	Anzahl der Haushaltsmitglieder im Alter zwischen 34 und unter 59
$n[59-)$	Anzahl der Haushaltsmitglieder im Alter über 59
Familienstand	Familienstand des HVs (5 Stufen: **verheiratet**; alleinstehend; verwitwet; geschieden; getrennt)
Gesamtarbeitsstunden	Anzahl der Arbeitsstunden aller Haushaltsmitglieder im Vorjahr
Gewerbe	Arbeitssektor des HVs (10 Stufen: **nicht erwerbstätig**; Landwirtschaft; Energie; Bergbau; Produktion; Baugewerbe; Handel; Transport; Bank/ Versicherung; Dienstleistungen)
Wohnverhältnis	Status des Eigentums der Wohnstätte (3 Stufen: **Eigentümer**; Hauptmieter; Untermieter)
Größe des Wohnsitzes (Größe des Wohnsitzes)^2	Größe des Wohnsitzes in Quadratmetern quadratischer Term
Schulabschluss	Höchster Schulbildungsabschluss des HVs (5 Stufen: **Hauptschulabschluss**; Realschulabschluss; Abitur oder Fachhochschulreife (FHSR); andere Abschlüsse; kein Schulabschluss)
n(Vollzeitbeschäftigung) n(Vollzeitbeschäftigung)^2	Anzahl der Haushaltsmitglieder in Vollzeitbeschäftigung quadratischer Term
n(Teilzeitbeschäftigung)	Anzahl der Haushaltsmitglieder in Teilzeitbeschäftigung
arm1,..., arm5	Binäre Variablen zur Beschreibung der Armut in den vergangenen 5 Jahren ($t-2,\ldots,t-6$) (jeweils 2 Stufen: **nicht-arm**; arm)
n(arm)	Anzahl der Perioden in Armut in fünf vergangenen Jahren ($t-2,\ldots,t-6$)

Tabelle 4.1: *Beschreibung der Variablen in der Regression. Die Referenzstufen sind fett hinterlegt.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Variable	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	18860.31	494.09	38.17	<0.001
Geschlecht: weiblich	-645.31	185.86	-3.47	<0.001
$n[0,18)$	-1603.88	103.94	-15.43	<0.001
$n[18,34)$	-1477.45	174.63	-8.46	<0.001
$n[34,59)$	-1139.13	187.16	-6.09	<0.001
$n[59-)$	-216.37	213.65	-1.01	0.311
Familienstand: alleinstehend	-89.85	289.37	-0.31	0.756
Familienstand: verwitwet	298.52	284.71	1.05	0.294
Familienstand: geschieden	-1137.05	286.36	-3.97	<0.001
Familienstand: getrennt	-982.72	620.49	-1.58	0.113
Gesamtarbeitsstunden	0.19	0.10	1.92	0.055
Gewerbe: Landwirtschaft	-1036.88	751.73	-1.38	0.168
Gewerbe: Energiesektor	3324.15	843.63	3.94	<0.001
Gewerbe: Bergbau	814.35	1431.17	0.57	0.569
Gewerbe: Produktion	1200.30	331.92	3.62	<0.001
Gewerbe: Baugewerbe	1032.68	362.76	2.85	0.004
Gewerbe: Handel	85.65	361.66	0.24	0.813
Gewerbe: Transport	1524.63	474.78	3.21	0.001
Gewerbe: Bank/Versicherung	3855.56	641.36	6.01	<0.001
Gewerbe: Dienstleistungen	1044.26	310.57	3.36	<0.001
Wohnverhältnis: Hauptmieter	-933.66	183.48	-5.09	<0.001
Wohnverhältnis: Untermieter	-1433.11	416.23	-3.44	<0.001
Größe des Wohnsitzes	68305.67	6201.06	11.02	<0.001
(Größe des Wohnsitzes)^2	-22455.11	4773.50	-4.70	<0.001
Schulabschluss: Realschulabschluss	1014.66	185.81	5.46	<0.001
Schulabschluss: Abitur/FHSR	2017.96	226.69	8.90	<0.001
Schulabschluss: andere Abschlüsse	-1535.13	474.83	-3.23	0.001
Schulabschluss: kein Abschluss	-1841.39	560.70	-3.28	0.001
n(Vollzeitbeschäftigung)	145517.55	12793.75	11.37	<0.001
n(Vollzeitbeschäftigung)^2	-37514.93	6852.48	-5.47	<0.001
n(Teilzeitbeschäftigung)	1145.58	260.20	4.40	<0.001
adjustiertes Betimmtheitsmaß	0.294			
Anz. Beobachtungen (ungewichtet)[14]	4493			
F-Statitsik	61.15	(df: 30,4301)		

Tabelle 4.2: *Ergebnisse der Regression für Modell $P_{1,96}$ [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

[14] 161 Beobachtungen haben ein Gewicht von 0.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Variable	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	9037.08	449.21	20.12	<0.001
Vorjahreseinkommen	0.59	0.01	47.78	<0.001
Geschlecht: weiblich	-98.63	150.67	-0.65	0.513
$n[0,18)$	-608.48	86.57	-7.03	<0.001
$n[18,34)$	-713.33	142.06	-5.02	<0.001
$n[34,59)$	-608.74	151.70	-4.01	<0.001
$n[59-)$	-441.87	172.76	-2.56	0.011
Familienstand: alleinstehend	263.21	234.03	1.12	0.261
Familienstand: verwitwet	287.59	230.14	1.25	0.212
Familienstand: geschieden	-438.70	231.94	-1.89	0.059
Familienstand: getrennt	-776.46	501.58	-1.55	0.122
Gesamtarbeitsstunden	-0.37	0.08	-4.54	<0.001
Gewerbe: Landwirtschaft	-28.93	608.02	-0.05	0.962
Gewerbe: Energiesektor	1995.84	682.50	2.92	0.003
Gewerbe: Bergbau	-89.94	1157.02	-0.08	0.938
Gewerbe: Produktion	454.61	268.75	1.69	0.091
Gewerbe: Baugewerbe	732.07	293.30	2.50	0.013
Gewerbe: Handel	264.75	292.36	0.91	0.365
Gewerbe: Transport	789.27	384.09	2.05	0.040
Gewerbe: Bank/Versicherung	1764.89	520.27	3.39	<0.001
Gewerbe: Dienstleistungen	354.80	251.46	1.41	0.158
Wohnverhältnis: Hauptmieter	-640.40	148.44	-4.31	<0.001
Wohnverhältnis: Untermieter	-1072.11	336.53	-3.19	0.001
Größe des Wohnsitzes	31493.88	5071.40	6.21	<0.001
(Größe des Wohnsitzes)^2	-8476.78	3869.65	-2.19	0.029
Schulabschluss: Realschulabschluss	334.75	150.87	2.22	0.027
Schulabschluss: Abitur/FHSR	676.58	185.38	3.65	<0.001
Schulabschluss: andere Abschlüsse	-606.28	384.31	-1.58	0.115
Schulabschluss: kein Abschluss	-761.70	453.80	-1.68	0.093
n(Vollzeitbeschäftigung)	102425.57	10380.87	9.87	<0.001
n(Vollzeitbeschäftigung)^2	-10615.55	5567.63	-1.91	0.057
n(Teilzeitbeschäftigung)	945.36	210.37	4.49	<0.001
adjustiertes Betimmtheitsmaß	0.539			
Anz. Beobachtungen (ungewichtet)	4493			
F-Statitsik	164.2	(df: 31,4300)		

Tabelle 4.3: *Ergebnisse der Regression für Modell $P_{4,96}$ (mit aktuellen Kovariablen) [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

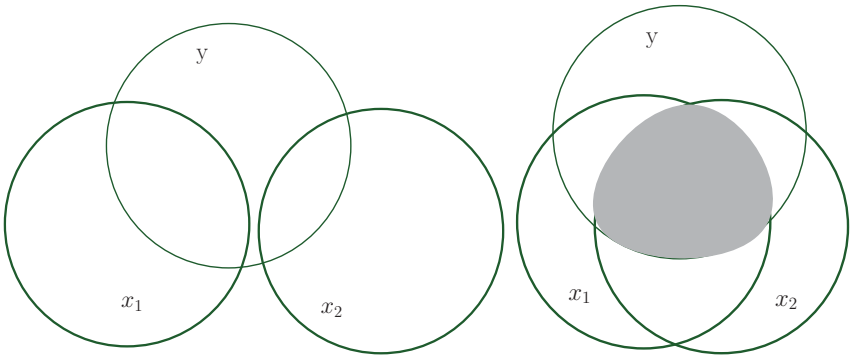
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Variable	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	6167.68	466.17	13.23	<0.001
Vorjahreseinkommen	0.61	0.01	47.65	<0.001
Geschlecht: weiblich	-23.98	154.07	-0.16	0.876
$n[0, 18)$	-311.02	88.74	-3.50	<0.001
$n[18, 34)$	-125.59	140.67	-0.89	0.372
$n[34, 59)$	1.32	150.25	0.01	0.993
$n[59-)$	218.94	172.50	1.27	0.204
Familienstand: alleinstehend	873.94	235.33	3.71	<0.001
Familienstand: verwitwet	325.65	235.60	1.38	0.167
Familienstand: geschieden	-172.52	239.76	-0.72	0.472
Familienstand: getrennt	768.38	454.72	1.69	0.091
Gesamtarbeitsstunden	0.15	0.08	1.86	0.063
Gewerbe: Landwirtschaft	-541.30	635.59	-0.85	0.394
Gewerbe: Energiesektor	1693.47	638.70	2.65	0.008
Gewerbe: Bergbau	95.04	986.31	0.10	0.923
Gewerbe: Produktion	544.80	274.63	1.98	0.047
Gewerbe: Baugewerbe	441.78	298.92	1.48	0.140
Gewerbe: Handel	252.16	299.53	0.84	0.400
Gewerbe: Transport	633.47	405.16	1.56	0.118
Gewerbe: Bank/Versicherung	1601.72	518.89	3.09	0.002
Gewerbe: Dienstleistungen	278.05	261.43	1.06	0.288
Wohnverhältnis: Hauptmieter	-620.23	152.85	-4.06	<0.001
Wohnverhältnis: Untermieter	-1437.91	339.71	-4.23	<0.001
Größe des Wohnsitzes	23364.76	5184.32	4.51	<0.001
(Größe des Wohnsitzes)^2	-5758.70	4013.95	-1.43	0.151
Schulabschluss: Realschulabschluss	256.92	153.55	1.67	0.094
Schulabschluss: Abitur/FHSR	701.04	190.56	3.68	<0.001
Schulabschluss: andere Abschlüsse Abschlüss	-698.80	389.88	-1.79	0.073
Schulabschluss: kein Abschluss	-749.81	461.76	-1.62	0.104
n(Vollzeitbeschäftigung)	17258.20	10605.28	1.63	0.104
n(Vollzeitbeschäftigung)^2	-11713.60	6134.41	-1.91	0.056
n(Teilzeitbeschäftigung)	158.85	218.85	0.73	0.468
adjustiertes Betimmtheitsmaß	0.520			
Anz. Beobachtungen (ungewichtet)	4493			
F-Statitsik	152.6	(df: 31,4300)		

Tabelle 4.4: *Ergebnisse der Regression für Modell $P_{6,96}$ (mit Kovariablen des Vorjahres) [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Das Phänomen der Multikollinearität wird im Venn-Diagramm in Abbildung 4.1 anhand einer bivariaten Regression veranschaulicht. In der Abbildung stellt der Kreis um y die Streuung der abhängigen Variablen und die Kreise um x_1 und x_2 die Streuungen der Kovariablen dar. Der sich überschneidende Bereich von y und x_1 stellt die Streuung dar, die durch x_1 erklärt wird (dasselbe gilt für x_2). Überschneiden sich aber auch die beiden Kreise von x_1 und x_2, so kann der überlappende Bereich (graue Fläche) der drei Kreise nicht vollständig einer der beiden Variablen zugeordnet werden. Diese Fläche symbolisiert die Multikollinearität. Auf der linken Seite (in 4.1(a)) liegt keine Multikollinearität vor (Kreise um x_1 und x_2 überlappen sich nicht), während auf der rechten Seite (in 4.1(b)) hohe Multikollinearität vorliegt (Stocker, 2012).



(a) Keine Multikollinearität (b) Hohe Multikollinearität

Abbildung 4.1: *Veranschaulichung von Multikollineariät im Venn-Diagramm [Darstellung nach Stocker (2012)].*

Multikollinearität ist hauptsächlich ein Problem, wenn das Ziel der Regression die Erklärung der abhängigen Variablen durch die Kovariablen ist. Die Interpretation der geschätzten Regressionskoeffizienten muss in solchen Fällen mit Vorsicht betrachtet werden. Multikollinearität ist weniger ein Problem, wenn das Ziel der Regression eine Vorhersage ist, solange die Korrelationen in etwa erhalten bleiben. In diesem Fall ist es weniger wichtig, wie groß der Einfluss bzw. die Korrelation zwischen den einzelnen Kovariablen und dem Response ist (Stocker, 2012). Im Folgenden wird auf die Ergebnisse des Modells $P_{1,96}$ eingegangen. Die Faktoren Geschlecht, Familienstand, Gewerbe, Wohnverhältnis und Bildung des HVs haben eine Wirkung. Die geschätzten Regressionskoeffizienten haben die erwarteten Vorzeichen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Folgende Haushaltseigenschaften haben eine negative Korrelation zum Einkommen:

- Haushalte mit weiblichem HV (in Bezug zur Referenzgruppe Haushalte mit männlichem HV),
- geschiedene Haushalte (in Bezug zur Referenzgruppe verheiratete Haushalte),
- Haushalte mit HV ohne Schulabschluss oder mit „anderen Abschlüssen“ (in Bezug zu Bildungsabschluss Hauptschulabschluss)
- Haushalte mit Schulabschluss Hauptschulabschluss des HVs (in Bezug zu Schulabschluss Realschulabschluss oder Abitur/FHSR)

Folgende Haushaltseigenschaften haben eine positive Korrelation mit dem Einkommen:

- Haushalte mit HVs in den meisten Gewerben (Produktion, Bank/Versicherung, Dienstleistungen (in Bezug auf die Referenzkategorie Haushalte mit nicht-erwerbststätigem HV))

Die drei Altersklassen (Anzahl der Haushaltsmitglieder in einer Altersklasse) $n[0,18)$, $n[18,34)$ und $n[34,59)$ sind hoch signifikant und haben eine negative Korrelation zum Einkommen, d.h., dass das Einkommen fällt, je mehr Mitglieder in diesen Klassen liegen. Das könnte u.a. auf die höhere Äquivalenzgewichtung zurückzuführen sein, ist vermutlich aber auch auf Multikollinearität zurückzuführen. Es scheint, als ob das Einkommen mit der Anzahl der Vollzeitbeschäftigten bis zu zwei Haushaltsmitglieder positiv korreliert ist. Ebenfalls eine positive Korrelation liegt zwischen dem Einkommen und der Anzahl der Teilzeitbeschäftigten im Haushalt vor (Faraway, 2002).

Die Vorzeichen für die Modelle $P_{4,96}$ und $P_{6,96}$ sind ähnlich wie für $P_{1,96}$. In diesen beiden Modellen wird zusätzlich die Variable des Vorjahreseinkommens verwendet, die hoch signifikant ist und eine wichtige Rolle bei der Schätzung einnimmt (Koeffizient von etwa 0.6).

Modell $P_{4,96}$ und $P_{6,96}$ haben wesentlich höhere Bestimmtheitsmaße als Modell $P_{1,96}$, nicht nur für das Jahr 1995 sondern für alle untersuchten Jahre. Das zeigt, dass das Einkommen in einem beliebigen Jahr eine wichtige Kovariable ist, um das Einkommen in dem darauffolgenden Jahr vorherzusagen. D.h., Vorhersagen, die nur auf Haushaltskovariablen von Querschnittsdaten beruhen, sind deutlich weniger genau als solche, die auf Paneldaten basieren. Die folgende ROC-Analyse wird diesen Unterschied graphisch veranschaulichen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Modell $P_{4,96}$ und $P_{6,96}$ erzeugen ähnliche Ergebnisse, siehe Tabellen 4.3 und 4.4. Eine Möglichkeit dafür ist, dass die meisten Eigenschaften sich von Jahr zu Jahr wenig verändern, wenn sich der Haushaltsvorstand nicht ändert. Es scheint, dass für die hier untersuchten Haushalte mit niedrigerem Einkommen, die vorhandenen Kovariablen wenig Erklärungsgehalt haben, so dass das Vorjahreseinkommen berücksichtigt wurde.

Wie in Abschnitt 2.4.1 angedeutet, wurden einige andere Verteilungen an die Regressionsmodelle angepasst. Da das Einkommen häufig als ungefähr logarithmisch normalverteilt (siehe z.B. Fahrmeir et al. (2007, S. 301)) angenommen wird, verwenden viele Autoren eine Logarithmustransformation des Einkommens. Dieses Vorgehen führte für die Daten dieser Studie nicht zu normalverteilten Residuen (siehe Anhang A.2.2); ihre Verteilung ist schief und die Kurtosis entspricht nicht 3. Dies kann dadurch erklärt werden, dass nur Haushalte betrachtet wurden, dessen Einkommen kleiner als 30 000 € ist. Wie die Histogramme in Anhang A.2.2 zeigen, führen auch die Regressionen mit Verteilungsannahme einer Normalverteilung nicht zu perfekt normalverteilten Residuen. Ein Vergleich der Regressionstabellen zeigt, dass die Regressionen ohne Transformation über ein geringfügig höheres Bestimmtheitsmaß verfügen wie die mit Transformation. Ein Vergleich zwischen den ROC-Kurven für die Modelle P_1, P_4 und P_6 zeigt, dass die Kurven ohne Transformation marginal besser sind als mit Transformation.

In dieser Arbeit wurden im Gegensatz zu anderen Studien (z.B. Chaudhuri et al. (2002)) homoskedastische Regressionsmodelle verwendet. Es wurde mit verschiedenen heteroskedastischen Regressionsmodellen experimentiert, in denen die Varianz als eine Funktion von Haushaltseigenschaften dargestellt wird. Eine Variante ist für die Modelle P_1 und P_6 in Anhang A.2.3 dargestellt. Die Anpassung erfolgte mit der Funktion `gamlss` aus dem gleichnamigen R-Paket von Stasinopoulos und Rigby (2007). Die Verbesserungen in der Anpassung waren unbedeutend. Weiterhin wurden heteroskedastische Regressionsmodelle mit anderen Verteilungsannahmen, wie die t-Verteilung, die Gammaverteilung und die Weibull-Verteilung angepasst. Wie Anhang A.2.4 zeigt, sind die Verteilungen der Quantilsresiduen schief; die Bestimmtheitsmaße verbessern sich ebenfalls nicht. Im Folgenden wird mit den Ergebnissen der homoskedastischen Modellen weitergearbeitet und eine Normalverteilung unterstellt.

Um die Güte der Schätzer für 1996 zu untersuchen, werden in Abbildung 4.2 die ROCs für die Modelle P_1-P_6 betrachtet. Die ROCs lassen sich in zwei Gruppen einteilen: vier Kurven, die das Einkommen zur Vorhersage verwenden und zwei Kurven, die diese Information nicht nutzen. Letztere sind deutlich weniger genau.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

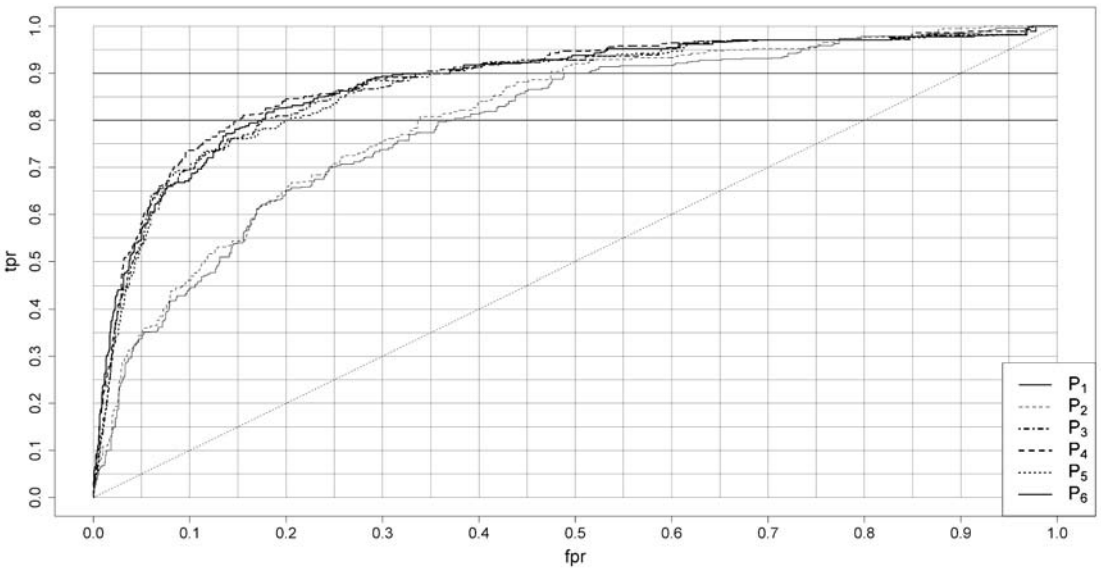


Abbildung 4.2: *ROC-Kurven für sechs verschiedene Vulnerabilitäts-Schätzungen für 1996 [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Ebenso wie der Vergleich der gesamten ROC-Kurven ist der Vergleich von FPRs bei einem gegebenen Niveau der TPR, z.B. bei 80% und 90% (Fienberg und Stern, 2005), von besonderem Interesse. Die zugehörigen FPRs dieser Punkte auf der ROC-Kurve bestimmen den Anteil der „false alarms“, d.h. den Anteil der Haushalte, die inkorrekt als armutsgefährdet geschätzt werden müssen, um 80% bzw. 90% der tatsächlich armen Haushalte richtig zu identifizieren.

Tabelle 4.5 zeigt die VPL und die FPR für jedes der Modelle für das Jahr 1996. Um 80% der Haushalte mit P_1 richtig zu identifizieren, die 1996 arm waren, müssen alle Haushalte, deren vorhergesagtes Einkommen unter 15 269 € lag, als armutsgefährdet deklariert werden. Die dazugehörige FPR ist 37%. Angesichts der Tatsache, dass die Armutsquote in 1996 bei etwa 8% lag, bedeutet das, dass 40% der Haushalte $(0.08 \cdot 0.8 + 0.37 \cdot 0.92)$ als armutsgefährdet klassifiziert werden müssen, um die vorgegebene TPR zu erreichen. Die FPR des besten Modells in 1996, P_4, liegt ungefähr bei 15%. Somit müssen 20% $(0.08 \cdot 0.8 + 0.15 \cdot 0.92)$ aller Haushalte als armutsgefährdet deklariert werden, ein prozentualer Anteil, der zwar nur halb so groß aber weiterhin ziemlich hoch ist. Wenn eine höhere TPR, etwa 90% erreicht werden soll, dann müssen sogar noch mehr Haushalte als armutsgefährdet deklariert werden.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	TPR=0.8		TPR=0.9	
Vorhersage	**FPR**	**VPL**	**FPR**	**VPL**
P_1	0.37	15269	0.50	16322
P_2	0.34	14964	0.47	16116
P_3	0.18	11894	0.34	14169
P_4	0.15	12751	0.37	15091
P_5	0.20	12300	0.37	14631
P_6	0.18	13018	0.37	15003

Tabelle 4.5: *FPR und VPL für TPR = 0.8 und TPR = 0.9 in 1996 [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Die Beziehung zwischen TPR und FPR misst die Genauigkeit einer Diagnostik Methode, unabhängig von der „Prävalenz“, d.h. in dieser Arbeit der Armutsquote in der Population. Hingegen hängt die Beziehung zwischen der *Anzahl* der false positives und der TPR von der Armutsquote ab. Der Zusammenhang wird in Tabelle 4.6 dargestellt, die die Anzahlen der richtig und falsch als armutsgefährdet klassifizierten Haushalte in einer hypothetischen Population von 1 000 Haushalten und drei verschiedenen Armutsquoten (10%, 20% und 40%) wiedergibt. Die Werte sind für die ROC jedes der Modelle P_1-P_6 für 1996 mit TPR =80% und 90% dargestellt

Die Zahlen in Tabelle 4.6 zeigen, dass für ein festes Paar (FPR, TPR), die *Anzahl* der false positives fällt, wenn die Armutsquote steigt. Im Falle einer niedrigen Armutsquote (10%) ist die Anzahl der false positives hoch. Um 80 der 100 armen Haushalte mit dem Modell P_1 zu identifizieren, muss die VPL so hoch fixiert werden, dass 333 von den 900 nicht-armen Haushalten falsch als armutsgefährdet deklariert werden müssen. Im Falle von 90 richtigen Klassifikationen steigt die Anzahl der false positives sogar auf 450. Wenn also nur Querschnittsdaten vorhanden sind, werden 333/(333+80) $\approx$ 81% (450/(450+90) $\approx$ 83%) der als armutsgefährdet klassifizierten Haushalten nicht arm. Im besten Fall, d.h. für Modell P_4 bei 80% und für Modell P_3 bei 90%, müssen 135 (306) von den im folgenden Jahr nicht-armen Haushalten als armutsgefährdet klassifiziert werden, um 80 (90) von 100 armen Haushalten richtig als armutsgefährdet zu klassifizieren, d.h. 63% (77%) der als armutsgefährdet deklarierten Haushalte werden nicht arm. Dagegen ist festzustellen, dass der Fehler, dass ein Haushalt nicht als armutsgefährdet deklariert aber im folgenden Jahr arm wird, relativ klein ist. Für Schätzer des Modells P_4 ergibt

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sich, dass 3% $(20/(20+765))$ bzw. 2% $(10/(10+594))$ der zukünftigen Armen als nicht-armutsgefährdet klassifiziert werden.

Armutsquote	10%				20%				40%			
	arm		nicht-arm		arm		nicht-arm		arm		nicht-arm	
Wahre Werte	100		900		200		800		400		600	
Modell	*TP*	*FN*	*FP*	*TN*	*TP*	*FN*	*FP*	*TN*	*TP*	*FN*	*FP*	*TN*
P_1	80	20	333	567	160	40	296	504	320	80	222	378
P_2	80	20	306	594	160	40	272	528	320	80	204	396
P_3	80	20	162	738	160	40	144	656	320	80	108	492
P_4	80	20	135	765	160	40	120	680	320	80	90	510
P_5	80	20	180	720	160	40	160	640	320	80	120	480
P_6	80	20	162	738	160	40	144	656	320	80	108	492
P_1	90	10	450	450	180	20	400	400	360	40	300	300
P_2	90	10	423	477	180	20	376	424	360	40	282	318
P_3	90	10	306	594	180	20	272	528	360	40	204	396
P_4	90	10	333	567	180	20	296	504	360	40	222	378
P_5	90	10	333	567	180	20	296	504	360	40	222	378
P_6	90	10	333	567	180	20	296	504	360	40	222	378

Tabelle 4.6: *True und false positives in einer hypothetischen Population von 1000 Haushalten für die Modelle P_1-P_6 und für drei verschiedene Armutsquoten [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Im Falle einer hohen Armutsquote (z.B. 40%) sind die Anzahl der false positives deutlich geringer. Z.B. mit P_4 (P_3 für 90%) werden 90 (204) Haushalte falsch als armutsgefährdet geschätzt, d.h. 22% (36%) von allen Haushalten, die als armutsgefährdet klassifiziert wurden. Allerdings steigt dann der Anteil der false negatives, d.h. der Anteil der Haushalte, die fälschlicherweise nicht als armutsgefährdet identifiziert wurden, auf 14% (10%). So ist in vielen Entwicklungsländern, in denen die Armutsquote zwischen 40% und 60% liegt, das Problem der falsch als armutsgefährdet klassifizierten Haushalte kleiner, während das Problem der nicht identifizierten armen Haushalte größer ist.

Die relative Performance der Modelle kann auch anhand der VPLs, die in Euro vorhergesagt werden, interpretiert werden. Diese können sich von Modell zu Modell und über die Zeit unterscheiden. Die VPLs sind für die Modelle P_1 und P_2 für eine gegebene TPR höher als für die Modelle P_3-P_6. Die ersten müssen mehr Haushalte als armutsgefährdet klassifizieren, um denselben Anteil der armen Haushalte richtig zu identifizieren.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Weiterhin wird untersucht, ob die Genauigkeit von Schätzern der Vulnerabilität erhöht werden kann, wenn die „Armutshistorie“ von Haushalten in den letzten fünf Jahren als Kovariable aufgenommen wird. Das wird auf zwei verschiedene Weisen durchgeführt. In der ersten Variante werden fünf binäre Dummy-Variablen (nicht-arm/arm) in das Regressionsmodell aufgenommen, d.h. eine für jedes der vorhergehenden Jahre. In der Alternative werden die Anzahl der Jahre gezählt, in denen ein Haushalt in den letzten fünf Jahren arm war. Beide Varianten führen zu signifikanten Regressionskoeffizienten aber die resultierende Verbesserung im Bestimmtheitsmaß und den ROC-Kurven war gering (Ergebnisse in Anhang A.2.1). Das zeigt, dass es wenig Nutzen bringt, die Armutshistorie der Haushalte zusätzlich zu der Information, die in den Haushaltskovariablen und im Vorjahreseinkommen steckt, in das Modell aufzunehmen. Das ist positiv zu bewerten, da Panels mit mehreren Wellen dafür benötigt würden und diese in Entwicklungsländern selten vorhanden sind.

4.1.2 Stabilität der Schätzer im Zeitraum 1994-2008

Im vorherigen Abschnitt wurden die Schätzer der Vulnerabilität für das Jahr $t = 1996$ geschätzt. In diesem Abschnitt wird untersucht, wie stabil die Vulnerabilität über die Zeit geschätzt werden kann. In dieser Arbeit wird die Betrachtung der Stabilität auf die Modelle P_1 und P_6 beschränkt. Diese beiden Modelle basieren auf dem Typ von Daten, die am häufigsten verfügbar sind und für Daten von Entwicklungsländern verwendet werden können.

Die Abbildungen 4.3 und 4.4 zeigen die ROC-Kurven für alle Jahre im Beobachtungszeitraum für die Modelle P_1 bzw. P_6. Es ist zu beobachten, dass die ROCs für die beiden genannten Modelle von Jahr zu Jahr variieren, wobei die ROC-Kurven basierend auf Schätzungen von P_1 stärker variieren als solche basierend auf P_6.

Bisher wurde die VPL ex post festgelegt und zwar so, dass sich eine TPR von 80% ergab. Um die Vulnerabilität zu schätzen, ist es natürlich notwenig, die VPL ex ante zu fixieren und zwar möglichst so, dass eine TPR von 80% bzw. 90% erreicht wird. Es wird an dieser Stelle eine konstante VPL fixiert. Eine Diskussion dazu erfolgt im Abschnitt 4.3. Die Niveaus der VPLs werden anhand der historischen VPLs, die zu einer TPR von 80% bzw. 90% führen, fixiert, was in den Abbildungen 4.5 und 4.6 dargestellt ist.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

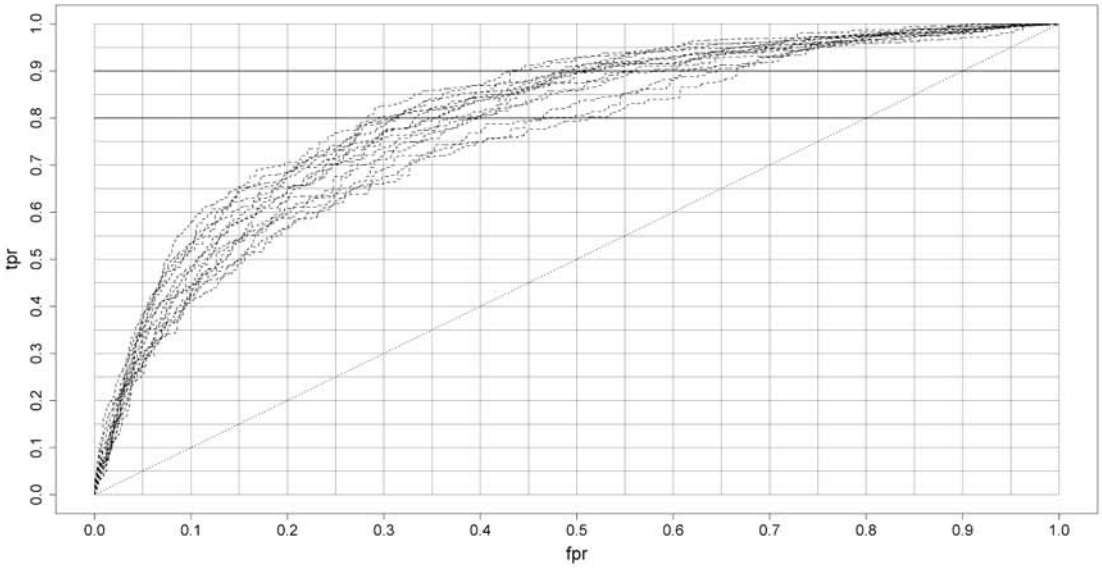


Abbildung 4.3: *ROC-Kurven erzeugt durch Modell P_1 (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

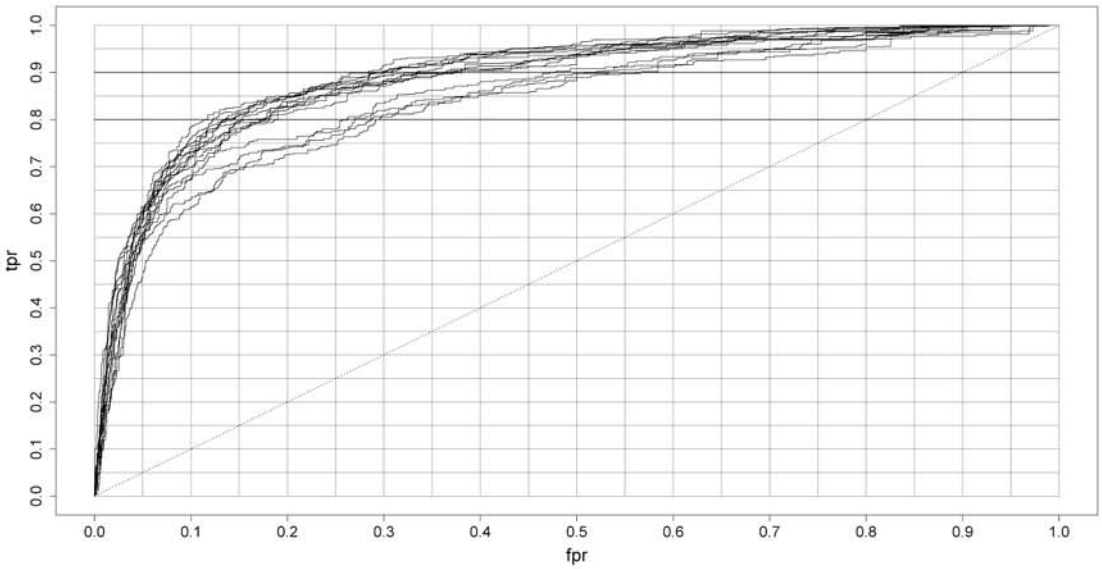


Abbildung 4.4: *ROC-Kurven erzeugt durch Modell P_6 (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

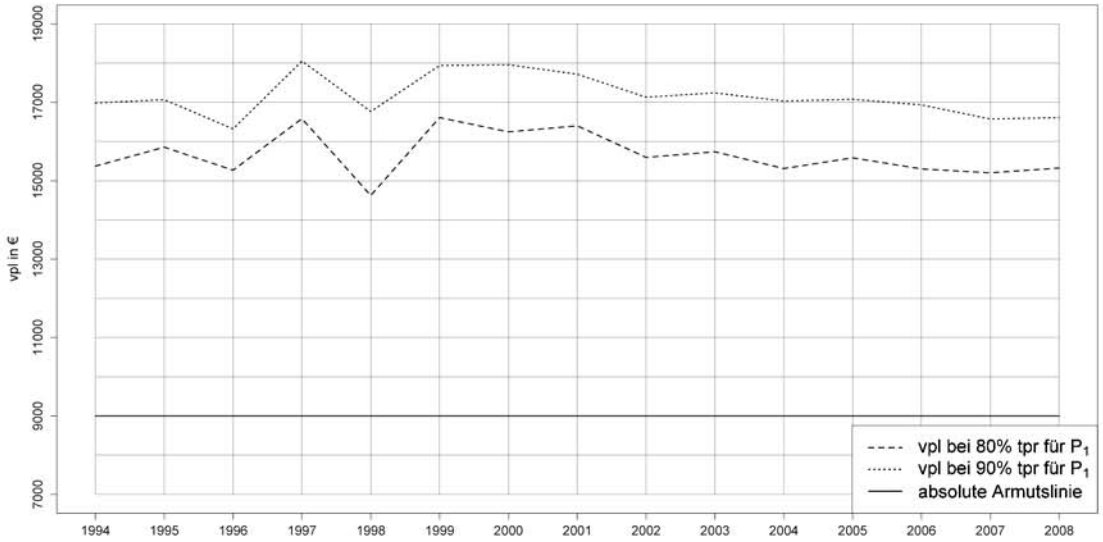


Abbildung 4.5: *Historische VPLs bei $TPR = 80\%$ und 90% erzeugt durch Modell P_1 (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

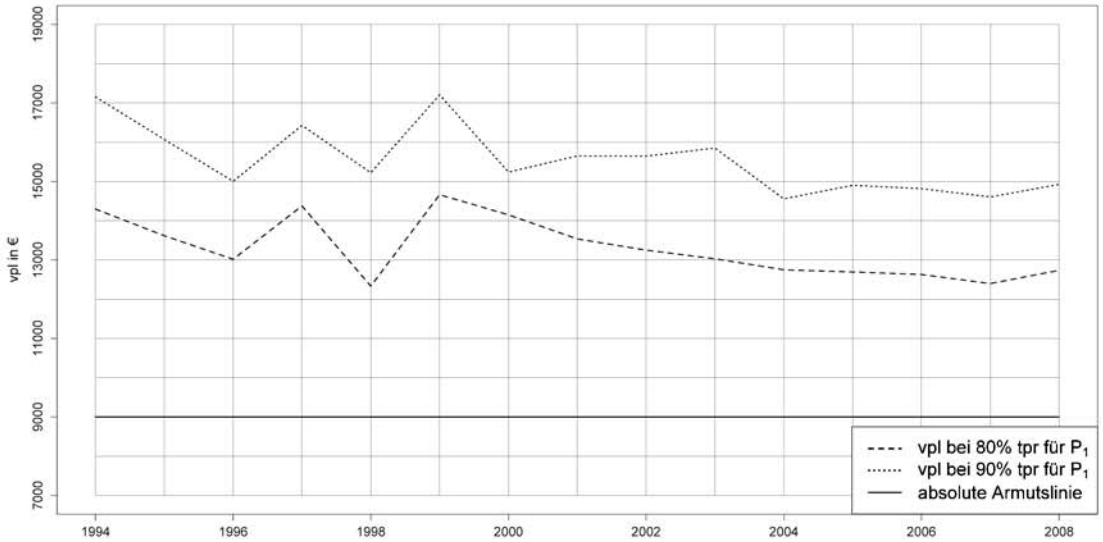


Abbildung 4.6: *Historische VPLs bei $TPR = 80\%$ und 90% erzeugt durch Modell P_6 (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Für das Modell P_1 schwankt die historische 80%-VPL im Beobachtungszeitraum um den Wert 15 500 €, während die 90%-VPL um den Wert 17 000 € schwankt.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

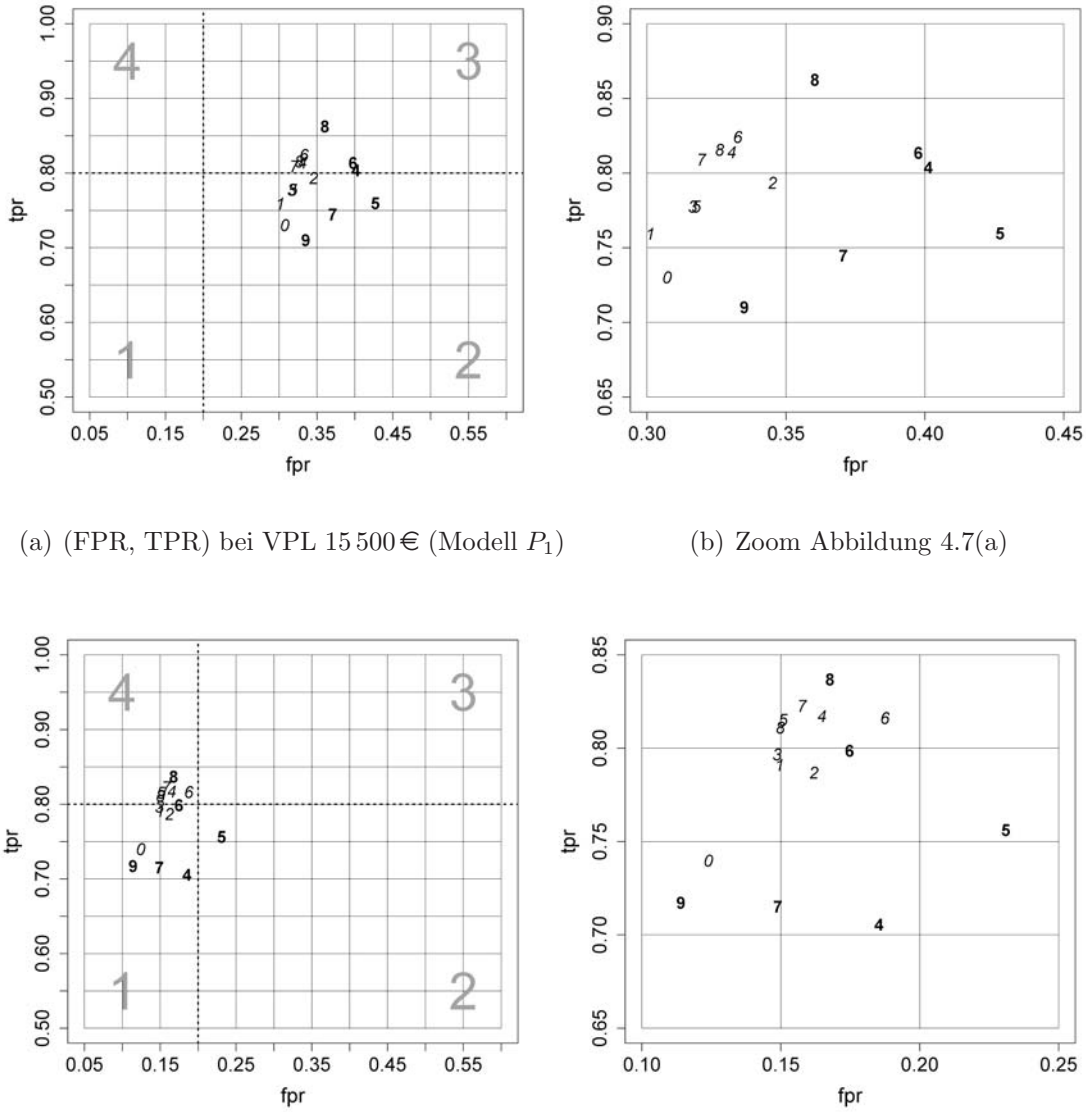
Die Niveaus der historischen VPLs für das Modell P_6 liegen etwas niedriger, für 80% ergibt sich 13 000 € und für 90% etwa 15 500 €. Somit müssen, um 80% (90%) der armen Haushalte zu identifizieren, alle Haushalte, deren prognostiziertes Einkommen mit Modell P_1 kleiner als 15 500 € (17 000 €) und mit Modell P_6 kleiner als 13 000 € (15 500 €) sind, als armutsgefährdet angesehen werden.

Die fixierten VPLs sollen nun genutzt werden, um die Vulnerabilität im Beobachtungszeitraum zu schätzen. Die Performance besteht dann nicht mehr ausschließlich aus der FPR sondern, da nun auch die TPR variiert, aus einem Paar (FPR, TPR). Die Performance wird in den folgenden Abbildungen für die VPLs 15 500 €/13 000 € und 17 000 €/15 500 € als Scatterplot dargestellt. Anstelle von Punkten wird die jeweils letzte Zahl eines Jahres, für das die Vulnerabilität geschätzt wurde, verwendet. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen. Der Vorteil dieser Darstellung besteht in der Möglichkeit zur Interpretation der Ergebnisse. Erstens ist die Darstellung besser lesbar, da Überschneidungen von Zahlen reduziert werden. Zweitens hebt es optisch den Unterschied der Genauigkeit der Schätzer zwischen den Jahren 1994-1999 einerseits und 2000-2008 andererseits hervor. Bei der Betrachtung von Armut ist festzustellen, dass seit dem Jahr 2000 die Ungleichheit zwischen den Haushalten und die Armut von Haushalten kontinuierlich angestiegen ist, so dass vermutet werden kann, dass sich auch die Genauigkeit der Vulnerabilitäts-Schätzung verändert hat.

Die Klassifizierung der Schätzer erfolgt über zwei Kriterien, die durch die Horizontale $TPR = 80\%$ (90%) und die Vertikale $FPR = 20\%$ ($FPR = 30\%$) beschrieben werden. Die Geraden teilen den Graphen in vier Quadranten 1-4 (graue Zahlen in der Abbildung) ein. Ein Schätzer wird als *genau* bezeichnet, wenn die $FPR \leq 20\%$ ($FPR \leq 30\%$) (Kriterium 1) und die $TPR \geq 80\%$ (90%) (Kriterium 2) ist (Quadrant 4). In diesem Fall werden viele der als armutsgefährdet geschätzten Haushalte tatsächlich arm, während „relativ“ wenige Haushalte (absolute Anzahl hängt von der Armutsquote ab), die als armutsgefährdet geschätzt wurden, nicht arm werden. Liegen die Punkte in Quadrant 1, so werden zwar relativ wenige Haushalte falsch aber auch wenige richtig als armutsgefährdet geschätzt. Quadrant 3 zeigt den anderen Fall; es werden sehr viele richtig aber auch viele falsch als armutsgefährdet geschätzt. Schätzer, die im Quadranten 2 liegen, werden *ungenau* genannt, da hier weder das Kriterium 1 noch das Kriterium 2 erfüllt ist.

Vergleicht man die Schätzer der Vulnerabilität in Abbildung 4.7(a) und 4.7(c), so werden die Ergebnisse aus Abschnitt 4.1.1 bestätigt. Bei der Schätzung der Vulnerabilität mit Querschnittsdaten muss ein deutlich höherer Anteil falsch als

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) (FPR, TPR) bei VPL 15 500 € (Modell P_1)

(b) Zoom Abbildung 4.7(a)

(c) (FPR, TPR) bei VPL 13 000 € (Modell P_6)

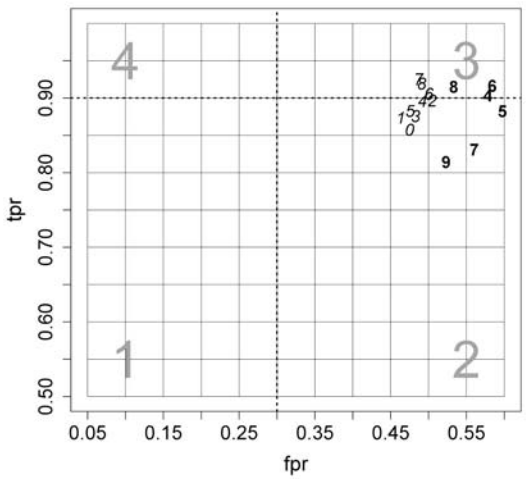
(d) Zoom Abbildung 4.7(c)

Abbildung 4.7: *Genauigkeit der Vulnerabilität im Zeitraum 1994-2008 (Niveau etwa bei 80% TPR). [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

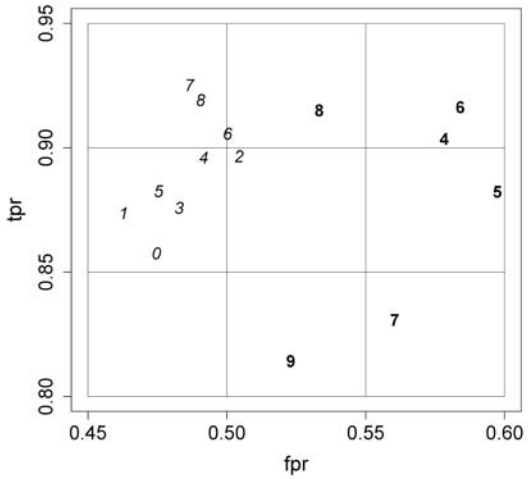
armutsgefährdet geschätzt werden als mit Paneldaten und Vorjahreseinkommen. Die Schätzer von P_6 sind, mit Ausnahme von vier Ausreißern in den 90er Jahren (1994, 1995, 1997 und 1999) und dem Jahr 2000, relativ stabil, während die durch P_1 erzeugten Schätzer stärker variieren. Soll ein höherer Anteil der zukünftig armen Haushalte richtig geschätzt werden (etwa 90%), müssen deutlich mehr falsch

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

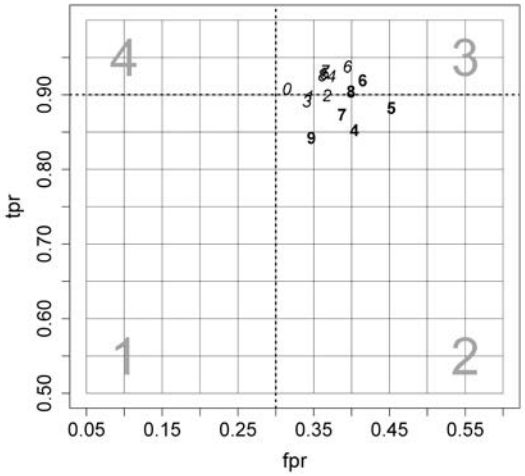
als armutgefährdet klassifizierte Haushalte in Kauf genommen werden, wie die Abbildungen 4.8(a) und 4.8(c) zeigen. Die Genauigkeit gemessen durch die TPR liegt relativ stabil bei 90% TPR, wobei die Jahre 1994, 1995, 1997 und 1999 wieder etwas größere Abweichungen zeigen. Ingesamt betrachtet scheint sich die Vulnerabilität seit etwa dem Jahr 2001 mit den verwendeten Methoden stabiler messen zu lassen.



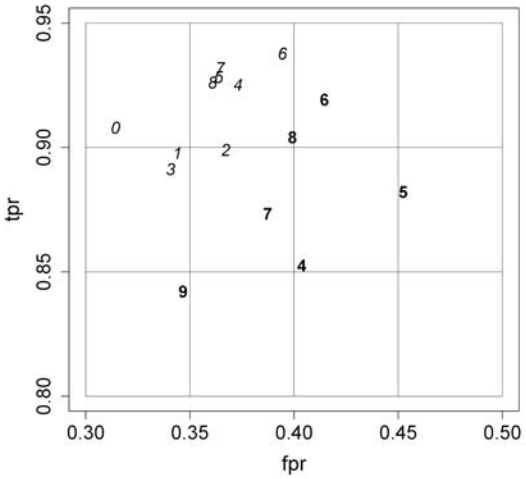
(a) (FPR, TPR) bei VPL 17 000 € (Modell P_1)



(b) Zoom Abbildung 4.8(a)



(c) (FPR, TPR) bei VPL 15 500 € (Modell P_6)



(d) Zoom Abbildung 4.8(c)

Abbildung 4.8: *Genauigkeit der Vulnerabilität im Zeitraum 1994-2008 (Niveau etwa bei 90% TPR) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.1.3 Profile von armutsgefährdeten Haushalten

Abgesehen von der Bewertung der Genauigkeit von Vulnerabilitäts-Schätzern ist es ebenfalls nützlich, die bestimmenden Faktoren der Vulnerabilität in Deutschland zu bestimmen. Um die armutsgefährdeten Gruppen von Haushalten zu identifizieren, wird der Status der Vulnerabilität von Haushalten unter Berücksichtigung einer Reihe von Haushaltseigenschaften (siehe z.B. Ligon und Schechter (2003) oder Chaudhuri et al. (2002)) untersucht. Es werden folgende Eigenschaften betrachtet: Familienstand, Beschäftigung des Haushaltsvorstandes (inklusive der Information (zum Befragungszeitpunkt) nicht erwerbstätig) und das Bundesland, in dem der Haushalt ansässig ist. Dazu werden die Einkommen mit den Modellen P_1 und P_6 geschätzt und ein Haushalt mit den zugehörigen VPLs, die zu einer TPR von 80% führen, als armutsgefährdet oder nicht-armutsgefährdet klassifiziert.[15] Für jeden Faktor wird der Anteil der (hochgerechneten) armutsgefährdeten Haushalte für jede Faktorstufe an allen Haushalten und nicht nur an denen mit einem Einkommen von 30 000 € berechnet. Die Profile für die Modelle P_1 und P_6 sind in den Tabellen 4.7-4.10 und den Abbildungen 4.9 und 4.10 dargestellt. Z.B. führt das geschätzte Einkommen von P_1 für 1996 und einer VPL korrespondierend zu $TPR = 80\%$ dazu, dass 33% der Haushalte, die in 1995 in Niedersachsen ansässig waren, als armutsgefährdet geschätzt werden (siehe Tabelle 4.7).

Generell werden mit dem Modell P_1 deutlich mehr Haushalte in den Faktorstufen als armutsgefährdet geschätzt als mit dem Modell P_6, da P_1 einen höheren Anteil falsch als armutsgefährdet schätzt.

Die Tabellen 4.7 und 4.8 zeigen den Anteil von Haushalten, die mit den Modellen P_1 und P_6 in jedem Bundesland für den Beobachtungszeitraum als armutsgefährdet klassifiziert werden. Allgemein sind die Anteile in den neuen Bundesländern höher. Für Schätzer des Modells P_6 fallen diese jedoch im Beobachtungszeitraum. Das zeigt, dass sich die Möglichkeit zur Schätzung der Vulnerabilität mit dem Modell P_6 verbessert hat.

Die Abbildungen 4.9 und 4.10 zeigen, dass bei der Schätzung durch beide Modelle ein höherer Anteil von geschiedenen oder getrennten Haushalten als armutsgefährdet geschätzt wird als verheiratete oder alleinstehende Haushalte.

[15] Ebenso wurden die Haushalte aufgrund der VPLs 15 500 € für Modell P_1 und 13 000 € für Modell P_6 geschätzt. Da sich die Ergebnisse für die Schätzer bei 80% übersichtlicher darstellen lassen, wurde in dieser Arbeit diese Betrachtungsweise gewählt. Beide Darstellungsweisen führen jedoch zu ähnlichen Ergebnissen und denselben Aussagen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Bundesland	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
Schleswig-Hols.	33	35	29	33	25	32	35	34	31	28	22	24	26	23	35
* Hamburg	52	52	54	61	38	50	43	51	38	36	37	48	45	39	42
Niedersachsen	32	36	33	41	25	35	30	27	24	24	21	22	21	19	19
* Bremen	53	62	58	68	33	60	45	39	37	44	28	22	22	27	20
Nordrhein-Westfa.	35	40	31	43	23	36	35	32	30	28	26	28	26	27	30
Hessen	35	39	29	43	23	34	27	31	28	27	25	32	23	23	25
R-Pfalz,Saarl.	29	34	26	44	25	37	31	33	30	30	28	27	27	27	27
Berlin (West)	33	36	33	37	29	43	38	28	34	31	26	29	32	35	35
Berlin (Ost)	36	35	30	44	27	41	37	44	40	40	41	40	45	38	42
Mecklenburg-V.	48	51	43	56	37	50	44	47	44	39	40	39	45	41	37
Brandenburg	40	43	36	50	30	44	38	40	37	35	32	37	35	31	32
Sachsen-Anhalt	49	51	43	56	35	51	42	46	38	39	34	43	36	36	40
Sachsen	47	50	46	51	36	47	40	45	40	42	42	41	42	41	41
Thüringen	49	57	52	61	40	61	44	48	41	40	37	41	44	38	38
Baden-Wuerttemb.	31	36	30	42	23	32	28	28	23	21	20	23	24	21	22
Bayern	26	34	27	37	21	32	23	25	25	27	22	24	22	21	23

Tabelle 4.7: *Anteil der als armutsgefährdet klassifizierten Haushalte nach Bundesland unter Verwendung von P_1 bei $TPR = 80\%$ in % (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Durch * gekennzeichnete Faktorstufen sind relativ dünn besetzt.*

Bundesland	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
Schleswig-Hols.	27	24	15	24	9	21	12	16	12	12	11	11	12	14	15
* Hamburg	37	32	30	45	19	33	28	21	20	15	20	22	19	16	13
Niedersachsen	26	22	18	27	16	21	16	14	15	14	14	14	17	12	13
* Bremen	32	33	24	32	9	32	10	23	22	37	20	13	6	7	7
N-Rhein-Westfa.	25	25	15	22	11	23	16	15	16	14	13	14	15	14	16
Hessen	21	20	16	23	13	20	16	14	19	16	17	18	15	14	15
R-Pfalz,Saarl.	27	28	18	30	17	22	19	20	21	19	18	16	19	17	23
Berlin (West)	16	18	18	16	10	17	18	16	18	22	16	15	20	23	20
Berlin (Ost)	30	27	19	28	16	18	14	29	23	20	25	25	31	21	29
Mecklenburg-V.	52	45	33	39	26	33	30	31	27	28	22	22	32	23	21
Brandenburg	43	35	23	34	10	28	24	21	23	18	19	21	22	18	20
Sachsen-Anhalt	48	38	28	35	19	34	27	28	25	21	23	22	25	20	21
Sachsen	50	46	31	36	17	31	26	27	23	22	23	24	29	22	22
Thüringen	49	44	33	43	24	40	29	27	24	21	25	25	32	24	22
Baden-Wuerttemb.	23	22	15	23	12	19	15	13	13	12	10	10	11	11	14
Bayern	21	22	14	18	13	18	14	13	16	16	13	12	13	10	13

Tabelle 4.8: *Anteil der als armutsgefährdet klassifizierten Haushalte nach Bundesland unter Verwendung von P_6 bei $TPR = 80\%$ in % (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Durch * gekennzeichnete Faktorstufen sind relativ dünn besetzt.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

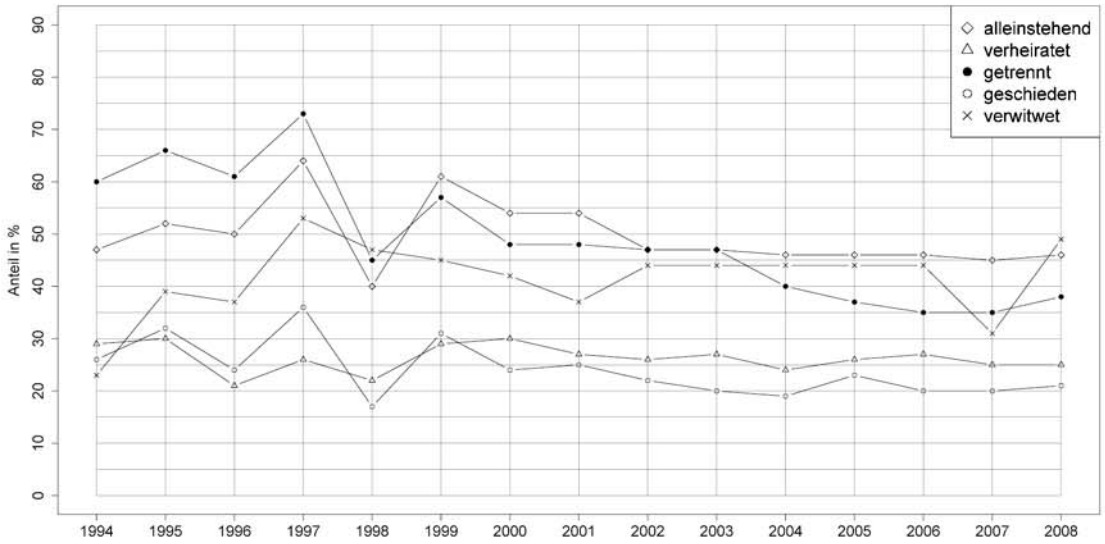


Abbildung 4.9: *Anteil der als armutsgefährdet klassifizierten Haushalte nach Familienstand unter Verwendung von P_1 bei $TPR = 80\%$ in % (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

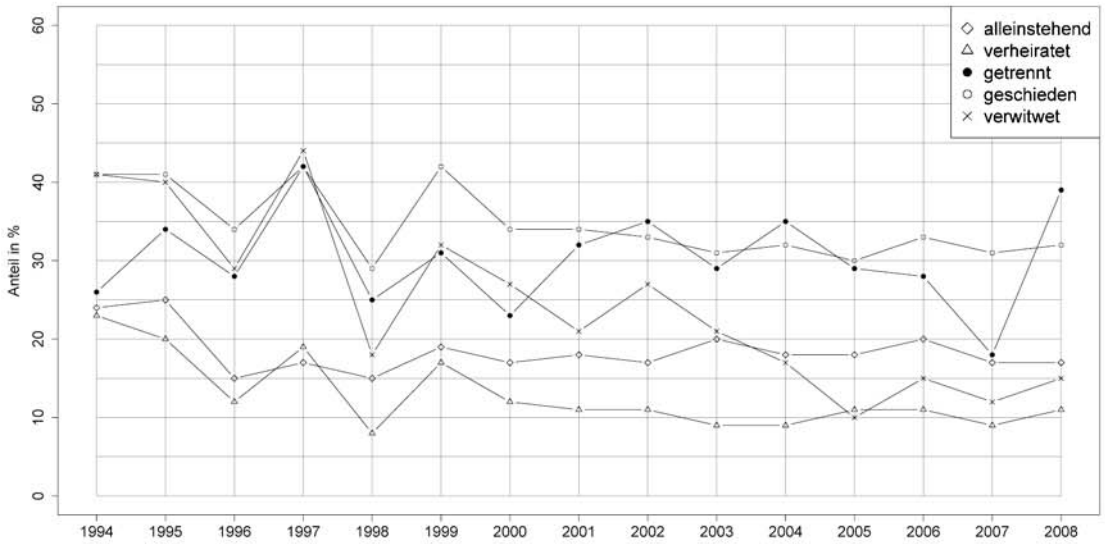


Abbildung 4.10: *Anteil der als armutsgefährdet klassifizierten Haushalte nach Familienstand unter Verwendung von P_6 bei $TPR = 80\%$ in % (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Gewerbe	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
nicht erwerbstätig	63	68	60	72	47	62	54	53	52	51	45	48	47	45	47
* Landwirtschaft	47	54	38	56	44	63	49	40	24	33	31	45	37	22	32
* Bergbau	24	22	21	13	9	34	28	39	14	1	28	20	23	32	49
Handel	18	33	20	38	15	35	29	27	27	25	20	22	25	19	28
Bau	15	18	13	22	6	26	25	26	14	9	12	9	13	12	8
Produktion	13	18	10	19	7	19	15	13	11	10	9	10	6	7	6
Transport	10	9	6	15	3	16	6	21	13	11	9	14	6	7	11
Dienstleistungen	6	7	8	13	4	13	10	9	11	10	11	12	10	10	13
Bank, Versicherung	6	4	5	7	4	1	2	3	2	2	1	6	3	1	2
* Energie	1	5	0	0	0	1	1	1	4	5	2	14	3	8	0

Tabelle 4.9: *Anteil der als armutsgefährdet klassifizierten Haushalte nach Gewerbe unter Verwendung von P_1 bei $TPR = 80\%$ in % (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Durch * gekennzeichnete Faktorstufen sind relativ dünn besetzt.*

Gewerbe	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
nicht erwerbstätig	46	43	32	43	24	37	31	28	30	28	26	25	28	25	26
* Landwirtschaft	56	44	16	40	32	33	34	16	19	23	32	33	21	5	13
* Bergbau	20	15	16	11	7	27	23	19	14	0	16	2	23	49	0
Handel	18	27	14	19	9	19	12	15	15	13	12	15	17	10	16
Bau	15	15	8	15	6	18	11	12	9	7	6	4	8	3	4
Produktion	17	15	7	10	3	13	5	6	7	5	5	4	4	5	3
Transport	13	8	5	9	3	12	4	6	5	6	3	9	6	3	3
Dienstleistungen	8	6	6	6	3	8	6	7	7	6	7	6	8	6	9
Bank, Versicherung	1	3	2	7	1	0	3	2	1	0	0	3	1	2	0
* Energie	2	8	0	0	0	5	1	3	2	5	2	14	3	8	0

Tabelle 4.10: *Anteil der als armutsgefährdet klassifizierten Haushalte nach Gewerbe unter Verwendung von P_6 bei $TPR = 80\%$ in % (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Durch * gekennzeichnete Faktorstufen sind relativ dünn besetzt.*

Der Anteil mit einem verwitweten HV ist zwischen 1994 und 1999 deutlich gefallen und dann nach einem leichten Anstieg etwa konstant geblieben. Aufgrund der höheren Lebenserwartung sind dies meist Haushalte mit weiblichem HV. Waren diese früher häufig nicht berufstätig und mussten von einer Witwenrente leben, so könnte der Rückgang darauf zurückzuführen sein, dass mittlerweile mehr Frauen erwerbstätig und solche Haushalte damit besser gestellt sind als noch vor einiger Zeit (BMFSFJ (2004), Goebel und Grabka (2011)).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Tabellen 4.9 und 4.10 quantifizieren den Effekt des Gewerbes des HVs auf die Vulnerabilität. Wie erwartet, wird ein hoher Anteil der Haushalte mit nicht-erwerbstätigem HV (arbeitslos oder keine Erwerbsperson) oder im Gewerbe der Landwirtschaft (z.B. saisonale Arbeit) als armutsgefährdet geschätzt. Erstgenannte sind ab etwa 2002 relativ konstant, während letztere relativ stark fluktuieren. Von den Haushalten, in denen der Haushaltsvorstand im Bereich Dienstleistungen, Bank/Versicherung oder Energiesektor beschäftigt war, wird nur ein kleiner Anteil als armutsgefährdet geschätzt. Es ist zu beachten, dass der Stichprobenumfang in einigen Gruppen recht klein ist und dass somit die Ergebnisse von diesen Gruppen mit Vorsicht betrachtet werden müssen.[16]

4.1.4 Schätzung der Vulnerabilität über n Jahre

Die bisher beschriebenen Ergebnisse bezogen sich auf die Schätzung der Vulnerabilität in der nächsten Periode. Dieser Abschnitt befasst sich mit der Schätzung der Vulnerabilität in n Perioden. Analog zu der Definition von Pritchett et al. (2000) wird die Vulnerabilität über n Jahre in dieser Arbeit als die Gefahr, in mindestens einer der nächsten n Perioden in Armut zu fallen, gemessen. Die Abbildungen 4.11 und 4.12 sowie Tabelle 4.11 zeigen die ROCs bzw. die Performance der „n-Jahres-Vulnerabilität“, geschätzt durch P_1 und P_6, wobei $n = 1$ (die bisherige Betrachtung), 2, 3, 4 gewählt wird. Die ROCs beziehen sich auf 1996 ($n = 1$) bis 1999 ($n = 4$) basierend auf verfügbaren Daten für sieben Wellen. Wenn n steigt, d.h., je weiter der Horizont der Vulnerabilität in die Zukunft reicht, desto kleiner wird die Fläche unter der ROC-Kurve und umso mehr verschlechtert sich die Performance der Schätzer von beiden Modellen.

	P_1		P_6	
n	FPR	VPL	FPR	VPL
1	0.39	15420	0.24	13852
2	0.5	16158	0.34	14837
3	0.53	16473	0.40	15554
4	0.58	16935	0.45	16075

Tabelle 4.11: *FPR und VPL für $TPR = 0.8$ für 1 bis 4 Jahre in der Zukunft [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

[16] Bzgl. des Gewerbes sind sehr wenig Daten für die Gruppe Bergbau vorhanden, was die starke Variation der Anteile erklärt. Dasselbe gilt auch für den Sektor Landwirtschaft und Energie.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

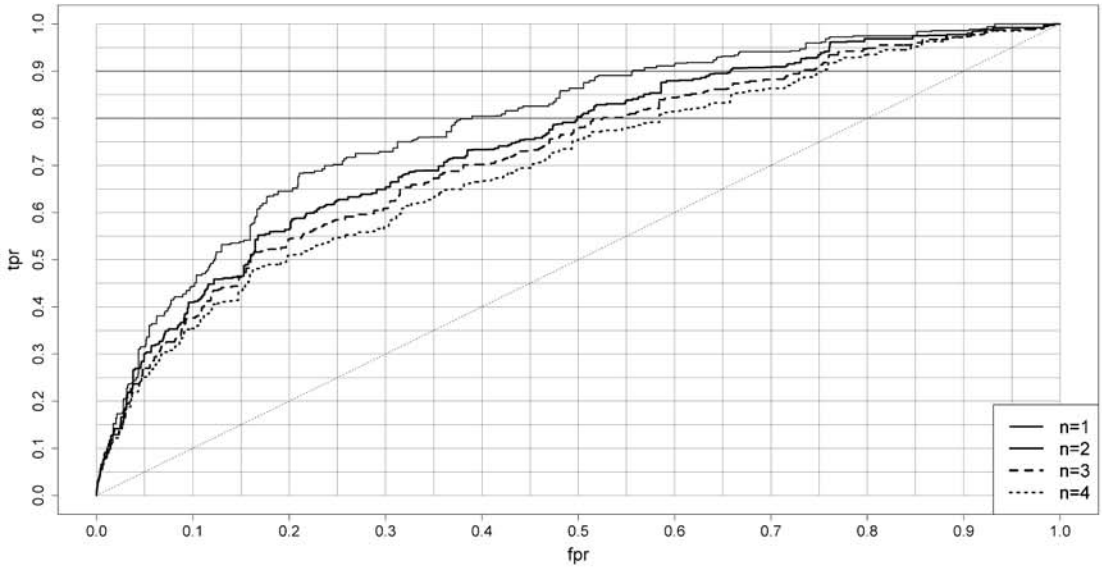


Abbildung 4.11: *ROC-Kurven für die n-Jahres-Vulnerabilität für* $n = 1, 2, 3, 4$, *erzeugt durch Modell* P_1 *(1996-1999) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

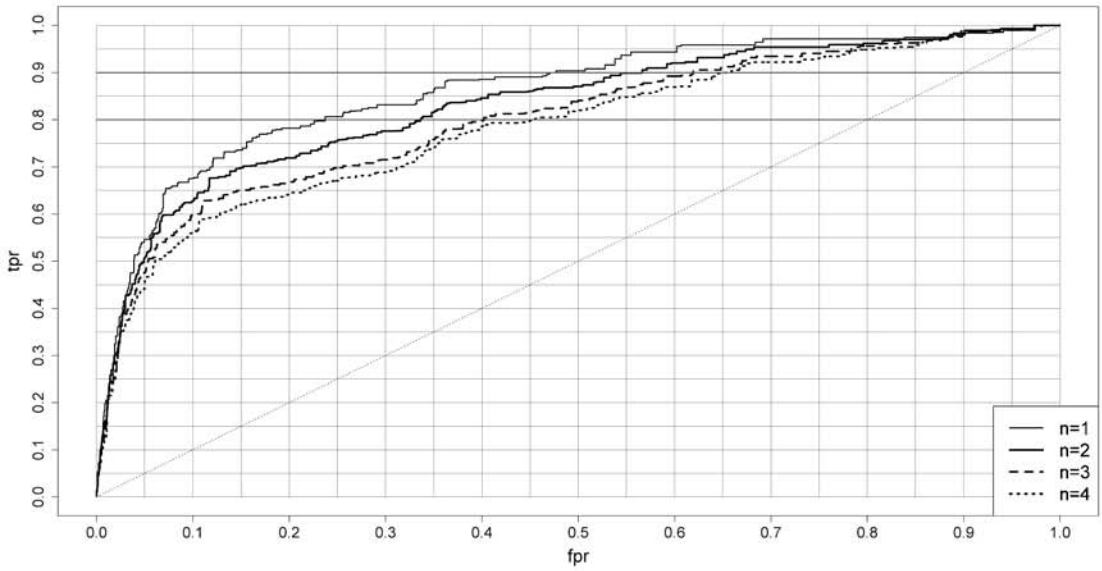


Abbildung 4.12: *ROC-Kurven für die n-Jahres-Vulnerabilität für* $n = 1, 2, 3, 4$, *erzeugt durch Modell* P_6 *(1996-1999) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.1.5 Zwischenfazit

In diesem Abschnitt wurde mit dem SOEP und den Modellen aus Abschnitt 2.4.1 untersucht, welche Kovariablen für die Genauigkeit der Schätzung der Vulnerabilität relevant sind, und wie genau sich die Vulnerabilität mit diesen Werkzeugen in Querschnitts- und in Paneldatensätzen schätzen lässt. Genauigkeit wurde im Kontext der ROC verstanden, d.h. die Schätzung soll möglichst viele der zukünftig armen Haushalte identifizieren und möglichst wenige Haushalte falsch als armutsgefährdet klassifizieren.

Die Ergebnisse der Analysen ergaben, dass sich die Vulnerabilität mit Paneldaten, wie erwartet, deutlich genauer schätzen ließ als mit Querschnittsdaten. Während die vorhandenen Haushaltseigenschaften kaum einen Einfluss auf die Schätzung der Vulnerabilität haben, verbessert das Vorjahreseinkommen als Kovariable die Schätzung erheblich. Allerdings sind selbst die Schätzer mit Paneldaten und Vorjahreseinkommen für Deutschland (und Ländern mit niedriger Armutsquote) nicht besonders genau. Absolut gesehen werden wenige Haushalte (aufgrund der niedrigen Armutsquote), die in der nächsten Periode arm werden, identifiziert und relativ viele (etwa das Doppelte der richtig als armutsgefährdet klassifizierten Haushalte) falsch als armutsgefährdet deklariert.

Hinsichtlich der Stabilität der Schätzer im Beobachtungszeitraum ist zu beobachten, dass sich insbesondere im Modell mit Paneldaten, die Schätzer ab dem Jahr 2001 für eine konstant gewählte VPL von 13 000 € relativ stabil verhalten. Durch die Berechnung von Vulnerabilitäts-Profilen für verschiedene Eigenschaften (Familienstand des Haushaltsvorstandes, Gewerbe, ansässiges Bundesland) ergab sich, dass einige Gruppen stärker von Vulnerabilität betroffen sind als andere. Zu den stärker betroffenen Gruppen zählen die geschiedenen bzw. getrennten Haushalte, die Haushalte mit nicht-erwerbstätigen Haushaltsvorständen sowie die Haushalte, die in Ostdeutschland ansässig sind.

Wird anstelle der Vulnerabilität in der nächsten Periode diejenige über n Perioden betrachtet, so nimmt die Genauigkeit der Schätzung, wie erwartet, ab je weiter n in der Zukunft liegt.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.2 Schätzung bei ungenauer Einkommensinformation

In diesem Abschnitt wird mit den in Abschnitt 2.4.2 beschriebenen Methoden empirisch untersucht, wie genau sich die Vulnerabilität schätzen lässt, wenn nur *ungenaue* Informationen über das Einkommen der Haushalte vorhanden sind. Unter ungenauer Information wird in diesem Kontext verstanden, dass nur die (grobe) Klassenzugehörigkeit der Haushalte zu einer von zwei bzw. drei Einkommensklassen bekannt ist. Vorausgesetzt wird, dass kurze Paneldaten, die Haushaltskovariablen aus dem vorhergehenden Abschnitt sowie die Information, ob ein Haushalt arm geworden ist, verfügbar sind. Die Genauigkeit der Schätzer wird im Vergleich zu der Genauigkeit der Schätzer von P_1 und P_6 des vorherigen Abschnittes bewertet.

Die Vulnerabilität wird wieder anhand von rollierenden 4-Jahres-Panels im Zeitraum 1992-2009 geschätzt. Für die Methoden M_{1K} und M_{2K} aus Abschnitt 2.4.2 wird die Information der ersten beiden Wellen genutzt, um die Wahrscheinlichkeit, in der dritten Welle in Armut zu fallen, zu schätzen. In Methode M_{2K} wird die Stichprobe der Haushalte in der ersten Welle nach der Zugehörigkeit zu einer Einkommensklasse in Teilstichproben zerlegt, und es werden für die Modelle der Einkommensklassen die Regressionskoeffizienten geschätzt. Mit diesen sowie den Haushaltsvariablen der zweiten Welle werden für die Zerlegungen nach Einkommensklassen in der zweiten Welle die Wahrscheinlichkeit der Armut in der dritten Welle getrennt geschätzt. Nach der Schätzung werden die Teilstichproben wieder zusammengeführt. Die Information, ob Haushalte arm geworden sind, ergibt sich aus Welle vier für Welle drei. Um die ROC zu konstruieren, werden Paare (FPR, TPR) für jede mögliche VPL auf Basis der vorhergesagten *Wahrscheinlichkeit* für die dritte Welle ermittelt. Wenn die VPL *kleiner* gewählt wird (Eintrittswahrscheinlichkeit, ab der ein Haushalt als armutsgefährdet klassifiziert wird, ist geringer, so dass mehr Haushalte als armutsgefährdet klassifiziert werden), so steigen sowohl die TPR als auch die FPR.

Da die verwendete Einkommensvariable aus dem CNEF stetig ist, besteht der erste Schritt in der Definition von Einkommensklassen (Abschnitt 4.2.1). Diese werden auf Basis der Vorhersage des Jahres 2006, das rein zufällig gewählt wird, vorgeschlagen. In dieser Arbeit werden zwei bzw. drei Einkommensklassen berücksichtigt. Die erste Klasse wird bei 0 € (L_{1fix}) begonnen, während die letzte Klasse bei 30 000 € (L_{Kfix}) endet. Wie bereits im vorherigen Abschnitt werden nur Haushalte, die

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ein Äquivalenzeinkommen haben, das kleiner als 30 000 € ist, betrachtet. Für zwei Einkommensklassen muss noch eine Klassengrenze L zwischen der ersten und der zweiten Einkommensklasse gewählt werden, während bei drei Einkommensklassen zwei Grenzen L_1 und L_2 zwischen der ersten und der zweiten sowie zwischen der zweiten und der dritten Einkommensklasse festgelegt werden müssen. Diese zu wählenden Klassengrenzen werden zunächst variabel gehalten. Für verschiedene L bzw. L_1 und L_2 sind in Abschnitt 4.2.1 ausgewählte Regressionstabellen dargestellt. Die Armutswahrscheinlichkeiten werden jeweils für verschiedene L bzw. L_1 und L_2 in Welle drei geschätzt und die ROC-Kurven konstruiert. Die Genauigkeit, mit der die Vulnerabilität bei den verwendeten Einkommensklassen geschätzt werden kann, wird wieder durch die FPR bei 80% TPR gemessen. Die Einkommensklassen, die zur kleinsten FPR führen, sind im Kontext dieser Arbeit eine sinnvolle Wahl. Das Vorgehen wird für zwei und drei Einkommensklassen für die Methoden M_{1K} und M_{2K} für alle Jahre durchgeführt und mit den Ergebnissen von P_1 und P_6 verglichen. In Abschnitt 4.2.2 wird untersucht,

- welche VPL zu einer TPR von 80% führt und
- ob die FPR bei konstant gewählter VPL über die Zeit stabil ist.

Abschnitt 4.2.3 untersucht für das beliebig gewählte Jahr 2001, um wieviel sich die Genauigkeit der Vulnerabilitäts-Schätzer verringert, wenn die Haushalte den Einkommensklassen fehlerhaft zugeordnet werden. Dazu werden verzerrte Einkommen simuliert, indem die stetigen Einkommen additiv mit einem gleichverteilten bzw. normalverteilten Fehler verknüpft werden. Es werden verschiedene Varianzen und Abhängigkeitsstrukturen der Fehler angenommen. Die fehlerbehafteten Einkommen werden dann auf die Einkommensklassen verteilt, die Armutswahrscheinlichkeit geschätzt und die ROC konstruiert. Um die Ergebnisse mit den vorherigen vergleichbar zu machen, wird festgelegt, dass Einkommen, die nach der Fehlerbehaftung negativ werden, der untersten und Einkommen von 30 000 € oder oberhalb dieser Grenze der obersten Einkommensklasse zugeordnet werden. Abschnitt 4.2.4 fasst die in diesem Abschnitt erzielten Ergebnisse zusammen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.2.1 Wahl sinnvoller Einkommensklassen

Die Wahl sinnvoller Einkommensklassen erfolgt auf Basis der Schätzung der Vulnerabilität für das Jahr 2006. Dieses Jahr ist beliebig gewählt, andere Jahre aus dem Beobachtungszeitraum führen zu ähnlichen Ergebnissen. Zunächst wird die Schätzung der Vulnerabilität basierend auf zwei Einkommensklassen betrachtet, d.h. mit den Methoden M_{12} und M_{22} aus Abschnitt 2.4.2. Die Klassengrenze L wird in Abhängigkeit der Armutslinie $z = 9\,000$ € definiert ($L(z) = kz$), wobei k als Sequenz zwischen 1 und 2 in Abständen von 0.1 gewählt wird. Um die Unterschiede in der Schätzung der Armutswahrscheinlichkeit zwischen den beiden Methoden und für verschiedene Einkommensklassen zu zeigen, werden beispielhaft die Regressionstabellen für das Modell $M_{12,06}$ (Beschreibung der neuen Einkommensvariablen siehe Tabelle 4.12; die Referenzstufen sind fett hinterlegt) mit $L = 1.3z$ und $L = 1.7z$ sowie für die Modelle mit getrennten Einkommensklassen in den Tabellen 4.13-4.16 dargestellt.

$$\widehat{P}(Z_{06} = 1|X_{05}, y_{05,2}) = \frac{\exp(X_{05}\widehat{\beta} + y_{05,2}\widehat{\gamma})}{1 + \exp(X_{05}\widehat{\beta} + y_{05,2}\widehat{\gamma})} \qquad (M_{12,06})$$

$$\begin{aligned} \widehat{P}(Z_{06} = 1|X_{05}^{(1)}, y_{05,2} = 1) &= \frac{\exp(X_{05}^{(1)}\widehat{\beta}_1)}{1 + \exp(X_{05}^{(1)}\widehat{\beta}_1)} \\ \widehat{P}(Z_{06} = 1|X_{05}^{(2)}, y_{05,2} = 0) &= \frac{\exp(X_{05}^{(2)}\widehat{\beta}_2)}{1 + \exp(X_{05}^{(2)}\widehat{\beta}_2)} \end{aligned} \qquad (M_{22,06})$$

Einkommensvariablen	Beschreibung
Einkommen in 2 Klassen	Faktor Einkommensklassen mit zwei Stufen (**Einkommen [L,30 000)**, Einkommen $[0, L)$)
Einkommen in 3 Klassen	Faktor Einkommensklassen mit drei Stufen (**Einkommen [$\mathbf{L_2}$,30 000)**, Einkommen $[L_1, L_2)$, Einkommen $[0, L_1)$)

Tabelle 4.12: *Beschreibung der Einkommensvariablen in den Regressionen* $M_{12,06}$-$M_{22,06}$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link	Variable	Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-4.86	0.01	-819.46	<0.001
	Einkommen $[0, L)$	2.41	0.00	1294.41	<0.001
	Geschlecht: weiblich	-0.15	0.00	-78.43	<0.001
	$n[0, 18)$	0.34	0.00	281.04	<0.001
	$n[18, 34)$	0.67	0.00	381.10	<0.001
	$n[34, 59)$	0.61	0.00	289.48	<0.001
	$n[59-)$	0.29	0.00	115.83	<0.001
	Familienstand: alleinstehend	0.99	0.00	289.18	<0.001
	Familienstand: verwitwet	0.47	0.00	130.08	<0.001
	Familienstand: geschieden	1.04	0.00	339.77	<0.001
	Familienstand: getrennt	0.35	0.00	75.52	<0.001
	Gesamtarbeitsstunden	0.00	0.00	-172.36	<0.001
	Gewerbe: Landwirtschaft	-0.37	0.01	-41.96	<0.001
	Gewerbe: Energiesektor	1.90	0.01	243.68	<0.001
	Gewerbe: Bergbau	-18.24	90.26	-0.20	0.84
	Gewerbe: Produktion	-0.65	0.01	-129.24	<0.001
	Gewerbe: Baugewerbe	-0.37	0.01	-67.84	<0.001
	Gewerbe: Handel	-0.11	0.00	-30.82	<0.001
	Gewerbe: Transport	-0.54	0.01	-87.42	<0.001
	Gewerbe: Bank/Versicherung	-0.42	0.01	-56.37	<0.001
	Gewerbe: Dienstleistungen	-0.27	0.00	-88.31	<0.001
	Wohnverhältnis: Hauptmieter	0.24	0.00	102.28	<0.001
	Wohnverhältnis: Untermieter	-0.37	0.01	-74.18	<0.001
	Größe des Wohnsitzes	-19.15	0.11	-172.25	<0.001
	(Größe des Wohnsitzes)^2	11.03	0.08	133.80	<0.001
	Schulabschluss: Realschulabschluss	0.19	0.00	90.36	<0.001
	Schulabschluss: Abitur/FHSR	0.06	0.00	24.34	<0.001
	Schulabschluss: andere Abschlüsse	0.38	0.00	125.82	<0.001
	Schulabschluss: kein Abschluss	0.69	0.00	189.59	<0.001
	n(Vollzeitbeschäftigung)	-29.98	0.18	-169.65	<0.001
	n(Vollzeitbeschäftigung)^2	4.65	0.13	35.81	<0.001
	n(Teilzeitbeschäftigung)	-0.16	0.00	-50.04	<0.001
				N	6095
				N_{gew}	25058614
				logL	-5516842
	Modellfamilie: binomial			AIC	11033749

Tabelle 4.13: *Regression für Modell $M_{12,06}$ mit $L = 1.3z$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link	Variable	Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-5.31	0.01	-903.19	<0.001
	Einkommen $[0, L)$	2.21	0.00	884.52	<0.001
	Geschlecht: weiblich	-0.09	0.00	-48.13	<0.001
	$n[0, 18)$	0.33	0.00	294.93	<0.001
	$n[18, 34)$	0.71	0.00	421.88	<0.001
	$n[34, 59)$	0.67	0.00	335.52	<0.001
	$n[59-)$	0.24	0.00	98.72	<0.001
	Familienstand: alleinstehend	1.05	0.00	327.89	<0.001
	Familienstand: verwitwet	0.11	0.00	32.19	<0.001
	Familienstand: geschieden	1.09	0.00	379.44	<0.001
	Familienstand: getrennt	0.52	0.00	115.64	<0.001
	Gesamtarbeitsstunden	0.00	0.00	-195.30	<0.001
	Gewerbe: Landwirtschaft	-0.07	0.01	-7.83	<0.001
	Gewerbe: Energiesektor	1.40	0.01	139.10	<0.001
	Gewerbe: Bergbau	-17.84	92.92	-0.19	0.85
	Gewerbe: Produktion	-0.93	0.00	-193.86	<0.001
	Gewerbe: Baugewerbe	-0.63	0.01	-119.25	<0.001
	Gewerbe: Handel	-0.19	0.00	-55.77	<0.001
	Gewerbe: Transport	-0.57	0.01	-96.79	<0.001
	Gewerbe: Bank/Versicherung	-0.03	0.01	-4.73	<0.001
	Gewerbe: Dienstleistungen	-0.42	0.00	-144.34	<0.001
	Wohnverhältnis: Hauptmieter	0.27	0.00	119.93	<0.001
	Wohnverhältnis: Untermieter	-0.57	0.00	-116.36	<0.001
	Größe des Wohnsitzes	-23.68	0.11	-225.24	<0.001
	(Größe des Wohnsitzes)^2	15.72	0.08	203.44	<0.001
	Schulabschluss: Realschulabschluss	0.13	0.00	64.51	<0.001
	Schulabschluss: Abitur/FHSR	-0.04	0.00	-14.51	<0.001
	Schulabschluss: andere Abschlüsse	0.41	0.00	143.51	<0.001
	Schulabschluss: kein Abschluss	0.94	0.00	275.90	<0.001
	n(Vollzeitbeschäftigung	-30.12	0.17	-181.02	<0.001
	n(Vollzeitbeschäftigung)^2	6.44	0.12	55.04	<0.001
	n(Teilzeitbeschäftigung)	-0.12	0.00	-40.78	<0.001
				N	6095
				N_{gew}	25058614
				logL	-5927152
	Modellfamilie: binomial			AIC	11854369

Tabelle 4.14: *Regression für Modell $M_{12,06}$ mit $L = 1.7z$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link	Variable	Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-1.39	0.01	-198.79	<0.001
	Geschlecht: weiblich	-0.27	0.00	-114.27	<0.001
	$n[0,18)$	0.38	0.00	256.56	<0.001
	$n[18,34)$	0.55	0.00	246.16	<0.001
	$n[34,59)$	0.58	0.00	218.38	<0.001
	$n[59-)$	0.40	0.00	124.26	<0.001
	Familienstand: alleinstehend	1.18	0.00	267.91	<0.001
	Familienstand: verwitwet	0.12	0.00	25.11	<0.001
	Familienstand: geschieden	0.93	0.00	240.84	<0.001
	Familienstand: getrennt	-0.14	0.01	-25.21	<0.001
	Gesamtarbeitsstunden	0.00	0.00	-135.88	<0.001
	Gewerbe: Landwirtschaft	-0.73	0.01	-70.89	<0.001
	Gewerbe: Energiesektor	18.50	103.90	0.18	0.86
	Gewerbe: Bergbau	-18.22	78.33	-0.23	0.82
	Gewerbe: Produktion	-0.57	0.01	-86.48	<0.001
	Gewerbe: Baugewerbe	-0.36	0.01	-46.44	<0.001
	Gewerbe: Handel	-0.24	0.00	-54.93	<0.001
	Gewerbe: Transport	-1.15	0.01	-132.88	<0.001
	Gewerbe: Bank/Versicherung	-1.10	0.01	-97.41	<0.001
	Gewerbe: Dienstleistungen	-0.56	0.00	-145.16	<0.001
	Wohnverhältnis: Hauptmieter	-0.38	0.00	-121.70	<0.001
	Wohnverhältnis: Untermieter	-1.03	0.01	-164.94	<0.001
	Größe des Wohnsitzes	-14.97	0.05	-274.49	<0.001
	(Größe des Wohnsitzes)^2	-2.60	0.06	-46.36	<0.001
	Schulabschluss: Realschulabschluss	0.18	0.00	68.79	<0.001
	Schulabschluss: Abitur/FHSR	0.01	0.00	2.07	0.04
	Schulabschluss: andere Abschlüsse	0.49	0.00	135.44	<0.001
	Schulabschluss: kein Abschluss	0.68	0.00	166.06	<0.001
	n(Vollzeitbeschäftigung	-5.93	0.07	-88.98	<0.001
	n(Vollzeitbeschäftigung^2	2.12	0.05	43.15	<0.001
	n(Teilzeitbeschäftigung	-0.33	0.00	-82.58	<0.001
				N	1151
				N_{gew}	5286036
				logL	-3162944
	Modellfamilie: binomial			AIC	6325951

Tabelle 4.15: *Regression für Modell $M_{22,06}$ mit $L = 1.3z$ (untere Einkommensgruppe) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link	Variable	Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-5.81	0.01	-555.30	<0.001
	Geschlecht: weiblich	0.05	0.00	14.71	<0.001
	$n[0,18)$	0.26	0.00	121.69	<0.001
	$n[18,34)$	0.73	0.00	249.40	<0.001
	$n[34,59)$	0.53	0.00	149.79	<0.001
	$n[59-)$	-0.26	0.00	-54.63	<0.001
	Familienstand: alleinstehend	0.61	0.01	103.32	<0.001
	Familienstand: verwitwet	1.01	0.01	172.49	<0.001
	Familienstand: geschieden	1.12	0.01	215.00	<0.001
	Familienstand: getrennt	1.32	0.01	185.52	<0.001
	Gesamtarbeitsstunden	0.00	0.00	-76.00	<0.001
	Gewerbe: Landwirtschaft	0.75	0.02	47.38	<0.001
	Gewerbe: Energiesektor	2.18	0.01	236.27	<0.001
	Gewerbe: Bergbau	-14.64	71.08	-0.21	0.84
	Gewerbe: Produktion	-0.65	0.01	-75.53	<0.001
	Gewerbe: Baugewerbe	-0.01	0.01	-1.46	0.14
	Gewerbe: Handel	0.12	0.01	18.18	<0.001
	Gewerbe: Transport	0.39	0.01	44.99	<0.001
	Gewerbe: Bank/Versicherung	0.42	0.01	42.70	<0.001
	Gewerbe: Dienstleistungen	0.08	0.01	14.89	<0.001
	Wohnverhältnis: Hauptmieter	1.17	0.00	266.96	<0.001
	Wohnverhältnis: Untermieter	0.77	0.01	89.77	<0.001
	Größe des Wohnsitzes	2.01	0.14	14.27	<0.001
	(Größe des Wohnsitzes)^2	10.43	0.09	122.75	<0.001
	Schulabschluss: Realschulabschluss	0.20	0.00	52.93	<0.001
	Schulabschluss: Abitur/FHSR	0.15	0.00	35.44	<0.001
	Schulabschluss: andere Abschlüsse	0.16	0.01	22.81	<0.001
	Schulabschluss: kein Abschluss	1.04	0.01	135.11	<0.001
	n(Vollzeitbeschäftigung)	-49.74	0.31	-162.64	<0.001
	n(Vollzeitbeschäftigung)^2	6.00	0.21	28.86	<0.001
	n(Teilzeitbeschäftigung)	-0.06	0.01	-11.39	<0.001
				N	4944
				N_{gew}	19772578
				logL	-2156734
	Modellfamilie: binomial			AIC	4313529

Tabelle 4.16: *Regression für Modell $M_{22,06}$ mit $L = 1.3z$ (obere Einkommensgruppe) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Die Regressionstabellen dokumentieren, dass alle Faktoren eine Wirkung auf die erwartete Armutswahrscheinlichkeit haben und die übrigen Kovariablen hoch signi-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

fikant sind. Die Ergebnisse sind allerdings aufgrund der Methode zur Berechnung der Standardfehler mit Vorsicht zu betrachten.[17]

Die Logit-Koeffizienten können interpretiert werden, indem das Odds Ratio (Effektkoeffizient) (=exp(β)) berechnet wird. Z.B. gibt das Odds Ratio bei der Interpretation der Ergebnisse für Modell $M_{12,06}$ für die untere Einkommensvariable an, um wieviel höher die Wahrscheinlichkeit für einen Haushalt in dieser Klasse ist, arm geschätzt zu werden, als wenn er in der oberen Einkommensklasse wäre. Für einen Haushalt in der unteren Einkommensgruppe (bei Wahl der Grenze bei $L = 1.3z$) ist die Armutswahrscheinlichkeit elf Mal so hoch als wenn er in der oberen Einkommensgruppe wäre. Bei der Grenze $L = 1.7z$ ist dieser Faktor etwas niedriger, nämlich neun Mal so hoch.

Hinsichtlich

- des Regressionskoeffizientens für Haushalte mit weiblichem Geschlecht unterscheiden sich die beiden Ergebnisse kaum. Die Variable hat wenig Einfluss.
- des Familienstands erhöht sich gegenüber der Referenzkategorie verheiratete Haushalte die Wahrscheinlichkeit bei beiden Modellen für alle anderen Kategorien zu der armen Gruppe zu gehören. Dies ist insbesondere für alleinstehende und geschiedene Haushalte der Fall (etwa drei Mal so hoch).
- der Erwerbststätigkeit vermindert sich die Wahrscheinlichkeit fast aller Berufsgruppen gegenüber der Referenzkategorie nicht-erwerbstätige HVs. Die Modelle unterscheiden sich im Hinblick auf einige Logit-Koeffizienten. Z.B. ist die Wahrscheinlichkeit, dass ein Haushalt mit HV, der in einer Bank beschäftigt ist und in der armen Kategorie liegt, geringer für das Modell mit k=1.3.
- des Wohnverhältnisses des Haushaltes und der Bildung überraschen die Ergebnisse auf den ersten Blick. Die Stufe Abitur hat kaum einen Einfluss. Für Untermieter sinkt die Wahrscheinlichkeit zur armen Kategorie zu gehören. Das könnte darauf zurückzuführen sein, dass in der unteren Einkommensklasse kaum Beobachtungen von höher qualifizierten oder Hausbesitzern sind. Weiterhin steigt die Wahrscheinlichkeit bei Zunahme der Personen im Haushalt (für alle Altersklassen) bei keinem vollbeschäftigtem Haushaltsmitglied. Sie sinkt bei steigender Anzahl der Teilzeitbeschäftigten und mit steigender Wohnungsgröße.

[17] Die Standardfehler werden hier nicht mit der Option `vcov` sondern mit der Option `qr`, also mit den individuellen Anpassungen der Parameter berechnet (Stasinopoulos und Rigby, 2007).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Eine Bewertung der Anpassungsgüte aufgrund des AIC Kriteriums ergibt, dass diese für das Modell mit Klassengrenze $L = 11\,700$ € höher ist als für das Modell mit $L = 13\,500$ €. Das AIC Kriterium misst jedoch nicht, mit welchem der beiden Modelle sich die Vulnerabilität genauer schätzen lässt. Das wird im Folgenden wieder mit der Methode ROC untersucht.

Beim Vergleich der Modelle der oberen und unteren Einkommensklasse muss beachtet werden, dass die Ergebnisse von den Eigenschaften der Haushalte in den Gruppen abhängt. In der unteren Einkommensgruppe sind vermutlich wenige Hausbesitzer oder HVs mit Abitur. Im Folgenden wird lediglich auf Unterschiede in den Ergebnissen der beiden Regressionen eingegangen. In der unteren Einkommensklasse sinkt die Wahrscheinlichkeit für Haushalte gegenüber der Referenzkategorien für Haushalte mit weiblichem HV, mit getrenntem HV, für Haupt- und Untermieter sowie für HVs mit Beschäftigung in den meisten Gewerben. Wohnungsgröße und die Anzahl der Vollbeschäftigten haben keinen Einfluss. Bei der oberen Einkommensgruppe sinkt die Armutswahrscheinlichkeit für steigende Anzahl der Personen über 60 Jahren im Haushalt. Dafür steigt die Wahrscheinlichkeit vieler Gewerbe gegenüber den nicht-erwerbstätigen HVs (möglicherweise zurückzuführen auf wohlhabendere HVs in Rente in dieser Gruppe). Die Anzahl der Vollbeschäftigten hat keinen Einfluss auf die Armutswahrscheinlichkeit. Sie steigt für Mieter und Untermieter gegenüber den Hausbesitzern und fällt mit steigender Größe des Wohnraums.

Die Untersuchung der Quantilsresiduen im Anhang (A.3.1) sind vereinbar mit der Annahme, dass die vier Modelle angemessen sind .

Abbildung 4.13 stellt die Genauigkeit (FPR bei $TPR = 80\%$) für elf verschiedene Klassengrenzen für die Methoden M_{12} und M_{22} dar. Die Ergebnisse der beiden Methoden unterscheiden sich nicht wesentlich. Mit Methode M_{12} werden etwas genauere Schätzer erzeugt, besonders wenn $k > 1.3$ gewählt wird. Die Klassengrenze spielt bei der Schätzung der Vulnerabilität mit Einkommensklassen eine wichtige Rolle. Verglichen mit der Genauigkeit der Schätzer von P_1 sind die Schätzer basierend auf (zwei) Einkommensklassen genauer und zwar für jede betrachtete Klassengrenze. Die Genauigkeit der Schätzer von P_6 wird nur für die Klassengrenzen $L = 2z = 10\,800$ € erreicht und von $L = 1.3z = 11\,700$ € sogar minimal übertroffen. Alle anderen Klassengrenzen führen zu weniger genauen Schätzern. Die Genauigkeit der Schätzer mit zwei Einkommensklassen nimmt umso mehr ab, je weiter die Klassengrenze (sowohl links als auch rechts) von 11 700 € entfernt gewählt wird.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

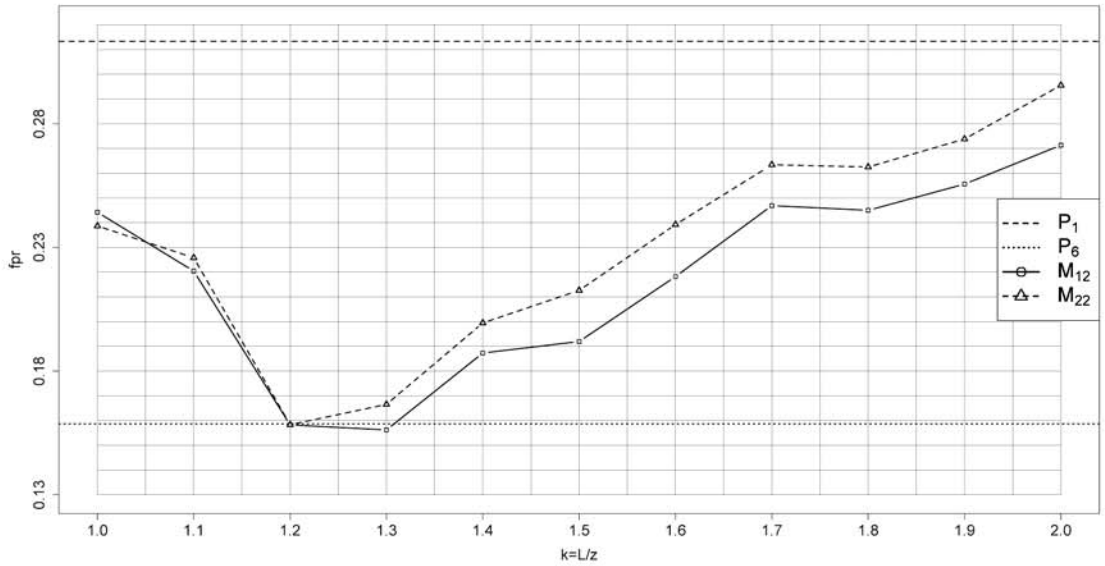
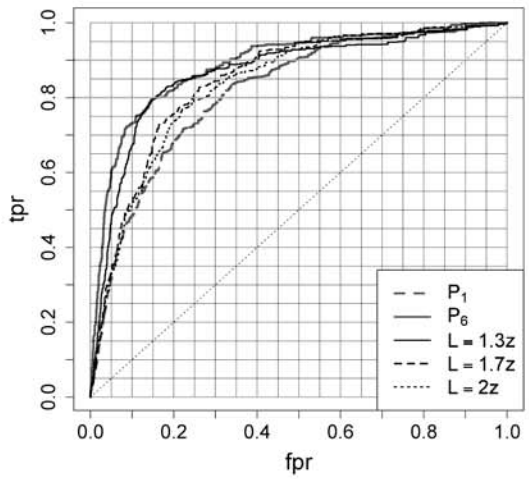
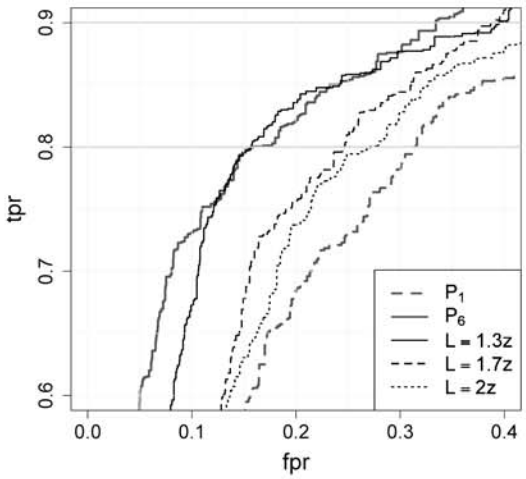


Abbildung 4.13: *Performance für zwei Einkommensklassen für 2006 [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Abbildungen 4.14(a) und 4.15(a) stellen ausgewählte ROC-Kurven für die Methoden M_{12} und M_{22} dar. Zur präziseren Darstellung der Genauigkeit werden in 4.14(b) und 4.15(b) Vergrößerungen der Umgebung um 80% TPR präsentiert.



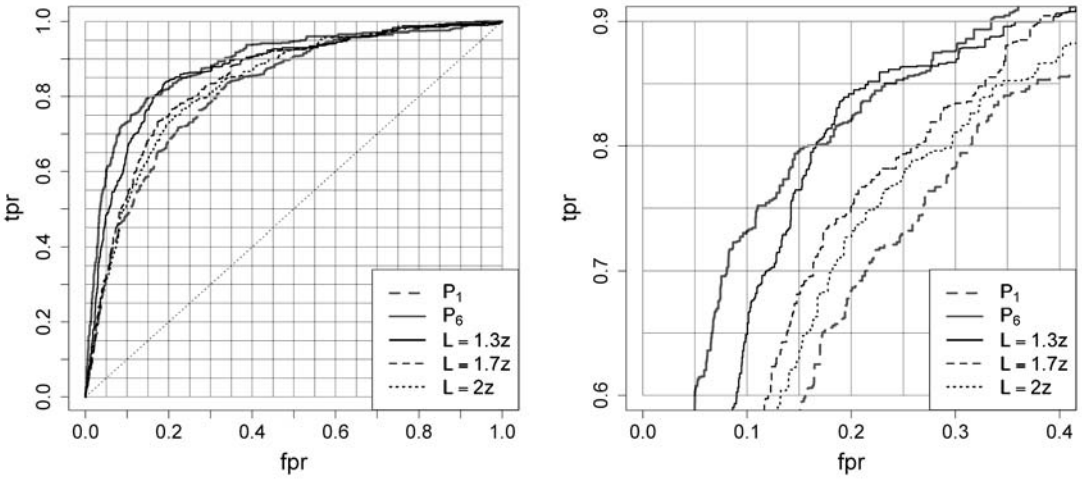
(a) Ausgewählte ROC-Kurven



(b) Umgebung $TPR = (0.6, 0.9)$ und $FPR = (0, 0.4)$

Abbildung 4.14: *Vergleich der Genauigkeit von Schätzern von M_{12}, P_1 und P_6 für 2006 [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Ausgewählte ROC-Kurven

(b) Umgebung $TPR = (0.6, 0.9)$ und $FPR = (0, 0.4)$

Abbildung 4.15: *Vergleich der Genauigkeit von Schätzern von M_{22}, P_1 und P_6 für 2006 [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Die Ergebnisse aus Abbildung 4.13 werden bestätigt. Die ROC-Kurven für Schätzer mit zwei Einkommensklassen und den drei gewählten Klassengrenzen unterscheiden sich in der Umgebung von $TPR = 80\%$, wie 4.14(b) und 4.15(b) zeigen, erheblich. Das Modell mit der Klassengrenze bei $L = 1.3z$ ist von den drei Modellen mit zwei Einkommensklassen am besten. Die Genauigkeit der Schätzungen liegt bei 80% TPR sogar leicht über P_6. Dagegen sind die Genauigkeiten von Schätzungen mit den Klassengrenzen $L = 1.7z$ und $L = 2z$ erheblich geringer, allerdings noch über denen von Modell P_1.

Im Folgenden werden drei Einkommensklassen untersucht, d.h. die Methoden M_{13} und M_{23} angewandt. Die Einkommensgrenzen werden wiederum als Vielfaches der Armutslinie gewählt ($L_1 = k_1 z$, $L_2 = k_2 z$). Da die Einkommensklassen in der Realität nicht so eng/genau abgegrenzt werden können, werden für die Abgrenzung von unterer und mittlerer Einkommensklasse „kleinere“ k_1 ($\in \{1, 1.1, 1.2, 1.3\}$) und für die Abgrenzung der mittleren und der oberen Einkommensklasse „größere“ k_2 ($\in \{2, 2.5\}$) untersucht. Zusätzlich wird mit $k_2 = 1.5$ noch eine engere Abgrenzung für die Grenze zwischen der mittleren und der oberen Einkommensklasse betrachtet. Die Regressionstabellen und die Histogramme der Quantilsresiduen sind im Anhang (A.3.2) aufgeführt. Wie zu erwarten, steigt die Armutswahrscheinlichkeit in der unteren Einkommensklasse stärker als in der mittleren Einkommensklasse

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

und auch stärker als bei der Betrachtung von zwei Einkommensklassen (k=1.1). Für die Logit-Koeffizienten der übrigen Klassen gab es nur wenige Veränderungen.

Für die drei getrennten Modelle ergeben sich große Unterschiede hinsichtlich der Logit-Koeffizienten. In der oberen Einkommensklasse ist zu beobachten, dass die Wahrscheinlichkeit zur armen Kategorie zu gehören, für Haushalte mit weiblichem HV höher ist als für solche mit männlichem HV im Gegensatz zu den beiden anderen Klassen. In der hohen Einkommensklasse haben Abitur und Realschulabschluss einen vermindernden Effekt auf die Armutswahrscheinlichkeit. Allerdings steigt die Wahrscheinlichkeit für Beschäftigung der HVs in Bank/Versicherung und Transport. Die Wahrscheinlichkeit fällt für verwitwete und getrennte Haushalte, für letztere sogar recht stark. Die Anzahl der Vollzeitbeschäftigten hat in den Regressionen in allen Gruppen keinen Einfluss, während die Anzahl der Teilzeitbeschäftigten in der oberen und der unteren Einkommensgruppe zu sinkender Wahrscheinlichkeit führt. Die Wohnungsgröße hat nur in der mittleren und hohen Einkommensklasse bei sehr kleinen Wohnungen einen Einfluss (Anstieg der Armutswahrscheinlichkeit).

Die Abbildungen 4.16 und 4.17 vergleichen die Genauigkeit von Schätzungen der Vulnerabilität mit zwei und drei Einkommensklassen, basierend auf den Methoden M_{1K} und M_{2K}, für verschiedene Kombinationen der Klassengrenzen. Es ist erneut festzustellen, dass die Methode M_{1K} genauere Schätzer erzeugt als die Methode M_{2K}. Wenn L bzw. L_1 als $1.2z$ oder $1.3z$ gewählt wird, erbringen drei Einkommensklassen keine Verbesserung im Vergleich zu zweien.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

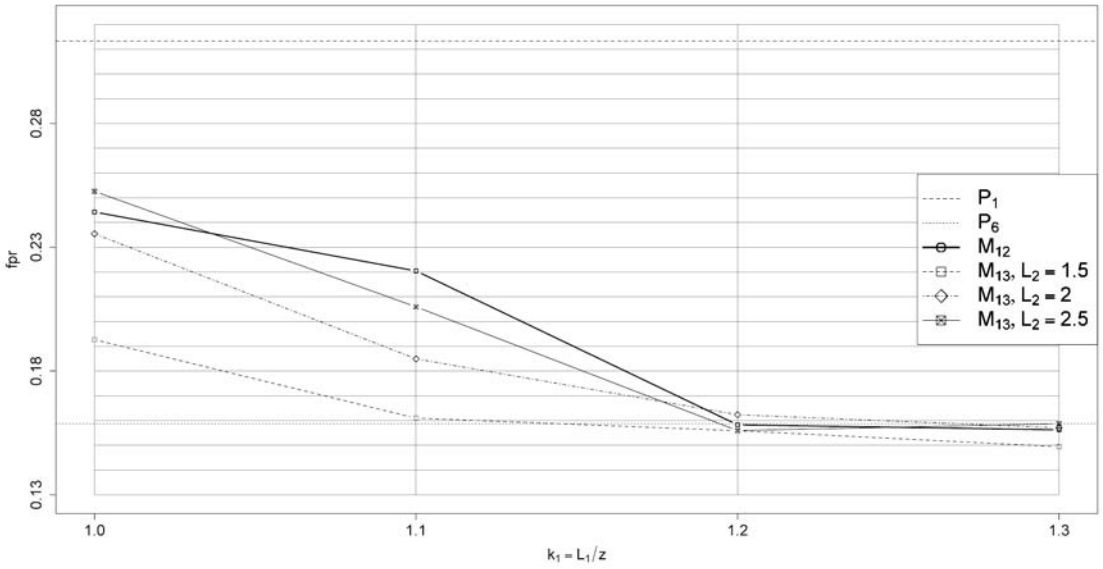


Abbildung 4.16: *Vergleich der Genauigkeit der Vulnerabilitäts-Schätzer mit zwei und drei Einkommensklassen (Methode M_{1K}) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

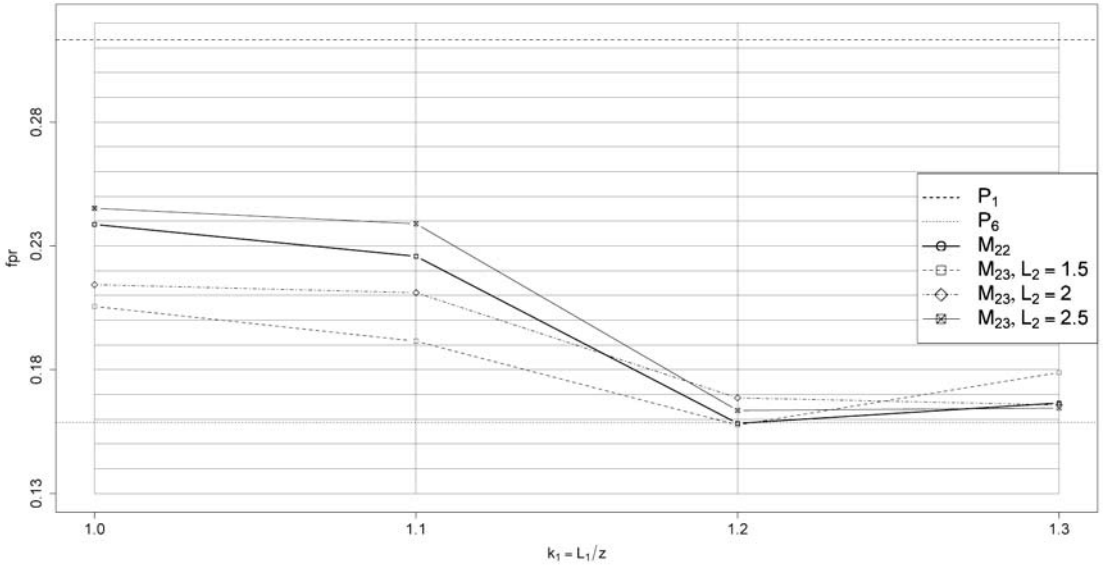


Abbildung 4.17: *Vergleich der Genauigkeit der Vulnerabilitäts-Schätzer mit zwei und drei Einkommensklassen (Methode M_{2K}) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Um die Genauigkeit der Schätzer von verschiedenen Modellen und Klassengrenzen bei zwei Einkommensklassen zu quantifizieren, werden Beispiele gegeben. In den Abbildungen 4.18 und 4.19 werden die Anzahl der falsch als armutsgefährdet klassifizierten Haushalte (false positives) bei $TPR = 80\%$ und drei verschiedenen Armutsquoten (10%, 20% und 40%) in einer hypothetischen Population von 1 000 Haushalten dargestellt. Auf der x-Achse sind die false positives gegen die drei Armutsquoten auf der y-Achse abgetragen. In der Abbildung ist zu beobachten, dass die false positives bei steigender Armutsquote für alle betrachteten Modelle linear fällt.

Die false positives (bei 80% TPR) der Schätzung basierend auf zwei Einkommensklassen unterscheiden sich für verschiedene Klassengrenzen L erheblich. Eine Klassengrenze von 11 700 € (= 1.3·9 000 €) führt zu vergleichbar genauen Schätzern wie die Schätzungen mit P_6. Um bei einer Armutsquote von 10% 80 von 100 Haushalten richtig als armutsgefährdet zu klassifizieren, müssen in dieser Variante des Modells M_{12} 140 der 900 nicht-armen Haushalte als armutsgefährdet toleriert werden, d.h. $140/(80 + 140) = 64\%$ der als armutsgefährdet klassifizierten Haushalte sind nicht arm geworden. Bei einer Armutsquote von 20% reduziert sich diese Quote auf 43% und bei 40% auf 22%. Wie bereits im vorherigen Abschnitt genannt, steigt dafür die Quote der nicht-identifizierten armen Haushalte. Für eine weniger gute Klassengrenze, z.B. 18 000 €, müssen 243 von 900 Haushalten falsch als armutsgefährdet klassifiziert werden, um 80 von 100 armen Haushalten richtig zu identifizieren, d.h. 75% der als armutsgefährdet klassifizierten Haushalte sind nicht arm geworden. Das sind etwa 35 false positives weniger als durch das Modell P_1. Bei steigender Armutsquote sinkt die Anzahl der nicht-armen Haushalte, so dass die Unterschiede kleiner werden. Die Modelle M_{12} und M_{22} unterscheiden sich bei gleicher Klassengrenze nicht gravierend. Modell M_{12} ist etwas besser.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

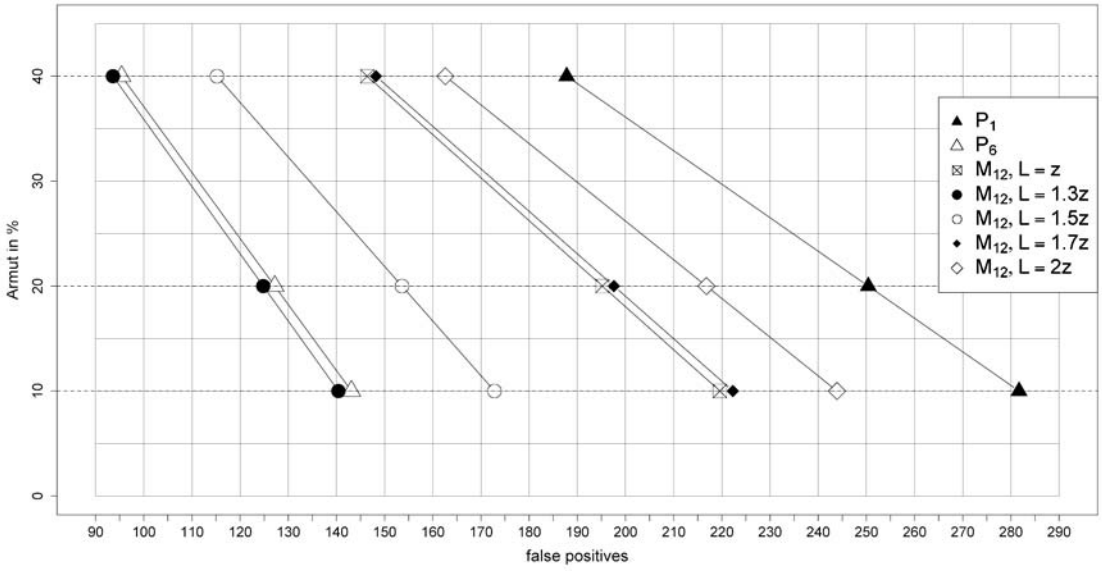


Abbildung 4.18: *False positives in einer Stichprobe von 1 000 Haushalten (Schätzung mit $M_{12,06}$) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

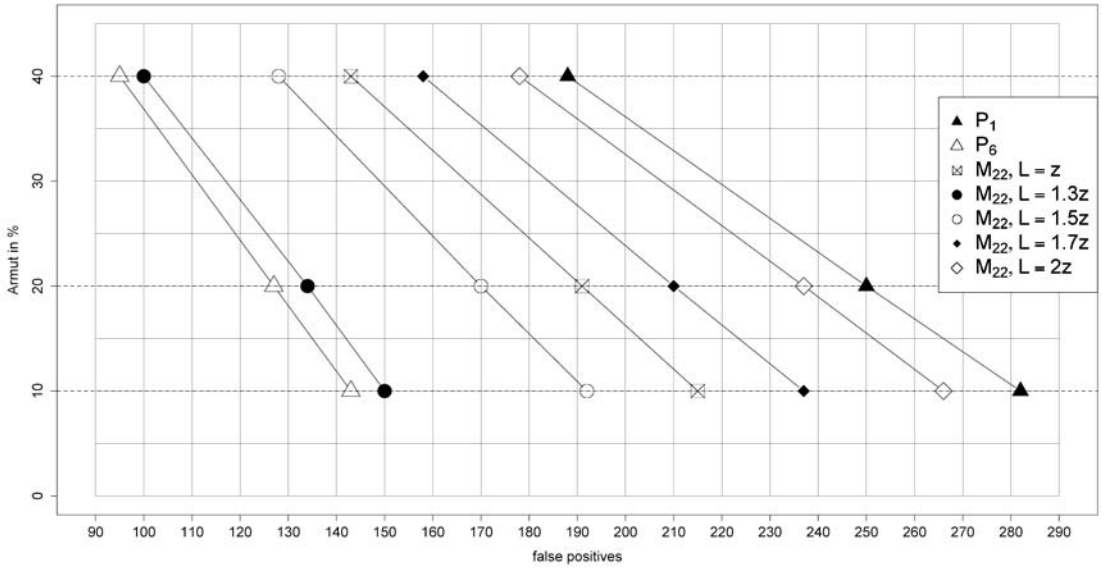


Abbildung 4.19: *False positives in einer Stichprobe von 1 000 Haushalten (Schätzung mit $M_{22,06}$) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Zusammengefasst lässt sich die Vulnerabilität für das Jahr 2006 mit dem Modell M_{12} und einer Klassengrenze von 11 700 € vergleichbar gut schätzen wie mit dem

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Modell P_6. Je weiter die Klassengrenze von 11 700 € entfernt gewählt wird, desto mehr verschlechtert sich die Genauigkeit der Vulnerabilitäts-Schätzung. Für alle betrachteten Klassengrenzen lag diese allerdings über den Schätzern, die auf P_1, d.h. Querschnittsdaten und Haushaltsvariablen, basieren.

4.2.2 Stabilität im Zeitraum 1994-2008

In diesem Abschnitt wird die Stabilität von Schätzern mit zwei und drei Einkommensklassen bei fixierter VPL untersucht. Da im letzten Abschnitt ermittelt wurde, dass $L = 1.3z$ im Falle von zwei sowie $L_1 = 1.3z$ und $L_2 = 2z$ im Falle von drei Einkommensklassen eine gute Wahl für Deutschland ist, wird die Stabilität für diese Klassengrenzen untersucht. Die Festlegung der VPLs erfolgt wieder aufgrund der historischen VPLs bei 80% TPR. Die VPLs sind hier, im Unterschied zu Abschnitt 4.1, Wahrscheinlichkeiten.

Abbildung 4.20 stellt die VPLs für die genannten Modelle dar. Für alle vier Modelle schwanken die historischen VPLs im Beobachtungszeitraum und weisen zusätzlich einen leichten Trend auf. Die Vulnerabilität wird mit einer VPL von 10% geschätzt. Die Genauigkeit der Schätzer sind in den Abbildungen 4.21 und 4.22 dargestellt.

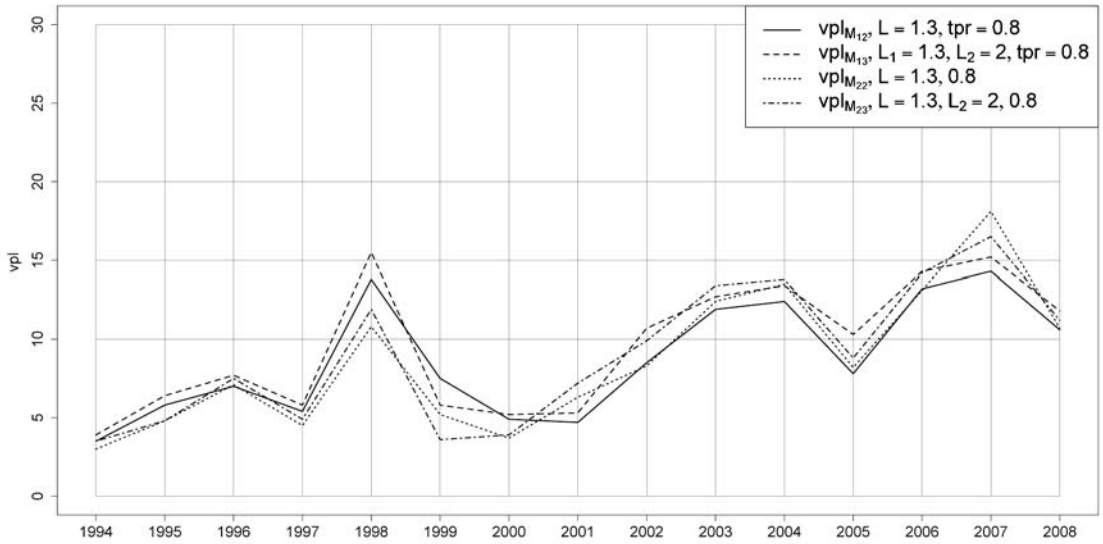
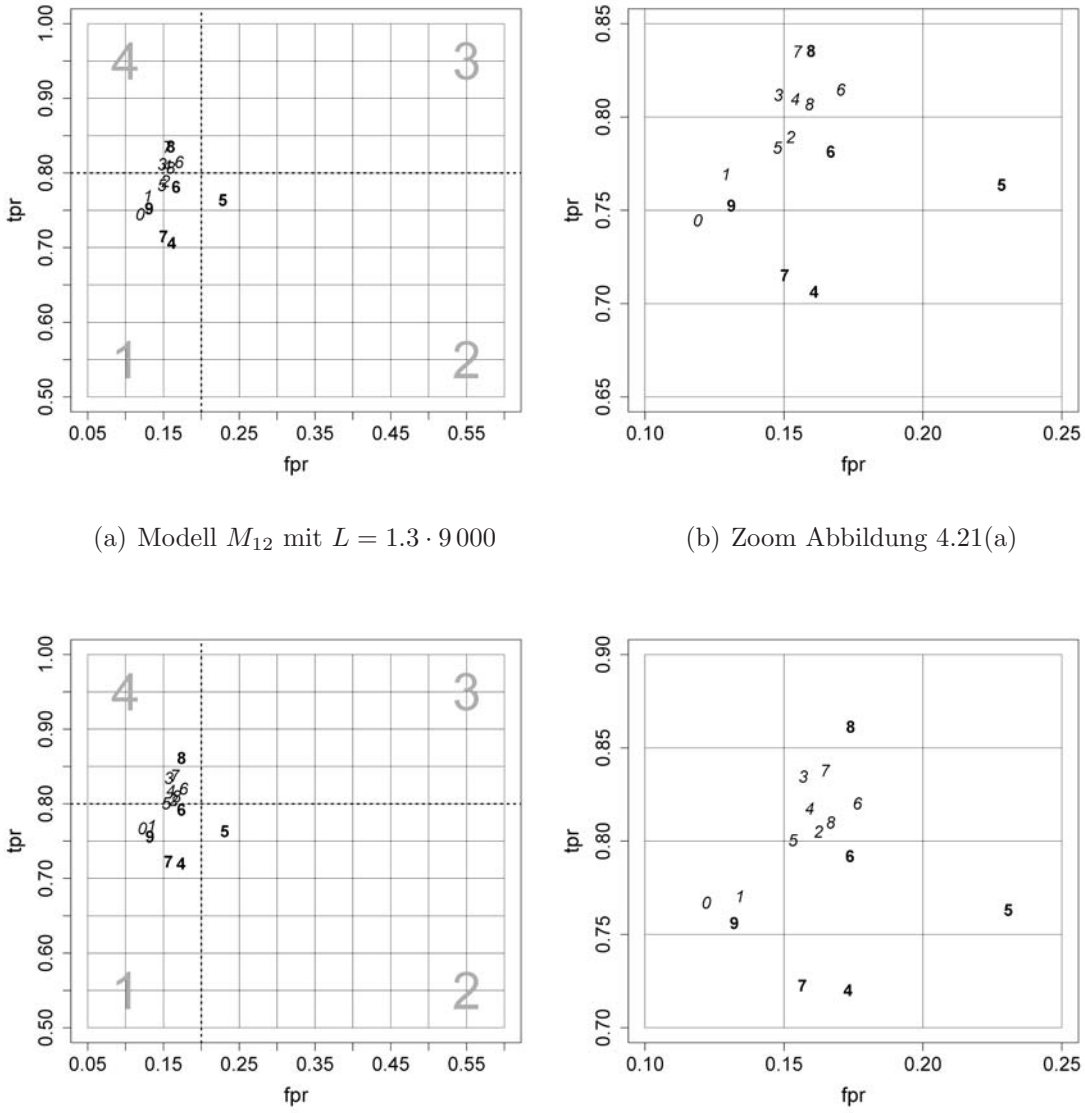


Abbildung 4.20: *VPL von drei Modellen bei 80% TPR (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Modell M_{12} mit $L = 1.3 \cdot 9\,000$

(b) Zoom Abbildung 4.21(a)

(c) Modell M_{13} mit $L_1 = 1.3 \cdot 9000$ und $L_2 = 2 \cdot 9\,000$

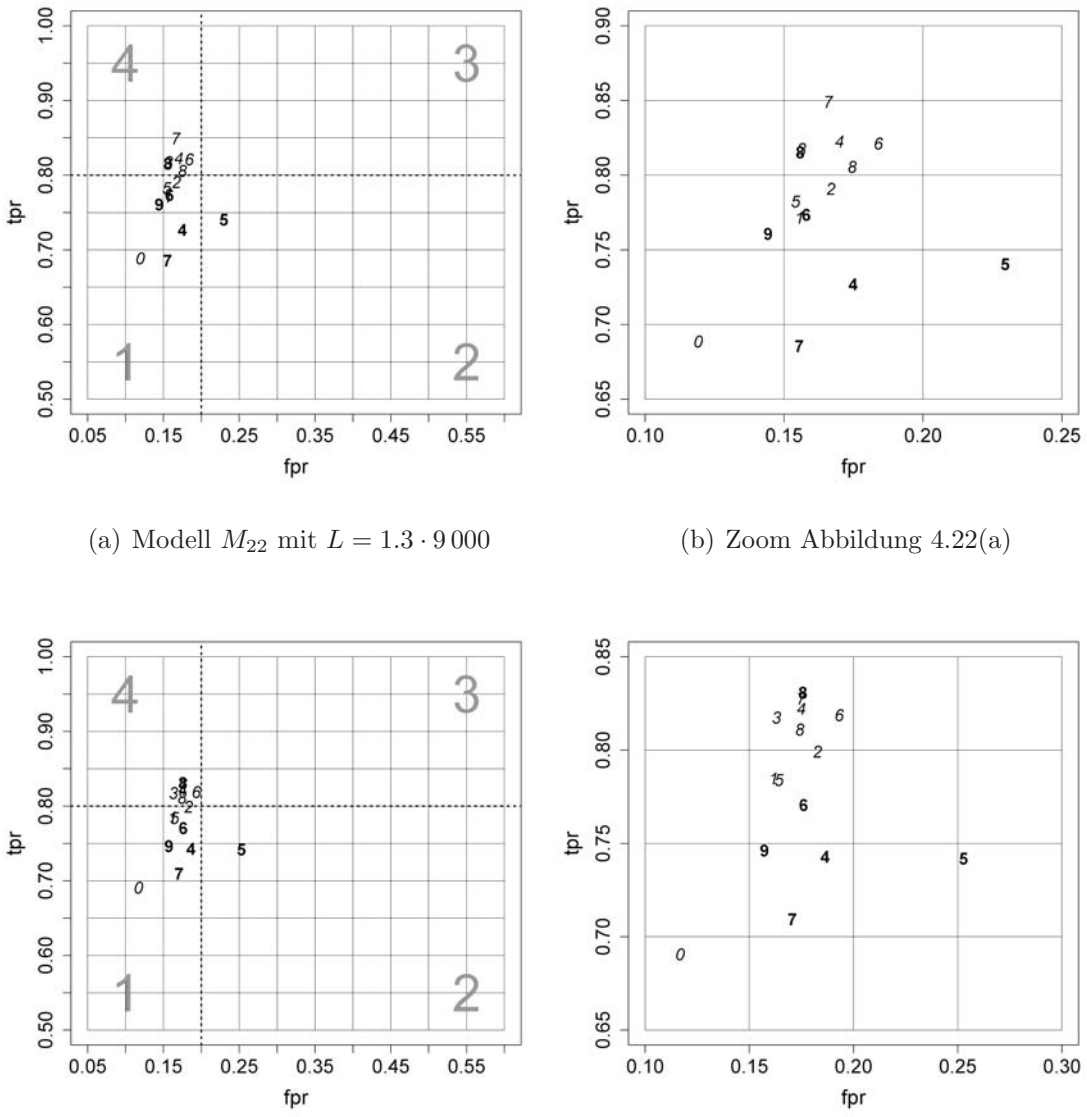
(d) Zoom Abbildung 4.21(c)

Abbildung 4.21: *Stabilität der Vulnerabilitäts-Schätzer von M_{1K} (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Die Abbildungen zeigen, dass alle Schätzer relativ stabil sind. Ausnahmen bilden, wie auch bei der Schätzung mit stetigem Einkommen, vier Jahre in den 90er Jahren und das Jahr 2000, die zu einer TPR unter 75% und/oder einer FPR über 20% führen. Insgesamt betrachtet erzeugt die Methode M_{1K} etwas genauere Schätzer als die Methode M_{2K}. Die Betrachtung von drei anstelle von zwei Einkommensklassen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

führt zu einer geringen Verbesserung der Genauigkeit der Schätzer. Die Genauigkeit ist in etwa vergleichbar mit Schätzern von P_6.



(a) Modell M_{22} mit $L = 1.3 \cdot 9\,000$

(b) Zoom Abbildung 4.22(a)

(c) Modell M_{23} mit $L_1 = 1.3 \cdot 9000$ und $L_2 = 2 \cdot 9\,000$

(d) Zoom Abbildung 4.22(c)

Abbildung 4.22: *Stabilität der Vulnerabilitäts-Schätzer von M_{2K} (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.2.3 Sensitivität

In diesem Abschnitt wird untersucht, wie sich die Genauigkeit der Vulnerabilitäts-Schätzer verändert, wenn Fehler bei der Zuordnung der Einkommen auf die Einkommensklassen gemacht werden (unter der Annahme, dass das im SOEP dokumentierte Einkommen unverzerrt ist).

Dazu werden aus den im SOEP gegebenen, stetigen Einkommen $y_{h,t}$ fehlerbehaftete Einkommen $y_{h,t,F}$ generiert, indem Fehler $m_{h,t}$ simuliert und additiv mit den tatsächlichen Einkommen verknüpft werden. Wie bereits erwähnt, wird der Datensatz bereits vor der Fehlerbehaftung auf die Haushalte mit einem Äquivalenzeinkommen unterhalb von 30 000 € beschränkt. Haushalte, die nach der Fehlerbehaftung ein negatives Einkommen haben, werden der unteren Einkommensklasse, solche mit einem Einkommen von größer oder gleich 30 000 € der oberen Einkommensklasse zugeordnet. Es wird angenommen, dass Fehler proportional vom Einkommen abhängen, d.h., dass hohe Einkommen stärker streuen als niedrige. Der Fehler ist folglich eine multiplikative Verknüpfung aus zwei Komponenten, dem Einkommen $y_{h,t}$ und einer Zufallszahl $r_{h,t}$, die den Fehleranteil des Einkommens bestimmt, siehe Gleichung (4.1). Die Streuungsbreite des Fehlers wird durch den Streuungsparameter ς festgelegt, der im Folgenden noch näher spezifiziert wird.

$$y_{h,t,F} = y_{h,t} \cdot (1 + r_{h,t}) \tag{4.1}$$

Untersucht werden verschiedene Varianten, die sich in der Größe der Fehler, ihrer Verteilung sowie ihrer Abhängigkeitsstruktur in aufeinanderfolgenden Wellen unterscheiden. Es wird unterschieden zwischen

- kleinen, mittleren und großen Fehlern (wird im Folgendem noch spezifiziert),
- Fehler aus einer Gleich- und einer Normalverteilung,
- unabhängige und perfekt korrelierte, d.h. identische Fehler von Welle zu Welle.

Die Höhe des *tatsächlichen* Einkommens sowie die Fehlervarianz determinieren, ob die fehlerbehafteten Einkommen einer falschen Einkommensklasse zugeordnet werden. Die Schätzung der Vulnerabilität eines Haushalts kann sich aufgrund des Fehlers wie folgt verändern:

1. Der Fehler führt nicht zu einer falschen Zuordnung der Haushalte, so dass sich die geschätzte Armutswahrscheinlichkeit des Haushalts und somit auch die Vulnerabilität nicht verändert.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2. Aufgrund des Fehlers wird ein Haushalt, der *richtig* der unteren Einkommensklasse zuzuordnen wäre, der oberen Einkommensklasse zugeordnet. Das führt möglicherweise dazu, dass für den Haushalt eine niedrigere Armutswahrscheinlichkeit geschätzt wird, so dass dieser nun nicht (mehr) als armutsgefährdet klassifiziert wird. In diesem Fall würde der Haushalt nicht als arm identifiziert. Falls der Haushalt arm geworden ist, so müsste, um 80% TPR zu erreichen, die VPL niedriger gewählt werden.
3. Der Fehler führt dazu, dass Haushalte aus der oberen in die untere Einkommensklasse eingeordnet werden und für diesen Haushalt eine höhere Armutswahrscheinlichkeit geschätzt wird. Aufgrund dieser wird er möglicherweise, im Gegensatz zu vorher, als armutsgefährdet klassifiziert. Wird der Haushalt nicht arm, so steigt die FPR.

Nachdem die Haushalte aufgrund der simulierten Einkommen $y_{h,t,F}$ den Einkommensklassen zugeordnet wurden, wird die Vulnerabilität unter Verwendung der *verzerrten* Einkommensklassen geschätzt. Die Information, ob ein Haushalt arm ist (sowohl als Response bei der Regression als auch beim Vergleich zur Konstruktion der ROC), beruhen auf dem *unverzerrten Einkommen*. Das Vorgehen wird sehr häufig wiederholt (hier 100 Mal), und es wird die mittlere Performance der Realisationen betrachtet. Um die Auswirkungen der Fehler besser beurteilen zu können, werden für jeden simulierten Fehler dieselben zufälligen Anteile $r_{h,t}$ bei den Realisationen gewählt, d.h.

$$r_{1,t,\varsigma=0.1} = r_{1,t,\varsigma=0.2} = r_{1,\varsigma=0.4}$$

$$r_{2,\varsigma=0.1} = r_{2,t,\varsigma=0.2} = r_{2,t,\varsigma=0.4}$$

$$\dots$$

$$r_{100,t,\varsigma=0.1} = r_{100,t,\varsigma=0.2} = r_{100,t,\varsigma=0.4}$$

Für abhängige Fehler sind die zufälligen Anteile $r_{h,t}$ für jede Welle des 4-Jahres-Panels gleich, während sie für unabhängige Fehler unterschiedlich (und ohne Korrelation zueinander) sind.

In der ersten Variante wird eine Gleichverteilung der $y_{h,t,F}$ angenommen, d.h.

$$y_{h,t,F} \sim y_{h,t} \cdot \mathcal{U}(1 - \varsigma_U, 1 + \varsigma_U)$$

wobei ς_U den Streuungsparameter der Gleichverteilung darstellt. Abbildung 4.23 veranschaulicht die Streuung gleichverteilter Einkommen von 1 000 € und 2 000 € bei einem Streuungsparameter von $\varsigma_U = 0.2$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Abbildung 4.23: *Verteilung des generierten Einkommens für $y_{h,t}$=1 000 € und $y_{h,t}$=2 000 € bei Annahme einer Gleichverteilung.*

In der zweiten Variante wird eine Normalverteilung unterstellt. In dieser Arbeit wird angenommen, dass die Einkommen im 95% Konfidenzintervall um das „wahre“ Einkommen streuen, siehe Abbildung 4.24.

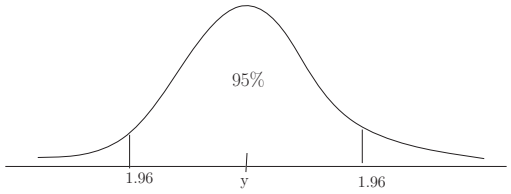


Abbildung 4.24: *Dichte des generierten Einkommens für $y_{h,t}$ bei Annahme einer Normalverteilung.*

Damit lässt sich die Beziehung zwischen dem Streuungsparameter ς_N und der Standardabweichung wie in (4.2) beschreiben.

$$1.96 \cdot \sigma_N + y_{h,t} = y_{h,t} \cdot (1 + \varsigma_N) \tag{4.2}$$

Die Streuungsparameter ς_U und ς_N werden über die Varianzen der Einkommen der beiden Verteilungen näher spezifiziert (siehe Gleichungen (4.3) und (4.4)). Gleichverteilte Einkommen haben auf einem festem Intervall eine größere Varianz als normalverteilte Einkommen auf diesem Intervall. Um die Genauigkeit der beiden Varianten vergleichen zu können, muss sie bei gleichen Varianzen betrachtet werden. In dieser Arbeit wird der Streuungsparameter für die Gleichverteilung spezifiziert (Notation: ς_U), der Streuungsparameter für die Normalverteilung (ς_N) ist dann eine Funktion von ς_U, siehe Gleichung (4.5).

$$\sigma_U^2(y_{h,t,F}) = \frac{((1 + \varsigma_U) \cdot y_{h,t} - (1 - \varsigma_U) \cdot y_{h,t})^2}{12} = \frac{\varsigma_U^2 \cdot (y_{h,t})^2}{3} \tag{4.3}$$

$$\sigma_N^2(y_{h,t,F}) = \left(y_{h,t} \cdot \frac{\varsigma_N}{1.96}\right)^2 \tag{4.4}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$$y_{h,t,F} \sim \mathcal{N}\left(y_{h,t}, \left(y_{h,t} \cdot \frac{\varsigma_N}{1.96}\right)^2\right)$$

bzw.

$$y_{h,t,F} = y_{h,t} \cdot \left(1 + \frac{\varsigma_N}{1.96} \cdot r_{h,t}\right) \quad mit\ r_{h,t} \sim \mathcal{N}(0,1)$$

$$\varsigma_N = \frac{1}{\sqrt{3}} \cdot \varsigma_U \cdot 1.96 \tag{4.5}$$

In dieser Arbeit wird ein kleiner Fehler mit $\varsigma_U = 0.1$, ein mittlerer Fehler mit $\varsigma_U = 0.2$ und ein großer Fehler mit $\varsigma_U = 0.4$ quantifiziert. Mit (4.5) ergibt sich dann $\varsigma_N \in \{0.11, 0.23, 0.45\}$. Nach dieser Festlegung ergibt sich für kleine Fehler eine Standardabweichung von 6%, für mittlere von 12% und für große von 23% des Einkommens.

Im Folgenden wird die Vulnerabilität ohne und mit den beschriebenen Arten von Fehlern mit M_{12} und $L = 1.3z$ für das Jahr 2001 geschätzt. Tabelle 4.17 zeigt die Einkommensverteilung der Haushalte auf die obere und untere Einkommensklassen für das Jahr 2000 bei einem großen Fehler (d.h. $\varsigma_U = 0.4$ und $\varsigma_N = 0.45$).

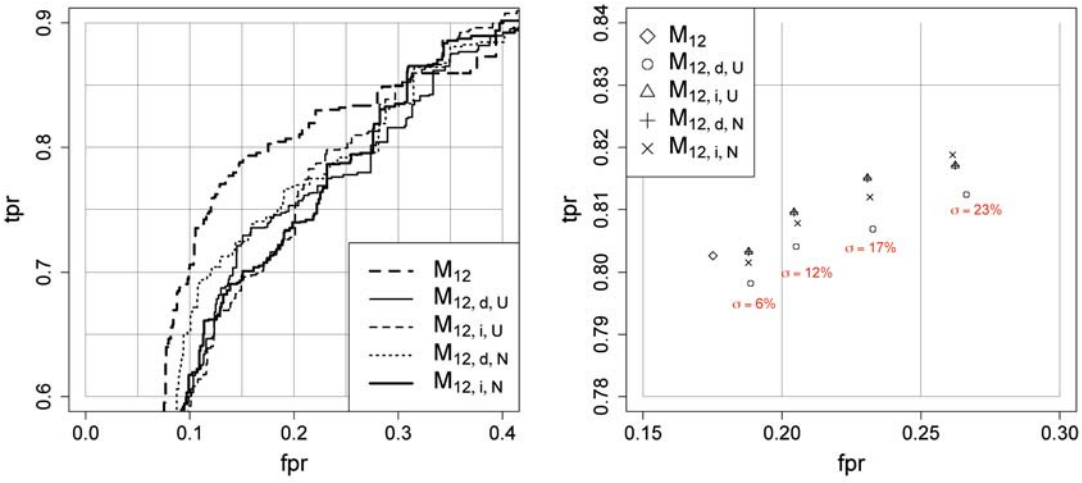
Fehler	Anzahl oben	Anzahl unten
keiner	3611	752
gleichverteilt, korreliert	3354	1009
gleichverteilt, unkorreliert	3313	1050
normalverteilt, korreliert	3351	1012
normalverteilt, unkorreliert	3388	975

Tabelle 4.17: *Einkommensverteilung der Haushalte in 2000 (Realisation mit Streuungsparameter $\varsigma_U = 0.4$ bzw. $\varsigma_N = 0.45$) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Die Ergebnisse zeigen, dass die obere Einkommensklasse der tatsächlichen Einkommensverteilung etwa 250 Haushalte mehr beinhaltet als die obere Einkommensklasse der fehlerbehafteten Einkommensverteilung. Die Verteilung der Haushalte bei den verschiedenen Arten unterscheiden sich. Die Verteilung mit gleichverteiltem, korreliertem Fehler und die mit normalverteiltem, korreliertem Fehler sind relativ ähnlich. Abbildung 4.25(a) zeigt die ROCs für die dargestellten Situationen. In der Legende wird die Verteilung des Fehlers durch den zweiten Teil im Index (U für die Uniformverteilung und N für Normalverteilung) angegeben. Die

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Abhängigkeitsstruktur der Fehler wird durch den dritten Teil im Index (d für dependent und i für independent) angegeben. Wird die VPL so gewählt, dass 80% der Vulnerabilitäts-Schätzung mit fehlerbehaftetem Einkommen richtig als armutsgefährdet klassifiziert werden, so führt die Fehlerverteilung mit gleichverteiltem und korreliertem Fehler zur geringsten Verschlechterung. Interessanter ist jedoch, wie sich die Genauigkeit unter einem Fehler verändert, wenn die VPL so gewählt wird, dass 80% der Vulnerabilitäts-Schätzer bei richtiger Klassenzuordnung als armutsgefährdet klassifiziert werden. Diese Darstellung wird in Abbildung 4.25(b) für kleine, mittlere und große Fehler gezeigt.



(a) ROC einer Realisation

(b) (FPR,TPR) für verschiedene Fehler

Abbildung 4.25: *Auswirkung eines Fehlers auf die Genauigkeit der Vulnerabilitäts-Schätzung in 2001 [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Die Standardabweichung in Prozent des Einkommens wird mit σ bezeichnet.*

Wie vermutet, verschlechtert sich die Genauigkeit, je größer der Fehler ist. Jedoch führt im Jahr 2001 weder ein kleiner noch ein mittlerer Fehler erheblich zu einer Verschlechterung der Genauigkeit. Bei großem Fehler steigt die FPR von etwa 17% ohne Fehler auf etwa 27%. Auch die TPR steigt von 80% auf 82% etwas an. Die Abhängigkeitsstruktur der Fehler führt nur zu geringen Unterschieden in der Genauigkeit. Auch die Verteilung der Fehler scheint für die Genauigkeit nur eine untergeordnete Rolle zu spielen. Die Vulnerabilitäts-Schätzung mit gleichverteiltem, korreliertem und normalverteiltem, korreliertem Fehler sind relativ ähnlich. Für die Schätzer von Modell P_1 ergab sich eine Genauigkeit von 40% bei 80% TPR.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Also verbessern Vulnerabilitäts-Schätzer basierend auf fehlerbehaftetem Einkommen solche basierend nur auf Haushaltskovariablen.

4.2.4 Zwischenfazit

In diesem Abschnitt wurde mit dem SOEP und den Modellen aus Abschnitt 2.4.2 untersucht, wie stark sich die Genauigkeit der Vulnerabilitäts-Schätzer verändert, wenn das Einkommen in Einkommensklassen anstelle von stetigem, unverzerrten Einkommen zur Schätzung verwendet wird. Zum Vergleich der Ergebnisse wurden Schätzer basierend auf Querschnittsdaten und Haushaltskovariablen sowie Paneldaten und stetigem Einkommen hinzugezogen.

Die Ergebnisse der Analysen ergaben, dass die Genauigkeit der Schätzer abhängig von der Definition der Einkommensklassen ist. Selbst wenn nur zwei Einkommensklassen verwendet werden, können für eine *günstige* Klassengrenze L, für die SOEP Daten bei 11 700 € = 1.3· 9 000 €, vergleichbar gute Ergebnisse erzielt werden wie mit stetigem Einkommen. Wird zusätzlich eine dritte Einkommensklasse mit Klassengrenze zwischen der zweiten und der dritten Klasse bei 18 000 € = 2· 9 000 € eingeführt, so verbessert sich die Genauigkeit noch ein wenig. Die Schätzer sind relativ stabil über die Zeit und weisen, ähnlich wie die Schätzer basierend auf stetigem Einkommen, nur in den 90er Jahren und im Jahr 2000 Ausreißer auf. Werden die Klassengrenzen *weniger günstig*, z.B. im Falle von zwei Klassen bei 18 000 € gewählt, so nimmt die Genauigkeit deutlich ab. Für alle betrachteten Klassengrenzen lag diese jedoch über denen von Schätzern, die nur auf Haushaltskovariablen basieren.

Bei der Analyse, welches Ausmaß Fehler verschiedener Art (Größe, Verteilung, Abhängigkeitsstruktur von Welle zu Welle) bei der Einordnung der Haushalte in die Einkommensklassen auf die Genauigkeit haben, zeigte sich nur bei großen Fehlern eine spürbare Verschlechterung der Ergebnisse. Die Genauigkeit übertraf jedoch weiterhin die Genauigkeit von Schätzern, die nur auf Haushaltsvariablen basierten.

Die Analyse macht deutlich, dass die Methoden zu annehmbaren Ergebnissen führen, selbst wenn das Einkommen nicht immer den korrekten Einkommensklassen zugeteilt ist. Es empfiehlt sich daher, auch eine *grobe* Einteilung der Haushalte auf die Einkommensklassen für die Schätzung der Vulnerabilität zu verwenden.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.3 Analyse von Verbesserungsmöglichkeiten zur Schätzung der Vulnerabilität für Deutschland

In diesem Abschnitt wird mit den Methoden aus Abschnitt 2.4.3 untersucht, ob sich die Schätzer der Vulnerabilität aus Abschnitt 4.1 konkret für Deutschland verbessern lassen. Dazu ist es zunächst notwendig, die Entwicklung und die Ursachen der Armut in Deutschland genauer zu betrachten. Laut OECD-Studien (2008 und 2011) haben Ungleichheit und Armut in Deutschland seit dem Jahr 2000 deutlich zugenommen. Die „Schere zwischen arm und reich“ ist weiter auseinandergegangen. Als Ursachen werden in den Berichten unter anderem der strukturelle Wandel des Arbeitsmarktes als auch die Veränderung der Haushaltsstruktur genannt. Diese strukturellen Veränderungen führen auch zu Veränderungen in der Vulnerabilität der Haushalte.

Der Abschnitt 4.3.1 widmet sich der Möglichkeit der Erklärung und Prognose von Armut durch volkswirtschaftliche Veränderungen, die durch volkswirtschaftliche Indikatoren (Makrovariablen) für relevante Bereiche (Arbeitsmarkt, Steuern, Wirtschaft und Finanzmarkt) abgebildet werden. Es wird ex post für den Beobachtungszeitraum diskutiert, welche der makroökonomischen Veränderungen Auswirkungen auf die Entwicklung der Armutsquoten gehabt haben könnten. Darauf aufbauend wird untersucht, ob die makroökonomischen Indikatoren sich als Prognosevariablen für die Armutsquoten eignen. Diese liefern zwar nur die Aussage, wie viele Haushalte und nicht welche Haushalte zukünftig arm werden, dennoch könnten diese Informationen die Prävention von Armut unterstützen, z.B. indem anhand von gruppenspezifischen Indikatoren (z.B. die Frauenarbeitslosigkeit oder die Lohnentwicklung in Ostdeutschland) die Entwicklung der Armut von betroffenen Gruppen vorhergesagt wird.

Im Abschnitt 4.3.2 wird untersucht, inwieweit Veränderungen der Haushaltsstruktur zur Erklärung der Entwicklung der Armutsquote beitragen können. Dazu wird die Stichprobe nach Merkmalen zerlegt, z.B. Geschlecht mit den Ausprägungen männlich und weiblich, und für jede Ausprägung wird die Armutsquote über die Zeit (sowohl in der Gruppe, d.h. die bedingte Armutsquote als auch die Armutsquote der Gruppe gemessen an der Gesamtpopulation) dargestellt. In dieser Arbeit wird die Armutslinie für alle Gruppen bei 9 000 € gewählt. Außerdem wird jede Gruppierung nach Merkmalen getrennt betrachtet, d.h., es werden keine Interaktionen untersucht.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Anhand der Aspekte aus den Abschnitten 4.3.1 und 4.3.2 wird in den Abschnitten 4.3.3 und 4.3.4 versucht, die Schätzer der Vulnerabilität aus Abschnitt 4.1 zu verbessern. In den Abschnitten 4.1 und 4.2 war zu beobachten, dass die geschätzte Vulnerabilität über die Zeit unterschiedlich stark schwankte. Das ist darauf zurückzuführen, dass die historischen VPLs, die zu einer TPR von 80% führen, instabil über die Zeit waren und sich die Vulnerabilität aus diesem Grund nicht konstant mit 80% TPR durch eine absolute VPL schätzen ließ. Die Performance der Vulnerabilität als Funktion $FPR(TPR)$ könnte verbessert werden, wenn es gelingt, die VPL an geänderte Einflüsse anzupassen. In dieser Arbeit wird angenommen, dass Veränderungen in der Vulnerabilität, die nicht durch die betrachteten Kovariablen erfasst werden, die Ursache von volkswirtschaftlichen Veränderungen bzw. Schocks sind.

Um zu ermitteln, inwieweit Makrovariablen zur Schätzung der Vulnerabilität beitragen, werden die gepoolten Einkommen aller Haushalte und aller Jahre im Beobachtungszeitraum in Abhängigkeit von den volkswirtschaftlichen Indikatoren betrachtet. Besonders ist in diesem Zusammenhang von Interesse, wie stark die Einkommensniveaus in Nähe der Armutslinie bei 9 000 € und in der Nähe der VPL auf Veränderungen der Indikatoren reagieren. Dazu werden Quantilsregressionen durchgeführt und die Perzentile 7 und 8 (in etwa von Armut betroffene Haushalte), 22 und 24 (etwa in Höhe der VPL) sowie der Median betrachtet.

In Abschnitt 4.3.4 wird untersucht, ob strukturelle Veränderungen bei der Messung der Vulnerabilität besser erfasst werden können, wenn das Einkommen und/oder die Vulnerabilität für Merkmalsausprägungen von Haushalten getrennt geschätzt werden. Für diese Analysen werden wieder 4-Jahres-Panels verwendet, wobei jedoch die Haushalte im Datensatz *nicht* durch das Einkommen (d.h. durch 30 000 €) beschränkt werden. Eine Beschränkung würde dazu führen, dass ohnehin kleine Gruppen zu klein werden, um eine verlässliche Aussage über die Genauigkeit der Vulnerabilität zu machen. Die Schätzung der Vulnerabilität beruht in diesem Abschnitt, wie in Abschnitt 4.1, auf der Vorhersage des Einkommens. Für die Schätzungen werden die Modelle $P_{1,G}$ und $P_{6,G}$ betrachtet. Zur Schätzung der Vulnerabilität in Gruppen werden zwei Methoden untersucht.

In der ersten Methode werden die aktuellen Einkommen für Gruppen von Merkmalen *getrennt* (z.B. Haushalte mit weiblichem HV einerseits und Haushalte mit männlichem HV andererseits) geschätzt, während die Festlegung einer VPL (die zu 80% TPR aller Haushalte führt) auf den geschätzten Einkommen *aller Haushalte in allen Gruppen* basiert. Wird das Einkommen mit Querschnittsdaten in Gruppen ($P_{1,G}$) geschätzt, wird die Stichprobe nach Merkmalen (z.B. Geschlecht

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

des HVs) in der zweiten Welle in Teilstichproben aufgespalten, und es werden für die Teilmodelle (z.B. $P_{1,weiblich}$ und $P_{1,\text{männlich}}$) die Regressionskoeffizienten getrennt geschätzt und das Einkommen für die dritte Welle aufgrund der angepassten Werte in der zweiten Welle vorhergesagt. Werden Paneldaten (Modell $P_{6,G}$) verwendet, so wird die Stichprobe nach Merkmalen in der ersten Welle getrennt, und es werden für die Teilmodelle die Regressionskoeffizienten separat geschätzt. Mit diesen, sowie den Kovariablen aus der zweiten Welle, werden für Teildatensätze, aufgespalten nach Merkmalen in der zweiten Welle, die Einkommen für die Gruppen in der dritten Welle geschätzt. Die Teilstichproben werden nach der Schätzung wieder zusammengeführt und die ROC wird auf Basis aller Haushalte konstruiert. In diesem Fall wird angenommen, dass die Haushaltseigenschaften und die Einkommensverläufe in den Gruppen homogener sind und damit zu besseren Schätzungen führen, als wenn die Einkommen aller Haushalte gemeinsam geschätzt werden. Das Vorgehen beruht auf der Annahme, dass die unabhängigen Variablen das Einkommen unterschiedlich erklären und die Regressionsgeraden von verschiedenen Gruppen *nicht parallel* zueinander verlaufen. Der Ansatz stellt eine Gesamtbetrachtung dar, der direkt die Frage beantwortet, inwiefern die Schätzung des Einkommens in Gruppen zu einer Verbesserung in der Messung der Vulnerabilität führt.

Der zweite Ansatz dagegen schätzt sowohl das Einkommen als auch die Vulnerabilität für die Gruppen getrennt, d.h. bei 80% TPR von den geschätzten Einkommen jeder Gruppe eines Merkmals. Das bedeutet, dass in Gruppen, in denen absolut gesehen viele Haushalte arm sind, mehr Haushalte richtig als armutsgefährdet identifiziert werden müssen als in Gruppen mit weniger armen Haushalten, unabhängig von der Größe der Gruppe. Die Vorgehensweise beruht auf der Annahme, dass sich die Einkommen in getrennten Modellen besser schätzen lassen und auch die Verteilung der geschätzten Einkommen und damit die Vulnerabilität zwischen den Gruppen unterschiedlich ist. Die Methode liefert das Ergebnis, wie gut die Vulnerabilität in den Gruppen gemessen wird und ggfs. für welche Gruppen die Schätzung verbessert werden sollte.

Abschnitt 4.3.5 fasst die in diesem Abschnitt gewonnenen Erkenntnisse zusammen.

4.3.1 Makroökonomische Ursachen von Armut

In diesem Abschnitt wird diskutiert, welche makroökonomischen Entwicklungen Ursachen für die Entwicklung der Armutsquote im Zeitraum von 1992-2008 gewesen sein könnten. Dazu werden wirtschaftspolitische Veränderungen (unter dem Gesichtspunkt des Verlaufs der Armutsquote) in den Bereichen Arbeitsmarkt, Steu-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ern sowie Wirtschaft und Finanzmarkt *getrennt* und im Detail betrachtet. Auf die Betrachtung weiterer Bereiche der Sozialpolitik, wie etwa die Familien-, Renten- oder Gesundheitspolitik wird an dieser Stelle verzichtet.

4.3.1.1 Arbeitsmarkt

Einen zentralen Indikator des Arbeitsmarktes stellt die Arbeitslosenquote dar, die, gemessen an den abhängigen zivilen Erwerbstätigen (sozialversicherungspflichtige und geringfügig Beschäftigte, Beamte, Arbeitslose), in Abbildung 4.26 sowohl für Deutschland insgesamt als auch für West- und Ostdeutschland dargestellt ist (Bundesagentur für Arbeit, 2011). Zusätzlich sind die Regierungsperioden sowie einige bedeutende arbeitsmarktpolitische Veränderungen in der Abbildung dokumentiert. Für weitere Details sowie mögliche Auswirkungen von Ereignissen und Reformen auf die Arbeitslosenquote siehe Wewer (1998, S. 145-181), Egle et al. (2003, S. 215-238), Egle und Zohlnhöfer (2007, S. 271-294) sowie Egle und Zohlnhöfer (2010, S. 279-301).

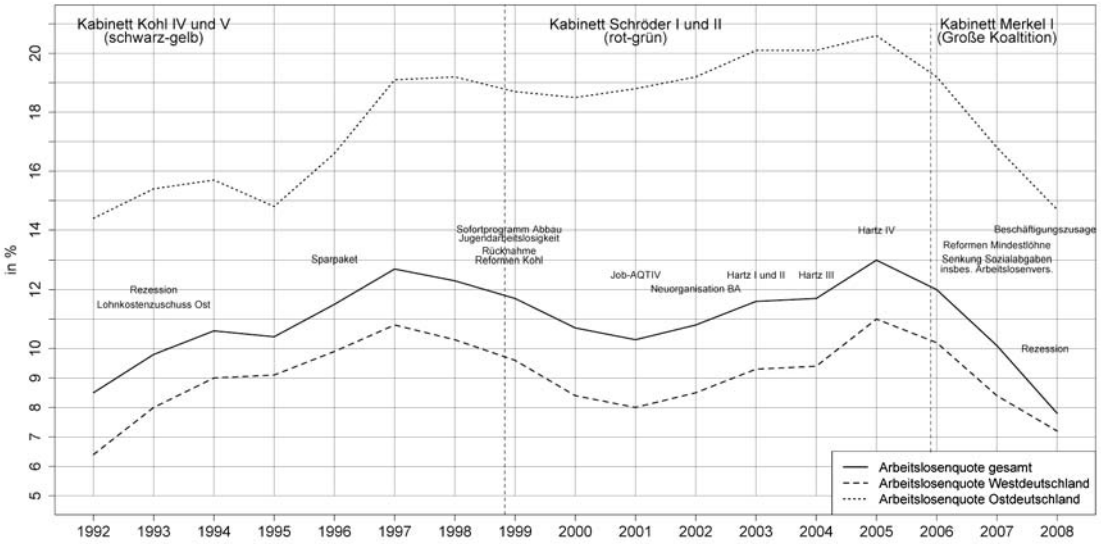


Abbildung 4.26: *Arbeitslosenquote (gemessen an den abhängigen zivilen Erwerbspersonen) (1992-2008) [Darstellung wie in Bundesagentur für Arbeit (2011)].*

Die Arbeitslosenquote für Deutschland ist, mit Ausnahme der ersten und der letzten beiden Jahre, im Beobachtungszeitraum mit 10%-13% relativ hoch. Der Verlauf der Arbeitslosenquote ist grob durch zwei Zyklen gekennzeichnet: einem Anstieg

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

von 1992-1997, einem leichten Rückgang bis 2001, einem starken Anstieg bis 2005 (Höhepunkt der Arbeitslosenquote im Beobachtungszeitraum) und einem rapiden Abfall am Ende des Beobachtungszeitraums.

Werden die Zyklen von Arbeitslosen- und Armutsquoten verglichen, so ist festzustellen, dass die Höhe der Arbeitslosenquote nur wenig zur Erklärung der Armutsquote beizutragen scheint. Lediglich in den Jahren 1997-1999 (fallend) und von 2001-2005 (steigend) sind die Größen gleichgerichtet. Ein konträrer Verlauf zeigt sich insbesondere zwischen 2005 und 2008. In diesen Jahren ist die Arbeitslosenquote trotz der Finanzkrise rapide auf das niedrigste Niveau im Beobachtungszeitraum gefallen, während die Armutsquote auf dem hohen Niveau von 2005 stagniert. Eine solche Entwicklung signalisiert, dass bei einem Teil der Bevölkerung das Einkommen aus der Erwerbstätigkeit nicht ausreicht (z.B. Zeitarbeit, Niedriglohn), um die Grundbedürfnisse zu decken. Dieses Phänomen wird in der Literatur mit den Begriffen „Erwerbsarmut“ und “working poor“ bezeichnet (z.B. IAB (2009)).

Zu berücksichtigen ist bei dieser Betrachtung allerdings, dass die Armutsquote auf der Haushaltsebene, die Arbeitslosenquote jedoch auf der Personenebene vorliegt. Die Arbeitslosenquote berücksichtigt also weder, wie viele Personen in einem Haushalt arm sind, noch erfasst sie die Länge der Arbeitslosigkeit. Dies sind aber wichtige Kriterien, die bestimmen, ob Haushalte, gemessen an der Höhe ihres Jahreseinkommens, arm werden.

Mit den Daten des SOEPs existieren bereits Studien, die versuchen, die Armut durch die Arbeitslosigkeit zu erklären (z.B. Frick et al. (2005), Frick und Grabka (2005), Biewen und Juhasz (2011)).

Frick et al. (2005) betrachten die Entwicklung von Armut und Konjunkturindikatoren, darunter auch die Arbeitslosenquote von 1985-2003 für Westdeutschland. Sie kommen zu dem Ergebnis, dass die Ungleichheit (gemessen durch den Gini-Koeffizient) und die Armutsquote bei sinkender Arbeitslosenquote rückläufig sind und umgekehrt. Die Autoren bemerken, dass bei der IT-Boom-Phase im Jahre 2000 das Ausgangsniveau der Arbeitslosigkeit höher war als in der Aufschwungphase Ende der 80er Jahre. Das wird dadurch begründet, dass der Aufschwung nicht so stark ausfiel, um das Niveau der Arbeitslosigkeit auf das der 80er Jahre zu senken. Zusätzlich hätten andere Entwicklungen zur Erhöhung von Ungleichheit und Armutsquoten beigetragen. Die Ergebnisse der Studie lassen vermuten, dass es nicht möglich ist, mit dem Indikator Arbeitslosigkeit als einzige Größe die Armutsquoten zu erklären.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Frick und Grabka (2005) untersuchen die Auswirkungen von Arbeitslosigkeit im Haushaltskontext auf die (personenbezogenen) Armutsquoten im Zeitraum 1992-2004. Dazu erzeugen sie einen Index für die Arbeitslosigkeit[18] und teilen die Haushalte in fünf Gruppen ein: Haushalte mit mindestens einer Person im erwerbsfähigen Alter ohne, mit geringer, mittlerer und starker Betroffenheit von Arbeitslosigkeit sowie Haushalte ohne erwerbsfähige Personen. In drei Konjunkturphasen (Schwächen in 1992-1998 und 2002-2004, Boom in 1999-2001) untersuchen die Autoren, wie sich die Arbeitslosigkeit auf die Armutsquote ausgewirkt hat. Sie kommen zu dem Ergebnis, dass „offensichtlich stärkere Arbeitslosigkeit zu höheren Armutsraten führt“. Der Tatsache, dass die Arbeitslosigkeit in den Gruppen niedrige und mittlere Betroffenheit z.T. geringer ist als ohne Betroffenheit, wird keine Aufmerksamkeit geschenkt. Das Ergebnis lässt keine Rückschlüsse zu, wie die Armutsquoten durch die Arbeitslosenquoten erklärt werden könnten.

Biewen und Juhasz (2011) untersuchen, zu welchem Anteil der Anstieg der Ungleichheit und Armut in Deutschland zwischen 2000 und 2005 auf die Veränderung in der Beschäftigtenstruktur (diese beinhaltet auch das Risiko der Arbeitslosigkeit) der Haushalte zurückzuführen ist. In der Studie werden die beiden Perioden 1999/2000 (Periode 0) und 2005/2006 (Periode 1) betrachtet. Ausgehend von der Dichte des Äquivalenzeinkommens in Periode 0 wird eine Einkommensdichte für 2005 erzeugt, bei der alle Faktoren gleich bleiben bis auf die bedingte Beschäftigtenstruktur (bedingt durch sozioökonomische Eigenschaften), die wie in Periode 1 ist (mit den Eigenschaften von Periode 0). Die verwendete Methode beruht auf DiNardo et al. (1996). Die Autoren kommen zu dem Ergebnis, dass die Veränderung des Arbeitseinkommens etwa 14% der Ungleichheit und 23% der Armutsquote erklären.

Im Folgenden soll der Wandel am Arbeitsmarkt und seine (möglichen) Folgen näher erläutert werden. Im Beobachtungszeitraum ist die Vollzeitbeschäftigung kontinuierlich um etwa 14% zurückgegangen, während die Teilzeitbeschäftigung um fast 50% gestiegen ist. Der Anteil der geringfügig Beschäftigten hat zwischen 1999 und 2009 stark zugenommen (um etwa 35%) und die befristete Beschäftigung (gemessen am Anteil der sozialversicherungspflichtigen Beschäftigten) ist seit 2001 ebenfalls gestiegen (Bundesagentur für Arbeit, 2011). Seit Jahren wächst die Beschäftigung durch Zeitarbeit mit nur kurzen Unterbrechungen rasant und im Vergleich zu anderen europäischen Ländern überdurchschnittlich (Bundesagentur für Arbeit, 2012).

[18] Der Index ist definiert als der im Vorjahr in Arbeitslosigkeit verbrachten Monate aller erwerbsfähigen Haushaltsmitglieder an allen potentiellen Erwerbsmonaten.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Tendenzen am Arbeitsmarkt deuten darauf hin, dass die soziale Unsicherheit zugenommen hat. Diese äußert sich zum einen im niedrigen Einkommen (Niedriglohnsektor, Zeitarbeit, Teilzeitbeschäftigung) und zum anderen in der Unsicherheit um den Arbeitsplatz (Zeitarbeit und befristete Verträge).

Weitere strukturelle Veränderungen haben sich durch die Zunahme der Selbstständigen um etwa 24% und die Abnahme der Beamten um etwa 16% im Beobachtungszeitraum ergeben (Bundesagentur für Arbeit, 2011). Reallohnsteigerungen gab es im Beobachtungszeitraum kaum; ab 2004 sind die Reallöhne sogar gefallen (Statistisches Bundesamt (2010), Bispinck (2007), Brenke (2009)). Ursachen dafür sind u.a. die sinkende Tarifbindung vieler Löhne, die moderate Lohnpolitik der Gewerkschaften zum Erhalt von Jobs und der durch die Globalisierung steigende Wettbewerbsdruck der Unternehmen (Bispinck (2007), Brenke (2009), Bundesagentur für Arbeit (2009)). Dagegen wachsen erwirtschaftete Kapitalerträge. Diese Entwicklung begünstigt die Wohlhabenden, da die ärmeren Personen nicht über Kapital zum Investieren verfügen (siehe z.B. Brenke und Grabka (2011)).

Es werden nun die Auswirkungen der Veränderungen auf verschiedene Indikatoren illustriert (siehe Abbildungen 4.27-4.35(c)), die später auf die Verwendbarkeit zur Vorhersage von Armut untersucht werden.

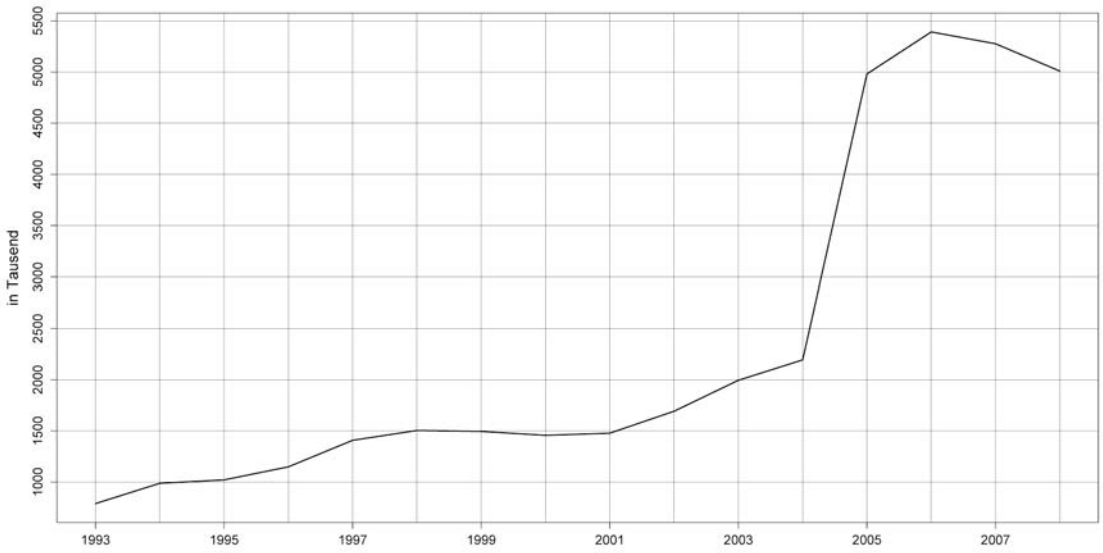


Abbildung 4.27: *Anzahl der Arbeitslosenhilfe/ Hartz IV-Bezieher (1992-2008) [Darstellung wie in Bundesagentur für Arbeit (2011)].*

Abbildung 4.27 zeigt, dass die Anzahl der Bezieher von Arbeitslosenhilfe/Hartz IV seit 2001 stark zugenommen haben. Ursachen liegen insbesondere in der hohen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Arbeitslosigkeit (bis 2006) und dem hohen Anteil der Beschäftigung im Niedriglohnsektor, der durch die Hartz-Reformen gefördert wurde, sowie im hohen Anteil der in Zeitarbeit Beschäftigten (Bundesagentur für Arbeit (2011), Bundesagentur für Arbeit (2012), Egle et al. (2003, S. 226ff.), Egle und Zohlnhöfer (2007, S. 272ff.)). Seit dem Jahr 2000 geht die steigende Anzahl der Bezieher von Arbeitslosenhilfe/Hartz IV mit einer steigenden Armutsquote einher. Jedoch ist ein Sprung von 2004 zu 2005, wie er bei der Anzahl der Bezieher zu beobachten ist, bei der Armutsquote nicht der Fall.

Die Anzahl der sozialversicherungspflichtig Beschäftigten (siehe Abbildung 4.28) ist, mit Ausnahme der Jahre 1999 bis 2001, bis 2005 stark gefallen. Ab 2006 zeichnet sich wieder ein deutlicher Anstieg ab. Ursachen für die Entwicklung der sozialpflichtig Beschäftigten sind u.a. die konjunkturelle Entwicklung und der Strukturwandel. Aber auch die Arbeitsmarktpolitik hat dazu beigetragen, da keine sozialversicherungspflichtigen Arbeitsbeschaffungsmaßnahmen gefördert wurden. Stattdessen wurden durch die Hartz-Gesetze der Niedriglohnsektor (sozialversicherungsfrei) und durch Existenzzuschüsse die Selbstständigkeit (überwiegend sozialversicherungsfrei) gefördert (siehe z.B. Bach et al. (2005)). Das Sinken der Anzahl der sozialversicherungspflichtig Beschäftigten könnte zur Erklärung des Anstiegs der Armutsquote von 2000 bis 2005 beitragen (Böhmer und Heimer, 2008).

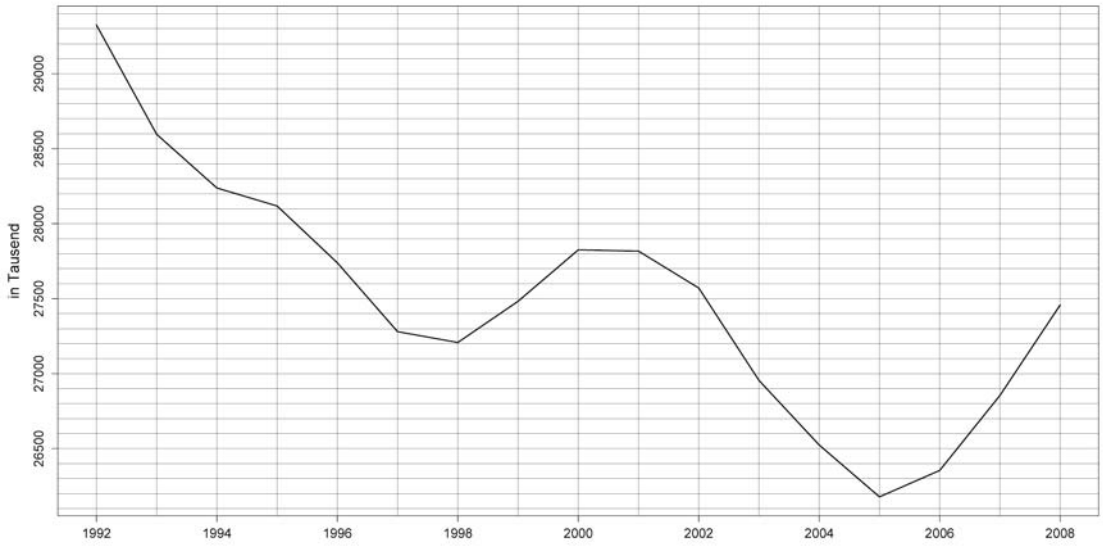


Abbildung 4.28: *Anzahl der sozialversicherungspflichtig Beschäftigten (1992-2008) [Darstellung wie in Bundesagentur für Arbeit (2011)].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Lohnquote, die definiert ist als Arbeitnehmerentgelte am gesamten Volkseinkommen (siehe Abbildung 4.29, Statistisches Bundesamt, 2009), ist im Beobachtungszeitraum, mit Ausnahme der Jahre 1999 und 2000 sowie 2008, gesunken und hat in 2007 mit 61% den niedrigsten Wert überhaupt erreicht. Ursache einer niedrigen Lohnquote ist die Verschiebung der Verteilung des Einkommens zu Kapital- und Selbstständigen-Einkünften. Die Lohnquote, wenn auch sehr umstritten, wird als Indikator für die ungleiche Verteilung des Einkommens angesehen. Je niedriger die Lohnquote, desto ungleicher sei die Einkommensverteilung. Der Grund dafür ist, dass Kapitaleinkünfte bei niedriger Lohnquote einen hohen Stellenwert einnehmen und Haushalte mit geringen Einkünften selten am Kapitalmarkt agieren (siehe z.B. Brenke und Grabka (2011)).

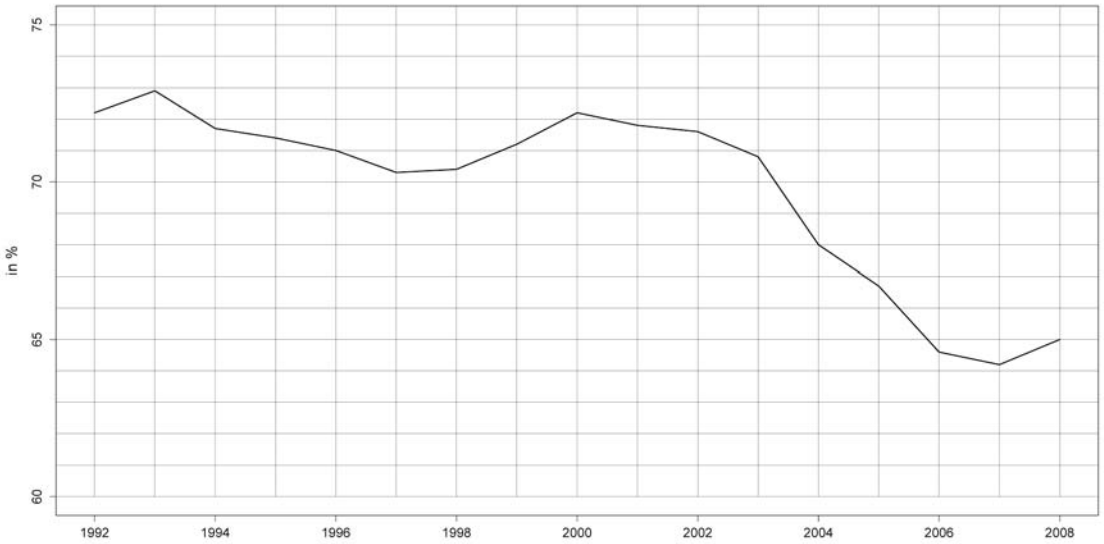


Abbildung 4.29: *Lohnquote (1992-2008) [Berechnung auf Basis der Quelle Statistisches Bundesamt (2010)].*

In Abbildung 4.30 ist die Lohnentwicklung als Veränderung des Indikators Reallöhne (netto pro Stunde) zum Vorjahr abgebildet (Statistisches Bundesamt (2010)). Die rückläufige Lohnentwicklung könnte zum Anstieg der Armut beigetragen haben.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

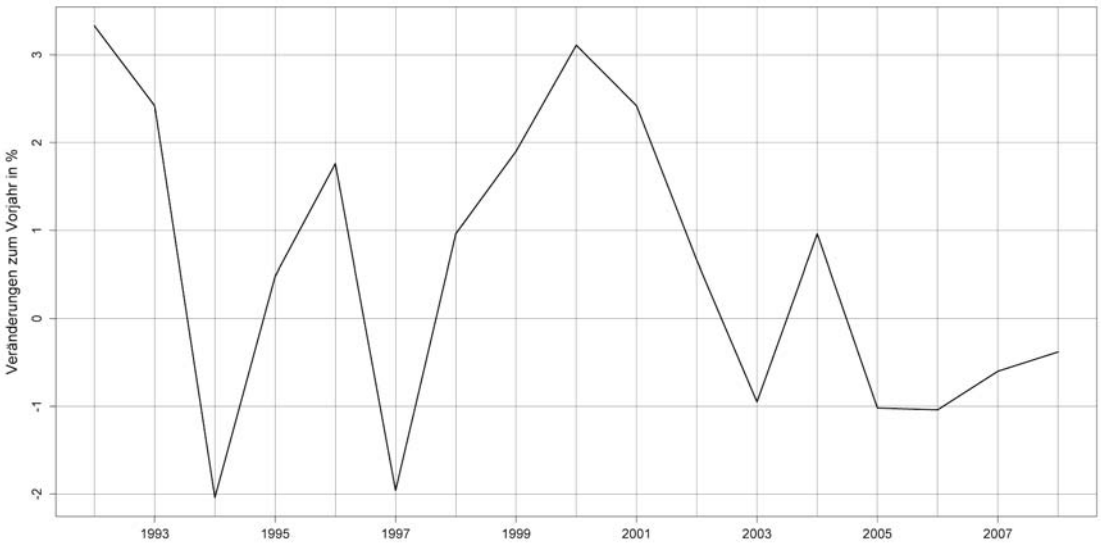


Abbildung 4.30: *Veränderung der preisbereinigten Nettolöhne und Gehälter je Stunde zum Vorjahr in % (1992-2008) [Darstellung basierend auf Statistisches Bundesamt (2010)].*

4.3.1.2 Steuern und Abgaben

In diesem Abschnitt wird auf verschiedenen Ebenen der Besteuerung diskutiert, inwieweit die Steuerreform 2000 zur Entlastung der Armen oder zur Begünstigung der Wohlhabenden und damit zur Verschärfung der Ungleichheit beigetragen hat.

Tabelle 4.18 zeigt die Entwicklung der Einkommenssteuer im Beobachtungszeitraum. Zum Ende der Regierungszeit Helmut Kohls waren sowohl der Eingangssteuersatz mit 25.9% als auch der Spitzensteuersatz mit 53% aufgrund der Kosten der Wiedervereinigung sehr hoch (Egle et al., 2003, S.194).

Zeitraum	Grundfreibetrag	Eingangssteuersatz	Ende Progressionszone	Spitzensteuersatz
-1995	2871	19%	61376	53%
1996-1997	6184	25.9%	61376	53%
1998	6322	25.9%	61376	53%
1999	6681	23.9%	61376	53%
2000	6902	22.9%	58643	51%
2001	7206	19.9%	54998	48.5%
2002-2003	7235	19.9%	55008	48.5%
2004	7664	16%	52152	45%
2005-2008	7664	15%	52152	42%

Tabelle 4.18: *Entwicklung des Steuersätze (1992-2008) [in Anlehnung an Biewen und Juhasz (2010), Bundesfinanzministerium (BMF)(2011)].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Im Zuge der Steuerreform 2000 wurde der Eingangssteuersatz stufenweise von 25.9% auf 15%, d.h. um 10.9 Prozentpunkte (entspricht 40%) gesenkt, während der Spitzensteuersatz von 53% auf 42%, d.h. um 11 Prozentpunkte (entspricht 20%) herabgesetzt wurde. Nach Butterwegge (2005, S. 170f.) kann man eine solche Steuerreform trotz der Senkung der Grundfreibeträge und der Entlastungen von Familien nicht als sozial bezeichnen. „Während die unteren Einkommensgruppen in 2005 kaum mehr Geld zur Verfügung hatten als 5 Jahre zuvor, mussten ledige und kinderlose Einkommensmillionäre jetzt über 90 000 € pro Jahr weniger Steuern zahlen als damals“. Die Verringerung der Progressionszone führte ebenfalls zu einer Begünstigung der Wohlhabenden. Mit dem Einkommen, das durch die Entlastungen gespart wird, können auf dem Kapitalmarkt Gewinne erwirtschaftet werden, die oberhalb des Freibetrags mit einer Steuer von 25% (also deutlich unter dem Spitzensteuersatz), der sogenannten Abgeltungssteuer, belegt werden (Butterwegge, 2005).

Zugunsten der ärmeren Bevölkerung wurde der Grundfreibetrag auf 7 664 € angehoben. Jedoch profitieren Individuen, deren Einkommen bereits vor der Reform unter der Grenze des Grundfreibetrages lagen, nicht davon. Als Beispiel seien Beschäftigte in Mini-Jobs genannt, die bis zu 400 € im Monat und damit maximal 4 800 € im Jahr verdienen, genannt.

Indirekte Steuern treffen alle Personen unabhängig ob arm oder reich im gleichen Maße. Seit einigen Jahren vollzieht sich ein Wandel von der direkten Steuer zur indirekten Steuer. Unter der indirekten Steuer wird die Umsatzsteuer und die Verbrauchssteuer (z.B. auf Energie, Strom, Tabak) verstanden, die nicht durch das Finanzamt sondern von Dritten erhoben wird. Im Beobachtungszeitraum wurde der Regelsteuersatz der Mehrwertsteuer von 14% in 1992 stufenweise auf 19% in 2007 angehoben, während der ermäßigte Satz bei 7% belassen wurde. Durch die ökologische Steuerreform wurden die Stromsteuer (pro kWh) sowie die Mineralölsteuer eingeführt. Letztere trug stark zur Preissteigerung der Kraftstoffe bei. Geringverdiener, die mit dem Auto zur Arbeitsstätte pendeln müssen, werden mit höheren Kosten belastet und sind in Hinsicht auf die Wahl eines neuen Arbeitsplatzes stark in ihrer Mobilität eingeschränkt. Die ökologische Steuerreform belastet die privaten Haushalte und nicht die Industrie, die für die Umweltverschmutzung die Hauptlast trägt (Butterwegge, 2005, S. 170).

Für Unternehmer und Großaktionäre wurden per Gesetz ebenfalls Steuerbegünstigungen eingeführt. Die Unternehmer wurden durch Senkung der Körperschaftssteuer von 40% bzw. 30% auf 25% und die Anteilseigner durch das Halbeinkünfteverfahren, das vorsieht, dass Anteilseigner nur noch die Hälfte der Ausschüttungen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

einer Kapitalgesellschaft bei der Einkommenssteuer versteuern müssen, entlastet (Butterwegge, 2005, S. 170ff.).

In Abbildung 4.31 wird für eine Gesamtbetrachtung der Indikator Abgabenquote betrachtet. Die Abgabenquote misst den Anteil von Steuern und Abgaben am Bruttoinlandsprodukt (BIP). Wie die Abbildung zeigt, sind die Abgaben nach der Steuerreform 2000 gefallen, die Armut ist jedoch gestiegen. Dies ist ein weiteres Indiz dafür, dass Personen ohne oder mit geringem Einkommen nicht von den Steuerentlastungen profitieren, die Wohlhabenden dagegen begünstigt werden.

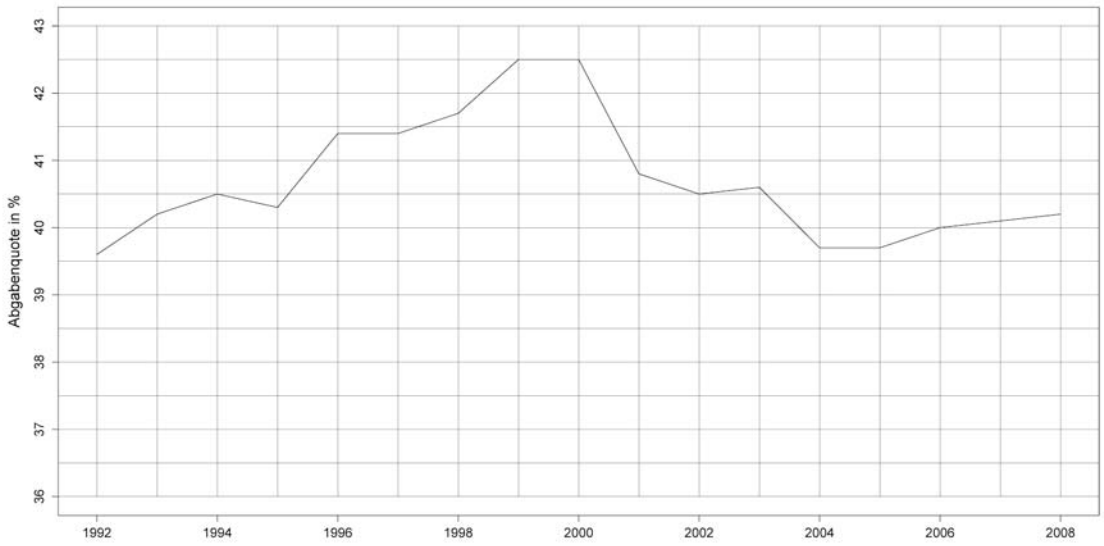


Abbildung 4.31: *Abgabenquote (1992-2008) (Abgrenzung der Volkswirtschaftlichen Gesamtrechnung (VGR)) [Darstellung basierend auf den Daten des BMFs (2011)].*

4.3.1.3 Wirtschaftswachstum und Finanzmarkt

In Abbildung 4.32 und 4.35(c) ist die Konjunkturentwicklung im Beobachtungszeitraum durch die Indikatoren Wirtschaftswachstum und Renditen der DAX-Jahresschlussstände dargestellt. Wird ein Konjunkturzyklus durch den absoluten Rückgang des Wirtschaftswachstums definiert (siehe z.B. BMF (2007))[19], so las-

[19]Es gibt bisher kein einheitliches Konzept zur Beschreibung und Erklärung von Konjunkturzyklen für Deutschland. Verschiedene nicht-parametrische und parametrische Methoden wie die „Presse-Methode“, die modifizierte „Presse-Methode“ nach Boldin (1994) und der BBQ Algorithmus sowie Markov Switching Modelle wurden u.a. von Schirwitz (2007) auf deutsche Daten angewendet. Diese führten z.T. zu unterschiedlichen Längen der Zyklen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sen sich im Beobachtungszeitraum das Ende eines Zyklus sowie zwei volle Zyklen betrachten. Die Enden der drei Zyklen sind durch die Rezession nach der Wiedervereinigung in 1993, das Platzen der Dot-Com Blase sowie die Terroranschläge 2001 (mit Schwächephase bis 2004) und der Finanzkrise in 2008 datiert.

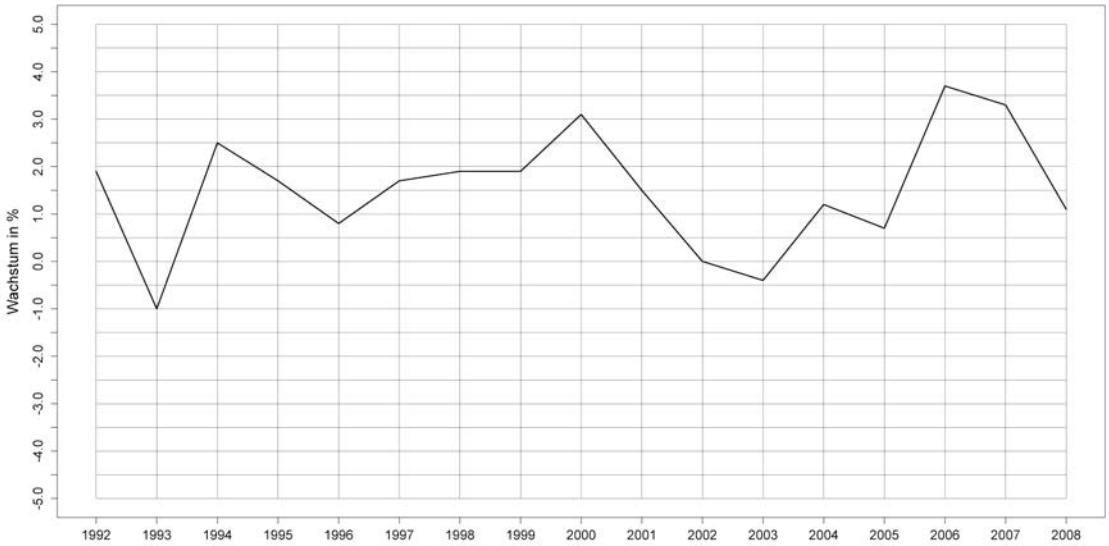


Abbildung 4.32: *Veränderung des Wirtschaftswachstums in Deutschland zum Vorjahr in % (1992-2008) [Darstellung wie in Statistisches Bundesamt (2010)].*

Abbildung 4.33: *Renditen Dax-Schlussstände (1992-2008) [Darstellung basierend auf den Daten von* $http://www.wifikon.de/index.php?article_id = 72$*].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Es zeigt sich, dass die Konjunkturindikatoren in den 90er Jahren zur Erklärung der Armutsentwicklung beigetragen haben könnten. So stieg die Armutsquote im Jahr nach der Rezession von 1993, während sie darauf in der wirtschaftlichen Hochphase bis etwa 1999 fiel, um dann nach dem Platzen der Dot-Com Blase wieder anzusteigen. Ab etwa 2003 stieg die Armutsquote dann jedoch weiter, während die Wirtschaft bis etwa 2006 gewachsen ist und der DAX hohe Renditen erzielte.

In der Literatur haben sich z.B. Frick et al. (2005) und Butterwegge (2005) mit der Frage beschäftigt, inwiefern sich die Armut durch die Konjunktur erklären lässt.

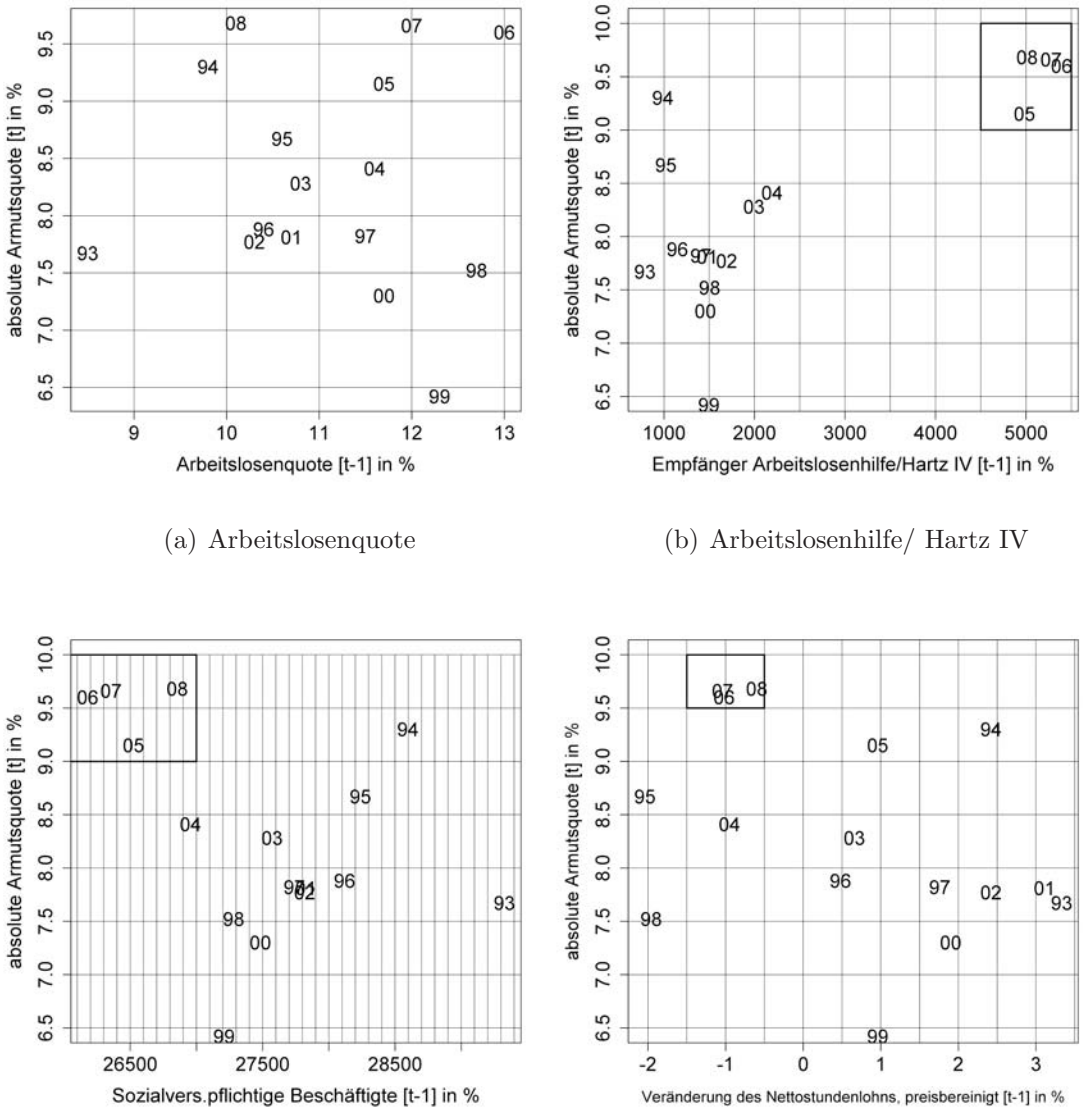
Frick et al. (2005) vergleichen die Zyklen von Armut und Wirtschaftswachstum für Westdeutschland im Zeitraum von 1985-2004. Sie kommen zu dem Ergebnis, dass steigendes Wirtschaftswachstum sinkende Armut impliziert und umgekehrt. Für den gemeinsamen Betrachtungszeitraum dieser Arbeit und der genannten Studie sind die Ergebnisse ähnlich.

Nach Butterwegge (2005, S. 222) sind hohe Armutsquoten dagegen nicht durch „externe Schocks“ wie das Platzen der Dot-Com Blase, die Terroranschläge vom 11. September oder schwachem Wirtschaftswachstum zu erklären. Der Autor stellt heraus, dass Armut ein „Struktur- und kein Konjunkturproblem“ ist.

Im Folgenden wird untersucht, ob die beschriebenen Indikatoren sich als Prognosevariablen für die Armutsquoten eignen. Diese liefern zwar nur die Aussage, wie viele Haushalte und nicht welche Haushalte arm werden, dennoch könnten diese Informationen die Prävention von Armut unterstützen und Anhaltspunkte für die Prognosefähigkeit der Vulnerabilität mit Makrovariablen liefern, auf die später eingegangen wird.

Die volkswirtschaftlichen Indikatoren sind zur Prognose von Armut geeignet, wenn sich die Armut als Funktion der Indikatoren darstellen lässt. Um einen Eindruck der Prognosefähigkeit von Armutsquoten durch genannte Variablen zu bekommen, werden die Armutsquoten (gesamt) in den Abbildungen 4.34 und 4.35 dargestellt.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Arbeitslosenquote

(b) Arbeitslosenhilfe/ Hartz IV

(c) Sozialversicherungspflichtig Beschäftigte

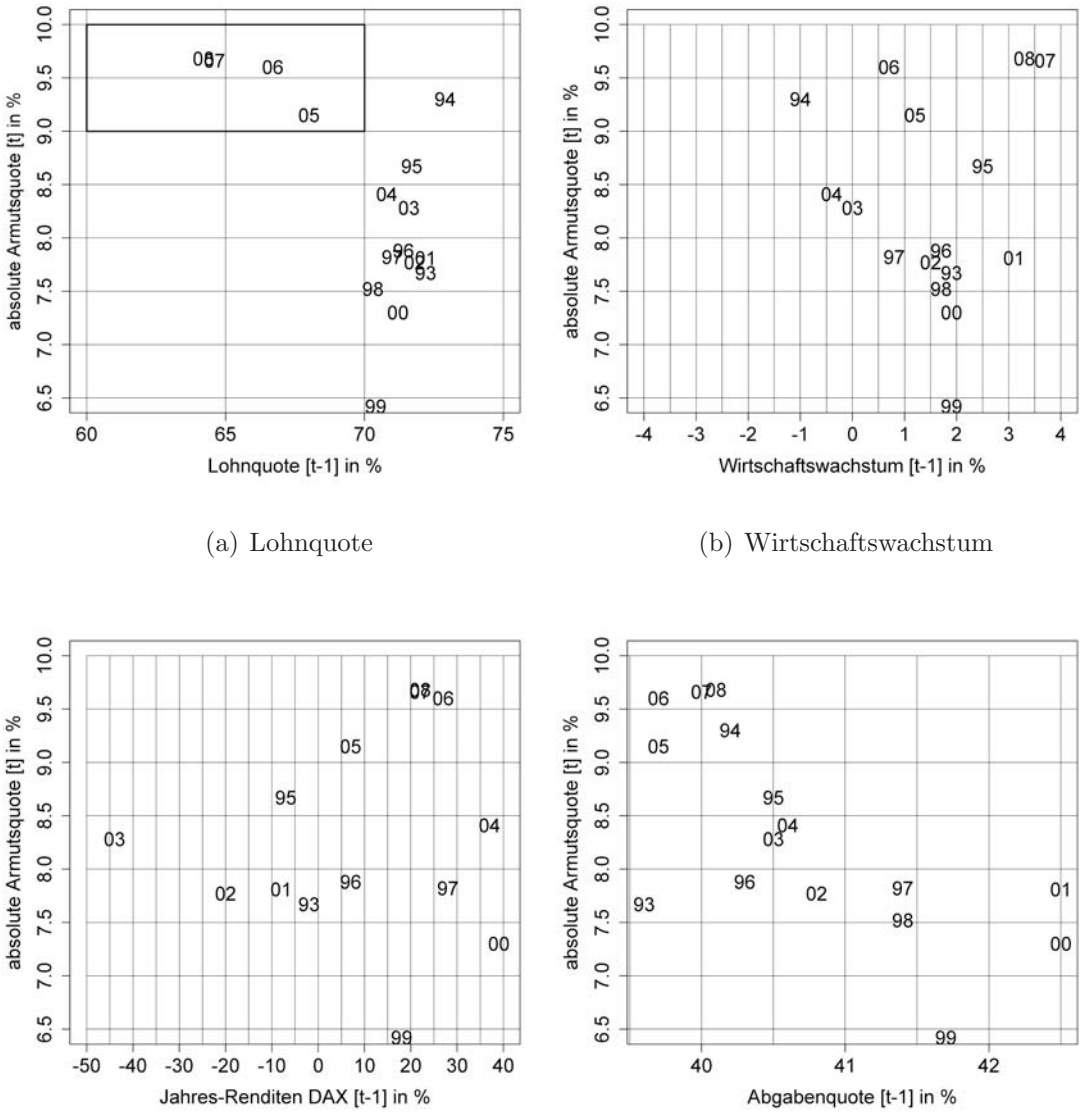
(d) Veränderung des Nettostundenlohns in %

Abbildung 4.34: *Armut im Jahr t in Abhängigkeit von Indikatoren des Arbeitsmarktes des Vorjahres (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten, Bundesagentur für Arbeit (2011), Statistisches Bundesamt (2010)]. Die Punkte werden durch die letzten beiden Ziffern des Jahres t dargestellt.*

Die Abbildungen zeigen, dass sich keiner der Indikatoren im Einzelnen als Kovariable zur Vorhersage der Armutsquote eignet. Die Arbeitslosenquote (Abbildung 4.34(a)) zeigt kein Verhaltensmuster, während die übrigen Indikatoren eine relativ hohe Korrelation zur Armutsquote haben (Abbildung 4.34(b)-4.35(d)). Diese sind

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

allerdings auf Trends und Zyklen zurückzuführen. Das Niveau der Armut ist in verschiedenen Zyklen bei konstantem Wert der Indikatoren unterschiedlich hoch. Da es sich sowohl bei der Armutsquote und den Indikatoren um Zeitreihen handelt, eignet sich ein Zeitreihenmodell vermutlich besser zur Prognose der Armut. Allerdings sind die betrachteten Zeitreihen sehr kurz.



(a) Lohnquote

(b) Wirtschaftswachstum

(c) Dax-Renditen

(d) Abgabenquote

Abbildung 4.35: *Armut im Jahr t in Abhängigkeit von weiteren volkswirtschaftlichen Indikatoren des Vorjahres (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten, Bundesagentur für Arbeit (2011), Statistisches Bundesamt (2010)]. Die Punkte werden durch die letzten beiden Ziffern des Jahres t dargestellt.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.3.2 Die Haushaltsstruktur als Ursache von Armut

Die OECD Studien von 2008 und 2011 nennen verschiedene Veränderungen in der Entwicklung der Haushaltsstruktur, die zum Anstieg von Einkommensungleichheit und Armut beigetragen haben, z.B. die Zunahme von Single-Haushalten und Alleinerziehenden sowie Partnerschaften in derselben Einkommensklasse.

Im Folgenden soll die Entwicklung der Haushaltsstruktur nach verschiedenen Merkmalen, deren Ausprägungen im Folgenden als Gruppen bezeichnet werden, näher untersucht werden (siehe Tabelle 4.19). Anschließend wird für jedes Merkmal die Armut der einzelnen Gruppen betrachtet. Auf die Berücksichtigung von Interaktionseffekten zwischen den Merkmalen wird verzichtet.

Merkmal	Ausprägungen (Namen der Gruppen)
Geschlecht (HV)	männlich, weiblich
Alter (HV)	$n[18, 25]$, $n[26, 45]$, $n[46, 65]$, $n[66, -)$
Region	Nord-, Ost-, West-, Süddeutschland
Familienstand (HV)	verheiratet oder mit Partner, alleinstehend, verwitwet, geschieden, getrennt
Gewerbe (HV)	nicht erwerbstätig, Landwirtschaft, Energie, Bergbau, Produktion, Baugewerbe, Handel, Transport, Bank/Versicherung, Dienstleistungssektor
Schulbildung (HV)	Hauptschulabschluss, Realschulabschluss, Abitur/FHSR, andere Abschlüsse, keinen Schulabschluss
Kinder im Haushalt	kinderlos, 1 Kind, 2 Kinder, 3 und mehr Kinder
Erziehung (HV)	alleinerziehend, zwei Elternteile
Migration (HV)	kein-, direkter-, indirekter Migrationshintergrund

Tabelle 4.19: *Gruppierungen nach Haushaltseigenschaften*

Während sich einige Merkmale auf den gesamten Haushalt beziehen, charakterisieren andere, wie schon zuvor, nur den Haushaltsvorstand (HV). Es wird dennoch angenommen, dass diese Merkmale charakteristisch für den gesamten Haushalt sind.

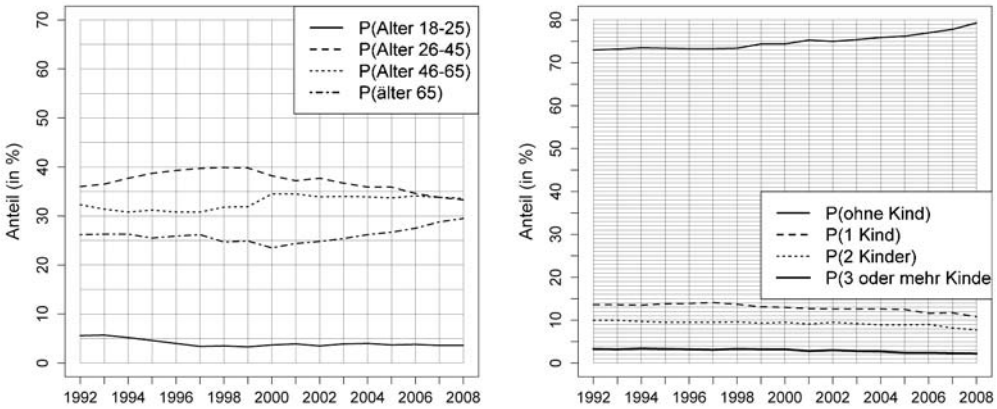
Folgende Festlegungen werden getroffen: Ein Haushalt wird als kinderlos definiert, wenn keine Kinder unter 18 Jahren im Haushalt leben, d.h., dass Kinder mindestens 18 Jahre alt sind und/oder den Haushalt bereits verlassen haben. Ein

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Haushalt, dem Kinder (unter 18) und nur eine erwachsene Person angehören, wird als alleinerziehend angesehen.

4.3.2.1 Entwicklung der Haushaltsstruktur nach Gruppen

In den Abbildungen 4.36 und 4.37 sind die Entwicklungen der Haushaltsstruktur von oben genannten Merkmalen über den Zeitraum 1992-2008 dargestellt. In den Abbildungen 4.36(a) - 4.37(b) sowie in Tabelle D.16 im Anhang A.4 sind die Faktoren des demografischen Wandels zu beobachten. Die Alterung der Gesellschaft zeigt sich im Rückgang der Haushalte mit HV im Alter zwischen 26 und 45 und die Zunahme der Haushalte mit HV über 65. Der Anteil der kinderlosen Haushalte hat ebenfalls zugenommen (deutet auf sinkende Geburtenzahl hin). Weiterhin sind die Haushalte mit weiblichem HV angestiegen (deutet auf Zunahme der alleinstehenden, geschiedenen bzw. alleinerziehenden[20] Haushalte hin, siehe Abbildungen 4.37(c) und 4.37(d)), wobei zwischen den Jahren 1999 und 2000 kurzzeitig wieder ein sprunghafter Rückgang zu beobachten war. Der Anteil der Haushalte mit indirektem und direktem Migrationshintergrund hat leicht zugenommen.

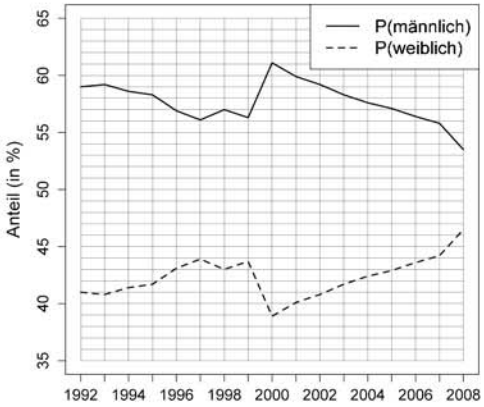


(a) Merkmal: Alter (b) Merkmal: Kinder

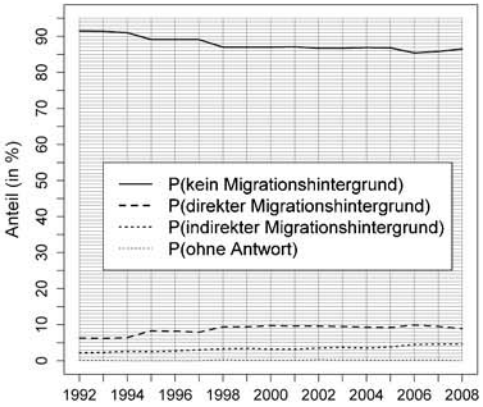
Abbildung 4.36: *Entwicklung der Haushaltsstruktur nach verschiedenen Merkmalen (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

[20] Aufgrund der Definition der kinderlosen Haushalte und der Alleinerziehenden ist letztere Gruppe trotz steigender Tendenz im vorliegenden Datensatz mit etwa 3% ($= 0.2 \cdot 0.15$) relativ klein

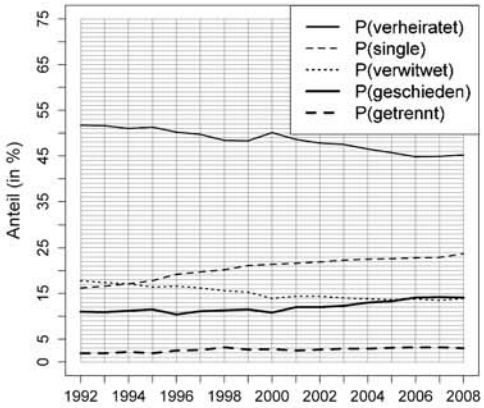
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



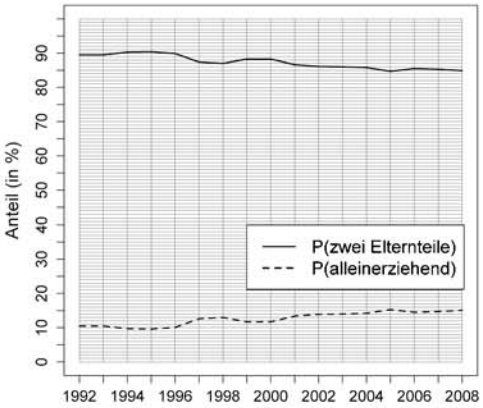
(a) Merkmal: Geschlecht



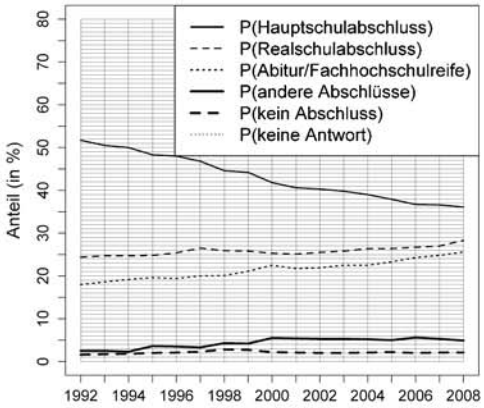
(b) Merkmal: Migration



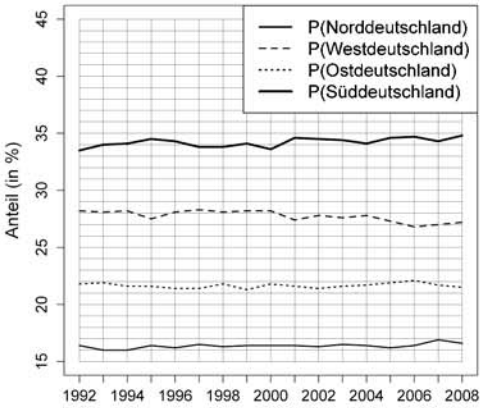
(c) Merkmal: Familienstand



(d) Merkmal: Alleinerziehend



(e) Merkmal: Bildung



(f) Merkmal: Region

Abbildung 4.37: *Fortsetzung Entwicklung der Haushaltsstruktur nach verschiedenen Merkmalen (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Ferner sind Haushalte, deren HV einen Hauptschulabschluss hat, zu Gunsten des Realschulabschlusses, des Abiturs bzw. der Fachhochschulreife sowie anderer Abschlüsse zurückgegangen (siehe Abbildung 4.37(e)). Bei der Aufteilung der Haushalte auf die vier Regionen sind nur leichte Wanderungen über die Zeit festzustellen (siehe Abbildung 4.37(f)).

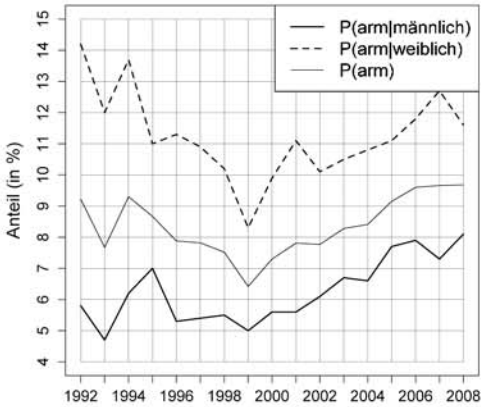
4.3.2.2 Armut von Haushalten nach Gruppen

In diesem Abschnitt wird untersucht, welche Gruppen in Deutschland besonders stark von Armut betroffen sind und welche Veränderungen in der Struktur der Haushalte möglicherweise zu einer Veränderung der Armutsquote beigetragen haben könnten. In den Abbildungen 4.38-4.40 wird die Armut nach Gruppen der Merkmale unter zwei verschiedenen Gesichtspunkten dargestellt. Die Abbildungen auf der linken Seite zeigen (für jede Gruppe eines Merkmals) die Armutsquoten von Haushalten in diesen Gruppen. Sie wurden berechnet als Anteil armer Haushalte einer Gruppe an allen Haushalten dieser Gruppe (z.B. $\frac{\sharp\text{arme Haushalte mit weiblichem HV}}{\sharp\text{Haushalte mit weiblichem HV}}$). Diese Armutsquoten werden im Folgenden als *bedingte Armutsquote* bezeichnet. In den Abbildungen auf der rechten Seite sind die Armutsquoten als Anteil von armen Haushalten einer Gruppe an allen existierenden Haushalten einer Population dargestellt (z.B. $\frac{\sharp\text{arme Haushalte mit weiblichem HV}}{\sharp\text{Haushalte in der Population}}$). Sie werden im Folgenden mit *Gruppenarmutsquoten* bezeichnet. Während die bedingte Armutsquote ein Maß dafür ist, wie stark die Haushalte in den Gruppen von Armut betroffen sind, veranschaulicht die Gruppenarmutsquote, wie stark die Gruppen gemessen an der Gesamtpopulation von Armut betroffen sind. Somit müsste die Summe der Gruppenarmutsquoten eines Merkmals die Gesamtarmutsquote ergeben. Allerdings liegen nicht für alle Haushalte, deren Einkommen zur Berechnung der Armutsquoten einbezogen wurde, die Ausprägungen aller Merkmale vor. Der Anteil für *Item-Nonresponse* ändert sich von Merkmal zu Merkmal und von Jahr zu Jahr. Bei Darstellung der Armutsquoten ist dieser Anteil in den folgenden Abbildungen und Tabellen nicht dargestellt. Deshalb muss die Summe der Gruppenarmutsquoten *nicht* unbedingt der Gesamtarmutsquote entsprechen.

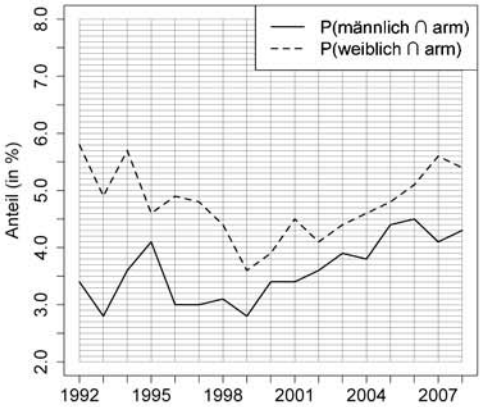
Zur Bestimmung der bedingten Armutsquoten und der Gruppenarmutsquoten wurde für alle Gruppen die gleiche Armutsgrenze bei 9 000 € verwendet. Es wurde beobachtet, dass die relativen Armutsgrenzen bei 50% des Medianeinkommens über die Zeit und zwischen den Gruppen wenig variieren.

Im Folgenden werden die beiden Armutsquoten für alle Gruppen pro Merkmal betrachtet.

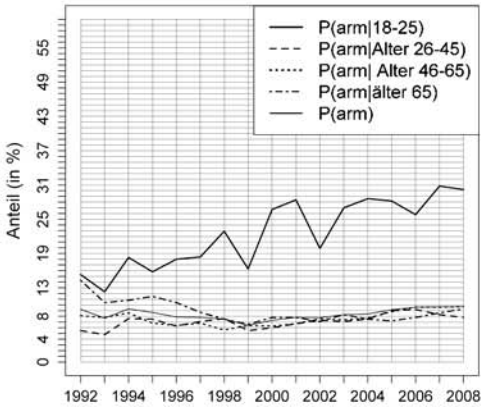
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



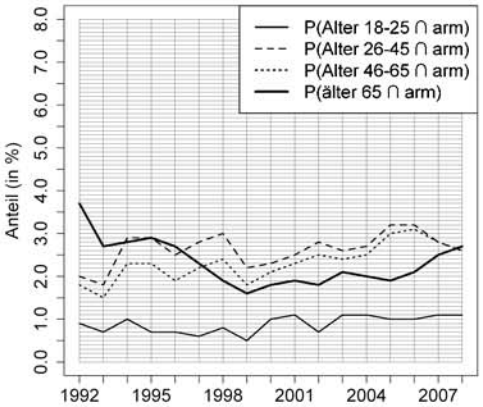
(a) Merkmal: Geschlecht



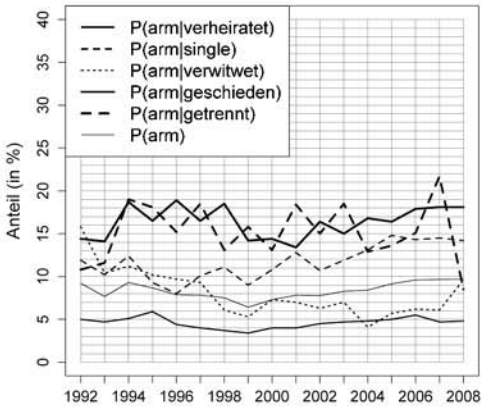
(b) Merkmal: Geschlecht



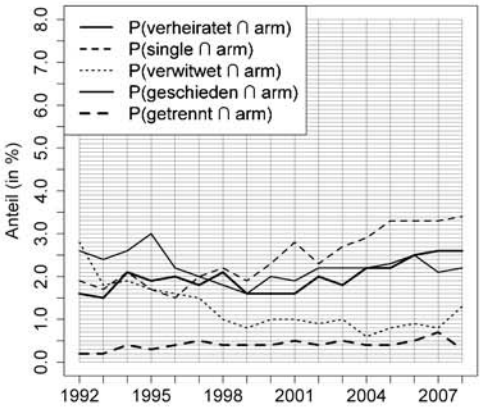
(c) Merkmal: Alter



(d) Merkmal: Alter



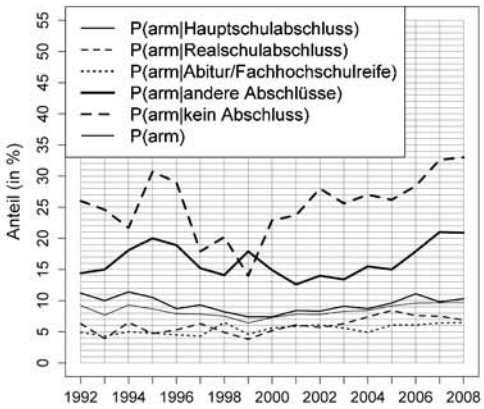
(e) Merkmal: Familienstand



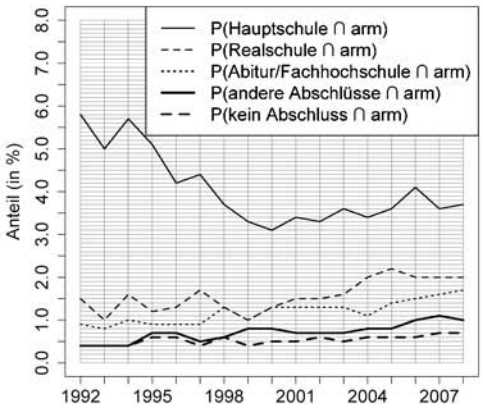
(f) Merkmal: Familienstand

Abbildung 4.38: *Entwicklung der Armut nach verschiedenen Merkmalen (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

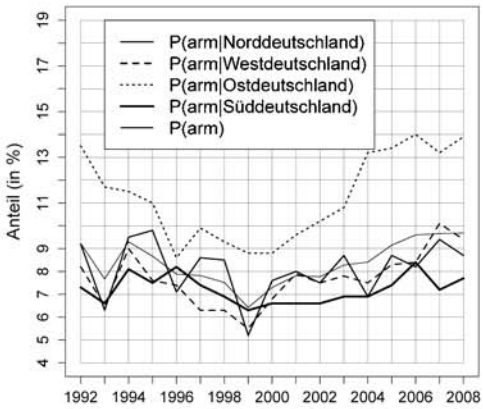
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



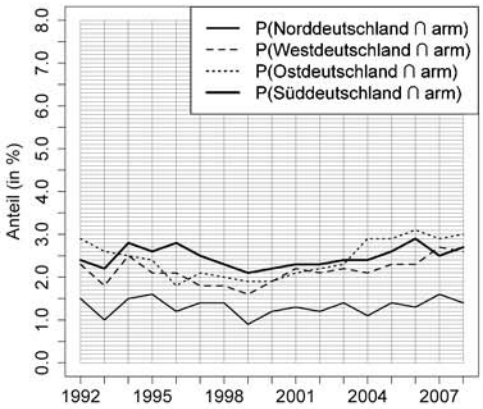
(a) Merkmal: Bildung



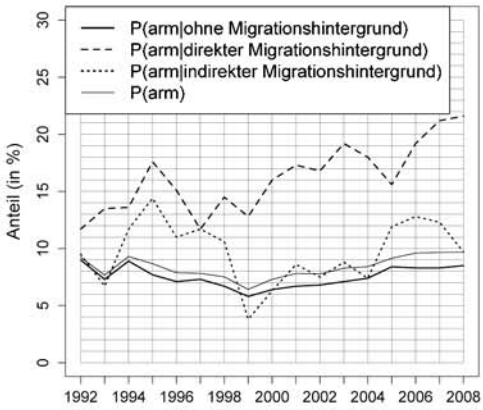
(b) Merkmal: Bildung



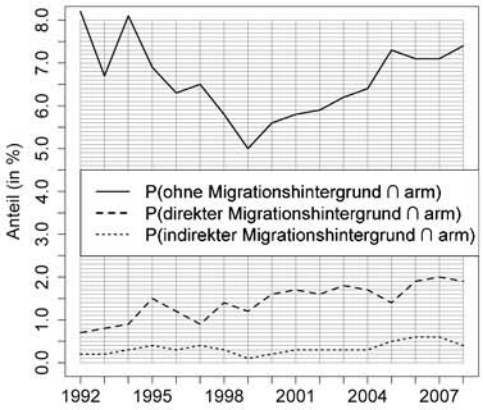
(c) Merkmal: Region



(d) Merkmal: Region



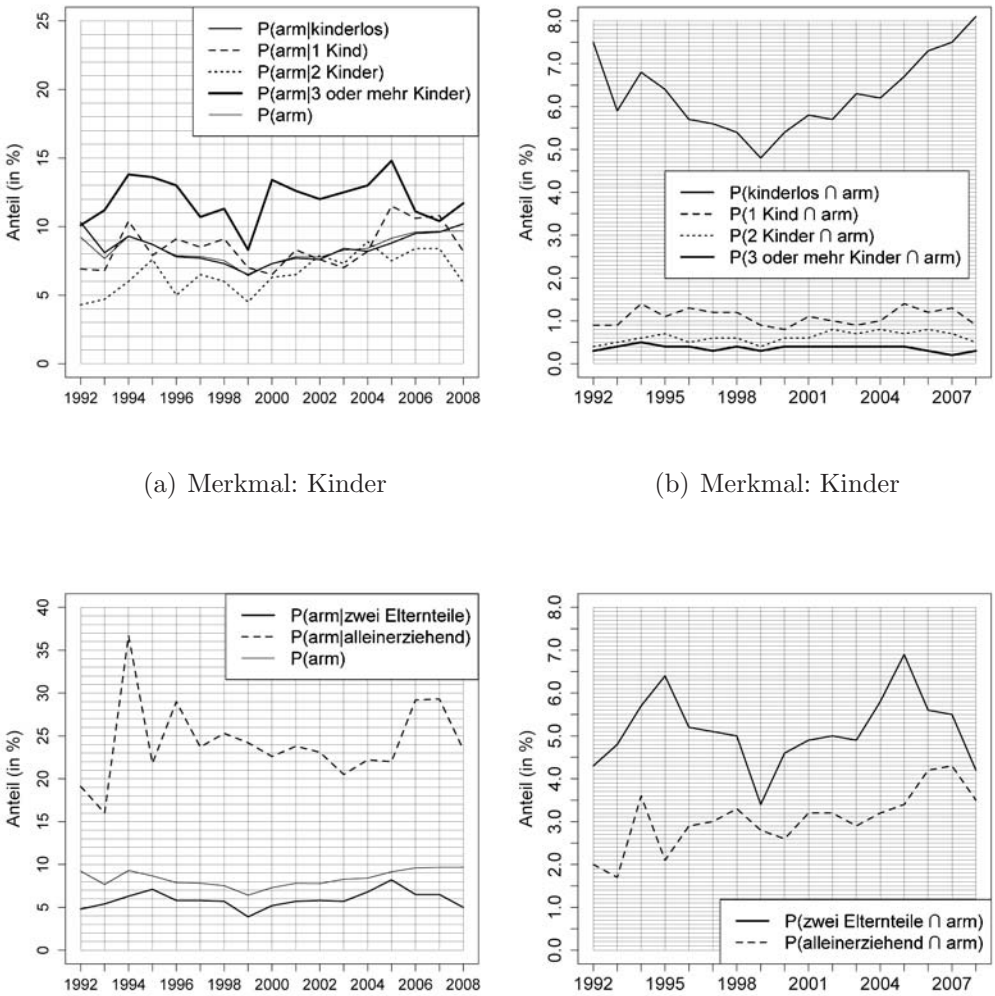
(e) Merkmal: Migration



(f) Merkmal: Migration

Abbildung 4.39: *Fortsetzung Entwicklung der Armut nach verschiedenen Merkmalen 1992-2008 [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Merkmal: Kinder

(b) Merkmal: Kinder

(c) Merkmal: Alleinerziehend

(d) Merkmal: Alleinerziehend

Abbildung 4.40: *Fortsetzung Entwicklung der Armut nach verschiedenen Merkmalen (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Gruppiert nach dem Merkmal Geschlecht zeigt sich, dass Haushalte mit weiblichem HV eine höhere Armutsquote aufweisen als Haushalte mit männlichem HV (Abbildung 4.38(a)). Die Unterschiede in den bedingten Armutsquoten können durch die üblichen genderspezifischen Nachteile von Frauen wie z.B. schlechtere Bezahlung oder Arbeitsunterbrechung durch Geburt und Erziehung von Kindern begründet werden (siehe dazu z.B. BMFSFJ (2004)). Zu Beginn und in der Mitte der 90er Jahre war die Differenz der bedingten Armutsquoten besonders groß. Ursache könnte die sehr hohe Frauenarbeitslosenquote in Ostdeutschland nach der Wende gewesen

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sein (siehe dazu z.B. BMFSFJ (2004)). Ebenso liegen die Gruppenarmutsquoten von Haushalten mit weiblichem HV über denen mit männlichem HV (Abbildung 4.38(b)). Die Armutsquoten weisen bzgl. der Trends einen ähnlichen Verlauf auf, so dass die steigende Ungleichheit und Armut seit dem Jahr 2000 zum einen auf den Anstieg der bedingten Armutsquoten von Haushalten mit männlichem HV und weiblichem HV und zum anderen auf die Zunahme von Haushalten mit weiblichem HV (und gleichzeitige Abnahme von Haushalten mit männlichem HV) zu begründen ist. Es zeigt sich hier ein doppelter Effekt: eine Zunahme der bedingten Armutsquoten und eine Verschiebung der Haushaltsstruktur.

Bei Betrachtung der bedingten Armut von Altersgruppen fallen insbesondere die Haushalte mit jungem HV auf. Deren bedingte Armutsquoten sind von 1992 bis 2008 linear von etwa 15% auf 30% angestiegen. Gründe dafür sind vermutlich Schwierigkeiten beim Berufseinstieg (unbezahlte Praktika, Zeitarbeit, befristete Stellen und Teilzeitstellen, sowie schlechte Ausbildung und geringe Flexibilität hinsichtlich der Berufswahl). Die Haushalte mit HV in den übrigen Altersklassen haben im Vergleich dazu moderate, bedingte Armutsquoten, jene der Haushalte mit HV älter als 65 sind zuletzt wieder etwas angestiegen. Für die Zukunft wird befürchtet, dass Altersarmut zunehmen könnte (Butterwegge (2009, S. 92ff.), Goebel und Grabka (2011)). Als Gründe dafür werden u.a. die hohen Arbeitslosenquoten, (verbunden mit einer Einzahlungsunterbrechung in die gesetzliche Rentenversicherung (GRV)), der zunehmende Anteil an geschiedenen und alleinerziehenden Haushalten, die Aussetzung der Rentenanpassung in 2004 und den folgenden Jahren, verbunden mit der Zunahme der Relevanz der privaten Vorsorge und die Anhebung des Renteneintrittsalters von 65 stufenweise bis 67 genannt. Die Gruppenarmutsquoten von Haushalten mit jungem HV machen aufgrund der kleinen Stichprobengröße auch nur einen kleinen Anteil der Gesamtarmutsquote aus. Die Gruppenarmutsquoten der Haushalte mit HV älter als 65 sind von 1992-1999 gefallen, dann aber wieder angestiegen und liegen nun wieder gleich auf mit denen der Haushalte mit HV in der Altersgruppe 26-45 und 45-65. Das Niveau ist mit 2.5% moderat. Basierend auf den bisherigen Erkenntnissen scheint die Gruppierung nach dem Merkmal Alter kaum zur Erklärung des Anstiegs von Ungleichheit und Armut ab dem Jahr 2000 beizutragen.

Abbildung 4.38(e) zeigt, dass in der Gruppierung nach Familienstand getrennte und geschiedene Haushalte eine hohe bedingte Armutsquote zwischen 15% und 20% aufweisen. Während die bedingten Armutsquoten von alleinstehenden Haushalten, seit etwa 2002 auf ein Niveau von 15% angestiegen sind, sind die bedingten Armutsquoten von verwitweten Haushalten (mit Ausnahme von 2008) kontinuierlich

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

gesunken. Mögliche Gründe für den Anstieg der bedingten Armut der Alleinstehenden wurden in Abschnitt 4.3.1.1 und für das Sinken der bedingten Armut der Haushalte mit verwitwetem HV in Abschnitt 4.1.3 genannt. In Abbildung 4.38(f) ist ein deutlicher Anstieg der Armutsquoten von alleinstehenden und ein mäßiger Anstieg von geschiedenen Haushalten erkennbar. Diese Ergebnisse bestätigen die Studien der OECD (2008) und (2011), dass eine Veränderung der Haushaltsstruktur zu mehr alleinstehenden und geschiedenen Haushalten zum Anstieg von Ungleichheit und Armut geführt hat. Bei den alleinstehenden Haushalten zeigt sich ein doppelter Effekt: Die Gruppengröße hat zugenommen, verbunden mit einem Anstieg der (bedingten) Armut in dieser Gruppe. Für die Gruppe der Haushalte mit verwitwetem HV ist das Umgekehrte der Fall; die Gruppengröße und die bedingte Armut sind gefallen. Die steigende Armut von Alleinstehenden lässt sich auch durch die gestiegene soziale Unsicherheit auf dem Arbeitsmarkt erklären (hohe Arbeitslosigkeit, Niedriglohn, Zeitarbeit und Teilzeitbeschäftigung, siehe Abbildung 4.3.1 und Betreuung der Kinder zumeist gegen Entgelt), deren Last sie alleine tragen müssen.

Die Gruppierung nach dem Merkmal Bildung zeigt erwartungsgemäß sehr hohe bedingte Armutsquoten für Haushalte mit HV ohne Schulabschluss und hohe bedingte Armutsquoten für HVs mit Hauptschulabschluss (4.39(a)). Die Gruppenarmutsquoten von HVs mit Hauptschulabschluss sind etwas gesunken, während die Armutsquoten der HVs mit Realschulabschluss moderat gestiegen sind. Die Gruppe der Schulabbrecher fällt aufgrund der kleinen Gruppengröße kaum ins Gewicht.

Die Tabellen D.17 und D.18 im Anhang A.4 zeigen die Armutsquoten gruppiert nach dem Merkmal Beruf. Die Haushalte mit (zum Zeitpunkt der Befragung) nicht-erwerbstätigem HV haben eine hohe bedingte Armutsquote. Da diese Gruppe sehr groß ist und die anderen Berufsgruppen im Vergleich relativ klein, macht die Gruppe der Haushalte mit nicht-erwerbstätigem HV einen sehr hohen Anteil der Gesamtarmutsquote aus.

Unter dem Gesichtspunkt des Merkmals Region ist die bedingte Armutsquote von Ostdeutschen oberhalb derjenigen der anderen Regionen (Abbildung 4.39(c)). Nach der Wende war es, vermutlich durch die Lohnkostenzuschüsse, gelungen, die bedingte Armut nahezu auf das Niveau der anderen Regionen zu verringern (Egle et al., 2003, S. 217). Ab dem Jahr 2000 erfolgte ein kontinuierlicher Anstieg von 9% auf 14%. Dies hat auch in der Gesamtbetrachtung zur Folge, dass die Armut von ostdeutschen Haushalten gestiegen ist und sogar die Armutsquote süddeutscher Haushalte, deren Stichprobe um etwa 15% (der Gesamtstichprobe) größer ist, überstieg.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die bedingte Armut von Haushalten mit direktem Migrationshintergrund ist seit 1992 von 10% auf fast 20% angestiegen (Abbildung 4.39(e)). Dies hat ebenfalls zu einem Anstieg der Armutsquote von Haushalten mit Migrationshintergrund in der Gesamtpopulation von 1% auf 2.5% geführt (Abbildung 4.39(f)). Es ist nicht überraschend, dass die Armutsquote von Haushalten ohne Migrationshintergrund sehr hoch und seit 1999 stark angestiegen ist, da diese Gruppe mit 85%-90% einen großen Anteil der Gesamtpopulation ausmacht. Die Information, dass bei den Haushalten ohne Migrationshintergrund die Armutsquote stark angestiegen ist, liefert wenig zusätzliche Erklärung zum Anstieg der gesamten Armutsquote. Vielmehr geben aber die Entwicklung, sowohl der Größe als auch der bedingten Armut, von Haushalten mit direktem Migrationshintergrund im geringe Maße eine Erklärung zum Anstieg der Armutsquote. Wie schon bei den Single-Haushalten ist auch bei den Haushalten mit Migrationshintergrund ein doppelter Effekt zu beobachten.

Abbildung 4.40(a) zeigt, in Übereinstimmung mit der Literatur (siehe z.B. Böhmer und Heimer (2008)), dass Haushalte mit vielen Kindern stärker von Armut betroffen sind als Haushalte ohne oder mit ein oder zwei Kindern. Da Großfamilien in Deutschland nur selten (durchschnittliche Geburtenrate 1.39 Kinder pro Frau im Alter 15-49) vorkommen, kann diese Gruppe kaum zur Erklärung der Gesamtarmutsquote beitragen (siehe Abbildung 4.40(b)).

Wie Abbildung 4.40(c) zeigt, sind Alleinerziehende eine der am stärksten von Armut betroffenen Gruppen in Deutschland. Die hohe bedingte Armut (siehe 4.40(d)) resultiert trotz der kleinen Gruppengröße in einer relativ hohen Armutsquote mit steigender Tendenz seit 1999. Es wird vermutet, dass der Anstieg des Anteils von Alleinerziehenden in der Bevölkerung ebenfalls zu einem Anstieg von Ungleichheit und Armut geführt haben könnte.

In diesem Abschnitt wurde untersucht, welche Strukturveränderungen in den Haushalten den Anstieg von Ungleichheit und Armut seit dem Jahr 2000 erklären könnten. Die obigen Darstellungen haben gezeigt, dass insbesondere die Zunahme von alleinstehenden und geschiedenen Haushalten sowie ein Wachstum der kleinen aber stark betroffenen Gruppen alleinerziehende Haushalte und Migranten dafür verantwortlich sein könnten. Weiterhin wurden aus den obigen Darstellungen Gruppen von Haushalten identifiziert, die in Deutschland besonders stark von Armut betroffen sind. Für Deutschland wurden bereits politische Maßnahmen ergriffen, um die Armut in diesen Gruppen zu verringern, wie etwa Regelungen zur Vereinbarkeit von Beruf und Familie.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.3.3 Möglichkeiten zur Vulnerabilitäts-Messung mit Makrovariablen

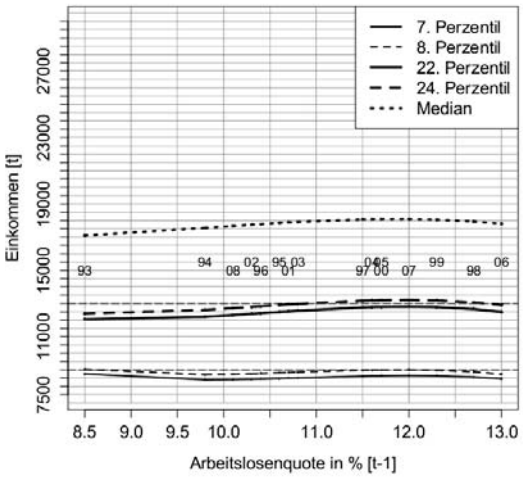
Um zu ermitteln, inwieweit Makrovariablen zur Schätzung der Vulnerabilität beitragen, werden die Einkommensniveaus in Abhängigkeit von den in Abschnitt 4.3.1 vorgestellten volkswirtschaftlichen Indikatoren betrachtet. In diesem Zusammenhang ist es von besonderem Interesse, wie stark die Einkommensniveaus in Nähe der Armutslinie bei 9 000 € auf Veränderungen der Indikatoren reagieren.

Es werden Quantilsregressionen, wie in Gleichung (4.6) beschrieben, durchgeführt. Dazu ist es notwendig, die Einkommensinformationen für alle betrachteten Jahre zu poolen. Es werden das 7. und 8. Perzentil (in etwa von Armut betroffene Haushalte), das 22. und 24. Perzentil (etwa in Höhe der VPL) sowie der Median betrachtet.

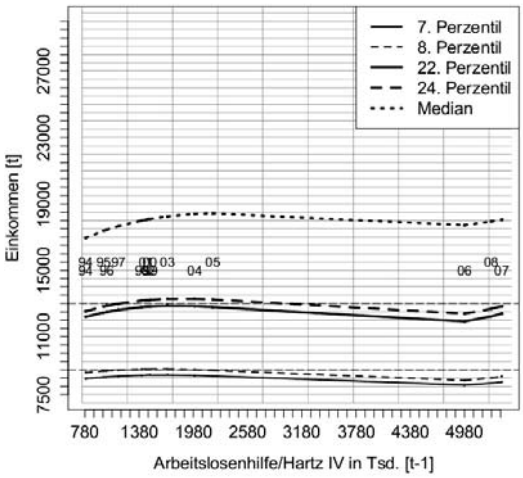
$$Q_\tau(y|x) = \psi_0 + \psi_1 x_{Indikator} \tag{4.6}$$

Die Quantilskurven sind in den Abbildungen 4.41 und 4.42 für den Beobachtungszeitraum dargestellt. Die Jahreszahlen über dem Wert der Makrovariablen (x-Achse) kennzeichnen den Wert der Makrovariablen, der im Vorjahr zu dem Wert der Quantilskurve in t geführt hat. Wie die Abbildungen zeigen, können die volkswirtschaftlichen Indikatoren der Vorperiode das Einkommen kaum erklären, insbesondere was die unteren Einkommensgruppen betrifft. Lediglich der Indikator Anzahl der Bezieher von Arbeitslosenhilfe/Hartz IV führt zu einer sichtbaren Veränderung der Quantilskurve für die unteren Quantile. Diese zeigt eine fallende Tendenz, d.h. ein sinkendes Einkommen für die unteren Einkommensgruppen bei steigender Anzahl der Arbeitslosenhilfe- bzw. Hartz IV-Bezieher. Ein Problem stellen jedoch wieder Trends und Zyklen dar. So ist etwa die Anzahl der Bezieher von 2005 zu 2006 stark gestiegen und ist dann in 2007 und 2008 in etwa konstant. Wäre der Indikator in 2009 stark gefallen (z.B. aufgrund von Reformen), so hätte für 2009 ein niedrigeres Einkommen für die unteren Einkommensgruppen prognostiziert werden müssen.

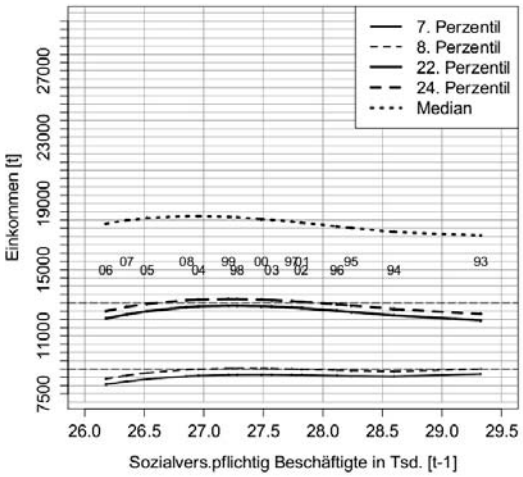
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



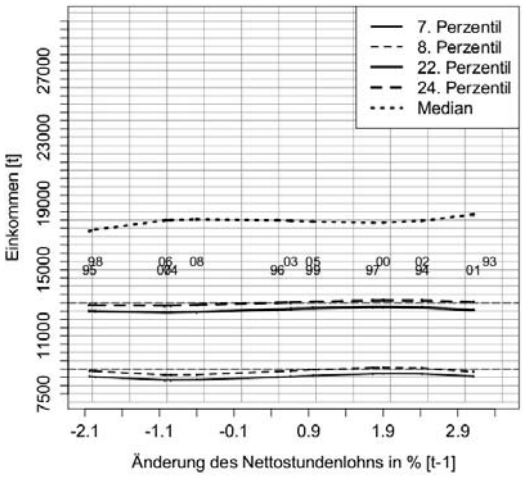
(a) Arbeitslosenquote



(b) Arbeitslosenhilfe/ Hartz IV



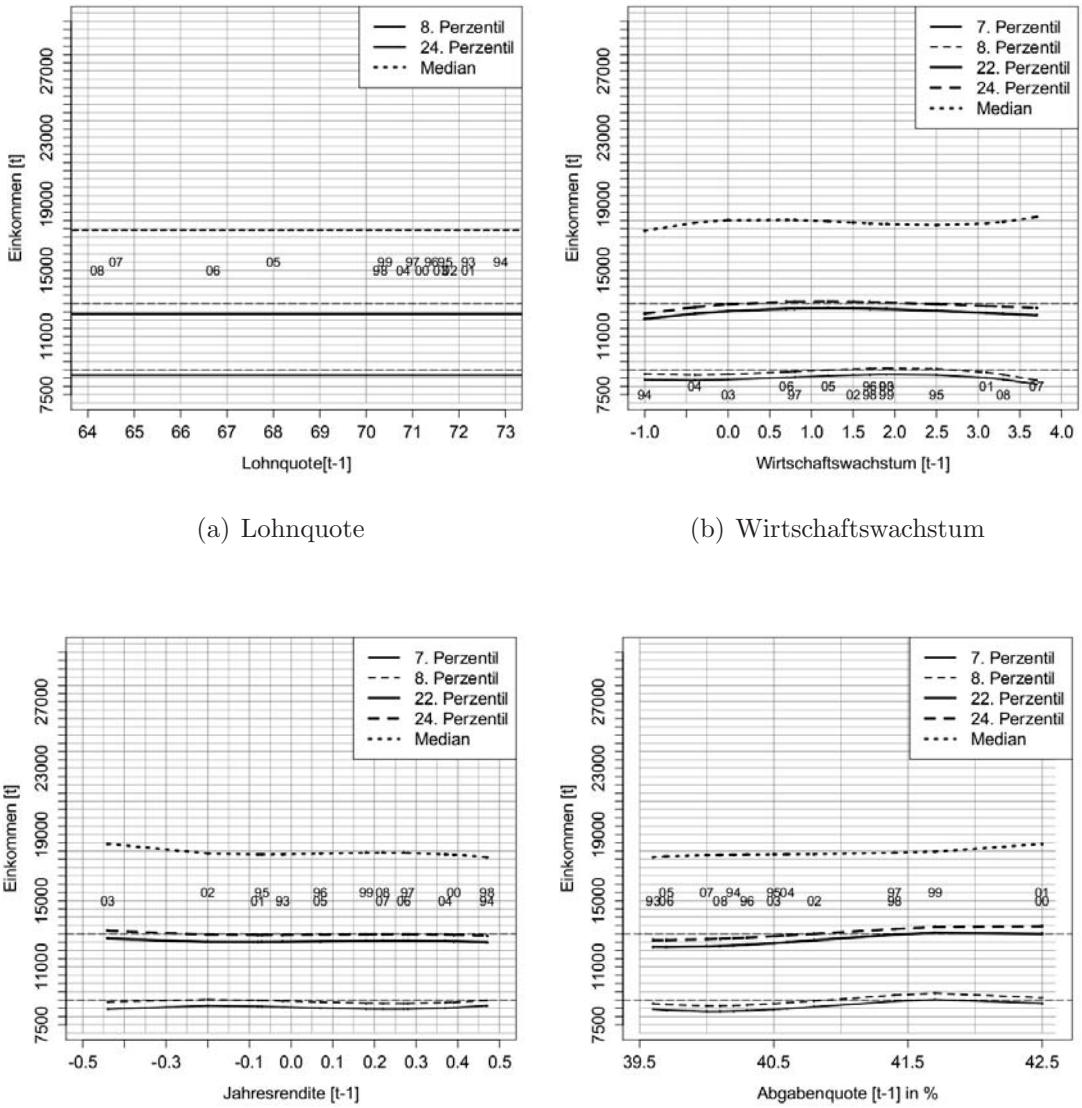
(c) Sozialversicherungspflichtig Beschäftigte



(d) Veränderung des Nettostundenlohns in %

Abbildung 4.41: *Erklärungsgehalt des Einkommens im Jahr t durch Arbeitsmarktindikatoren des Vorjahres für verschiedene Quantile (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten, Bundesagentur für Arbeit (2011), Statistisches Bundesamt (2010)]. Die Punkte werden durch die letzten beiden Ziffern des Jahres t dargestellt.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Lohnquote

(b) Wirtschaftswachstum

(c) Dax-Renditen

(d) Abgabenquote

Abbildung 4.42: *Erklärungsgehalt des Einkommens im Jahr t durch Arbeitsmarktindikatoren des Vorjahres für verschiedene Quantile (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten, Bundesagentur für Arbeit (2011), Statistisches Bundesamt (2010)]. Die Punkte werden durch die letzten beiden Ziffern des Jahres t dargestellt.*

Zusammenfassend kann gesagt werden, dass weder eine Möglichkeit zur Prognose von Armut noch eine Verbesserung der Schätzung der Vulnerabilität mit einzelnen makroökonomischen Kovariablen gefunden wurde.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Für die Zukunft wäre es sicher interessant zu analysieren, wie sich die Veränderung von volkswirtschaftlichen Indikatoren auf die Vulnerabilität von gewissen Gruppen von Haushalten auswirkt, wie etwa die Entwicklung der Arbeitslosigkeit für Haushalte nach Art des Familienstands. Um dies zu untersuchen, könnten in die verwendeten Modelle P_1-P_6 Makrovariablen in Interaktion mit Haushaltsvariablen ähnlich zum Vorgehen von den Autoren Christiaensen und Subbarao (2005) aufgenommen werden. Mit den erweiterten Modellen könnte dann wieder die Genauigkeit der Vulnerabilitäts-Schätzer bewertet werden.

4.3.4 Messung der Vulnerabilität von Gruppen

In diesem Abschnitt wird die Genauigkeit der Vulnerabilitäts-Schätzer mit zwei verschiedenen Methoden und den Modellen aus Abschnitt 2.4.3 geschätzt. Anders als im Abschnitt 4.1 wird in den Modellen $P_{1,G}$ und $P_{6,G}$ dieses Abschnittes eine logarithmische Transformation des Einkommens verwendet (sowohl für den Response als auch für die Kovariable Vorjahreseinkommen). In Abschnitt 4.1 war die Begründung gegen die Transformation, dass die Einkommensverteilung bei 30 000 € „abgeschnitten“ wurde und diese nicht als logarithmisch normalverteilt angesehen werden konnte. Keine der untersuchten Verteilungen (Normal-, Lognormal-, t-, Gamma- und Weibullverteilung) führte zu normalverteilten Residuen bzw. normalverteilten Quantilsresiduen. In diesem Abschnitt wird das Einkommen nicht durch 30 000 € sondern durch die Gruppenzugehörigkeit eingegrenzt. Als Verteilungsannahmen für die Regressionsmodelle wurden sowohl eine Normalverteilung als auch eine logarithmische Normalverteilung (bzw. Normalverteilung des transformierten Responses) untersucht. Die Regressionen mit untransformiertem Response führten zu einigen extrem hohen Residuen. Die Histogramme der Residuen sind in Anhang A.4.3 exemplarisch für das Merkmal Geschlecht dargestellt. Weder die untransformierten noch die transformierten Regressionsmodelle führten zu normalverteilten Residuen (Kurtosis größer als drei und nicht symmetrisch). Da die Transformation, insbesondere bei der Schätzung mit Modell $P_{1,G}$, zu genaueren Vulnerabilitäts-Schätzern führt, wird in diesem Abschnitt das transformierte Einkommen verwendet. Auch bei der Schätzung mit Modell $P_{6,G}$ führt die Transformation in den meisten Fällen zu einer leichten Verbesserung.

In Methode 1 wird das logarithmische Einkommen für jede Gruppe (Ausprägung eines Merkmals) getrennt vorhergesagt, während die ROC basierend auf allen prognostizierten (logarithmischen) Einkommen konstruiert wird und die Vulnerabilität somit gemeinsam geschätzt wird. Im Gegensatz dazu wird in Methode 2 sowohl

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

das logarithmische Einkommen als auch die ROC in den Gruppen getrennt erzeugt, d.h., die Vulnerabilität wird für die Gruppen einzeln bei 80% TPR in der Gruppe geschätzt.

Die verwendeten Kovariablen zur Schätzung des (logarithmischen) Einkommens mit $P_{1,G}$ und $P_{6,G}$ variieren von Merkmal zu Merkmal. Vereinfachend wird für die Modelle der Ausprägungen eines Merkmals dieselben Kovariablen verwendet, d.h., dass z.B. das (logarithmische) Einkommen in der Gruppe der Haushalte mit männlichem HV mit denselben Kovariablen geschätzt wird wie jenes der Haushalte mit weiblichem HV. Es ist offensichtlich, dass ein Merkmal, nach dem gruppiert wurde, keine Kovariable der jeweiligen Regressionsmodelle ist. Weiterhin wurden z.T. einige in den beiden vorangehenden Abschnitten gewählte Kovariablen nicht verwendet, da sie entweder zur Schätzung in den Gruppen nicht relevant waren oder das betrachtete Merkmal Ausprägungen mit sehr kleinen Stichprobenumfängen beinhaltete, so dass nicht so viele Variablen berücksichtigt werden konnten. In letzterem Fall wurden auch Stufen von Faktoren zusammengelegt. Die Schätzungen für Haushalte mit jungem HV und die alleinerziehenden Haushalte sind trotzdem mit Vorsicht zu betrachten, da Stichproben dieser beiden Gruppen sehr klein sind. Informationen über die verwendeten Kovariablen in den Gruppen werden im Anhang A.4.2 gegeben. In den Tabellen D.20-D.23 des Anhangs sind exemplarisch Details der Regressionen dargestellt. Die Regressionstabellen der übrigen Gruppen sind vom Autor erhältlich.

Die Regressionskoeffizienten können aufgrund der Transformation nicht direkt mit den Tabellen aus 4.1 verglichen werden. Die Regressionskoeffizienten müssen multiplikativ interpretiert werden, d.h., ein Anstieg von x_1 um eine Einheit multipliziert den erwarteten Response (in der untransformierten Skalierung) mit $\exp(\beta)$ (Faraway, 2002, S. 95f.). Den Regressionstabellen D.21 und D.20 wird entnommen, dass alle Faktoren (d.h. Bundesland, Familienstand, Gewerbe, Wohnverhältnis und Bildung) eine Wirkung haben. Die Effekte der Faktorstufen unterscheiden sich zwischen den beiden Gruppen:

- Für beide Gruppen ist die Ansässigkeit in ostdeutschen Bundesländern gegenüber der Referenzgruppe West-Berlin negativ korreliert mit dem Einkommen (insbesondere bei den Frauen für Mecklenburg Vorpommern).
- Für die Gruppe der Haushalte mit weiblichem HV wirken Scheidung und Trennung gegenüber der Referenzgruppe Lebensgemeinschaft stärker negativ als für Haushalte mit männlichem HV (mögliche Gründe: Frauen verdienen weniger Geld, arbeiten weniger oder gar nicht zur Erziehung von Kindern).

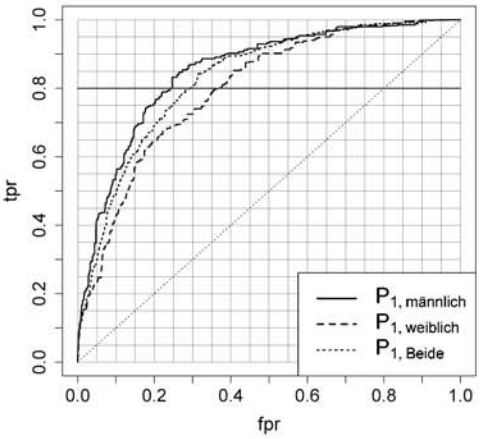
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

- Bzgl. des Faktors Gewerbe gibt es bei den weiblich geführten Haushalten gegenüber der Referenzgruppe keine Erwerbstätigkeit größere positive Effekte als bei den männlich geführten Haushalten (etwa die Beschäftigung in einer Bank/Versicherung, dem Baugewerbe oder dem Dienstleistungssektor).
- Das Wohnen zur Miete wirkt sich bei beiden Gruppen leicht negativ gegenüber der Referenzgruppe Hausbesitz aus.
- Ein hohes Bildungsniveau (Abschluss Abitur) hat bei den weiblichen HVs eine höhere (positive) Korrelation zum Einkommen als bei männlichen HVs.
- Der Anstieg der Haushaltsmitglieder hat bei beiden Gruppen in den Altersklassen eine negative Korrelation zum Einkommen mit Ausnahme der Altergruppe „älter als 59“ bei den weiblichen HVs.
- In der Gruppe der Haushalte mit männlichem HV ist das Einkommen höher (positiv) korreliert mit der Anzahl der Vollzeitbeschäftigten als bei Haushalten mit weiblichem HV.
- Die Korrelation zwischen Teilzeitbeschäftigung und Einkommen ist in der Gruppe der Haushalte mit weiblichem HV leicht negativ.
- Die Größe der Wohnung wirkt sich bei beiden Gruppen positiv auf das Einkommen aus.

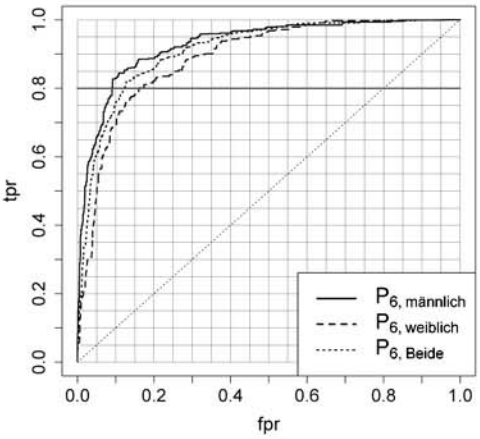
Wird zusätzlich das logarithmische Vorjahreseinkommen hinzugezogen (siehe Tabellen D.22 und D.23), so wird der Erklärungsgehalt der Haushaltskovariablen kleiner, während die Korrelation zwischen Vorjahreseinkommen und aktuellem Einkommen hoch ist. Die Bestimmtheitsmaße von den Modellen D.20 und D.21 einerseits und D.22 und D.23 andererseits sind sehr ähnlich. Da sich die Schätzung der Regressionskoeffizienten für die Gruppierung nach dem Merkmal Geschlecht unterscheidet, könnte die Gruppierung zu einer Verbesserung der Vulnerabilitäts-Schätzer basierend auf P_1 führen.

Die Abbildungen 4.43-4.45 und die Tabellen 4.20-4.22 bilden die Ergebnisse der Schätzung der Genauigkeit für 2005 ab. Mit den ROC-Kurven wird die gesamte Spanne von Paaren aus FPRs und TPRs der beiden Methoden gezeigt. In der Legende der Abbildungen werden die aus der Schätzung mit Methode 1 erzeugten ROCs mit $P_{1,Alle}$ bzw. $P_{6,Alle}$ (im Falle von zwei Gruppen $P_{1,Beide}$ bzw. $P_{6,Beide}$) bezeichnet, während die mit Methode 2 erzeugten Kurven mit $P_{1,Gruppenname}$ bzw. $P_{6,Gruppenname}$ bezeichnet werden.

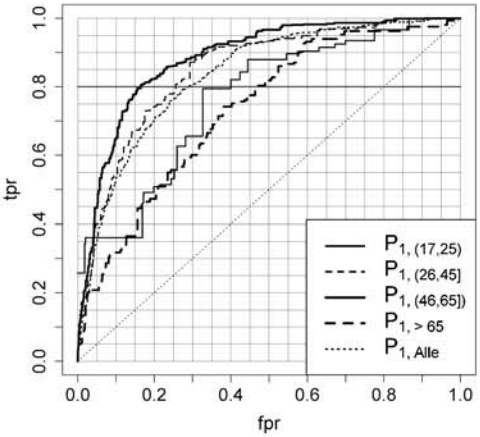
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



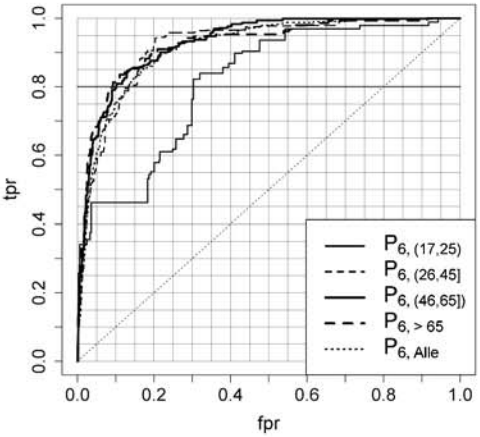
(a) ROC nach Geschlecht ($P_{1,G}$)



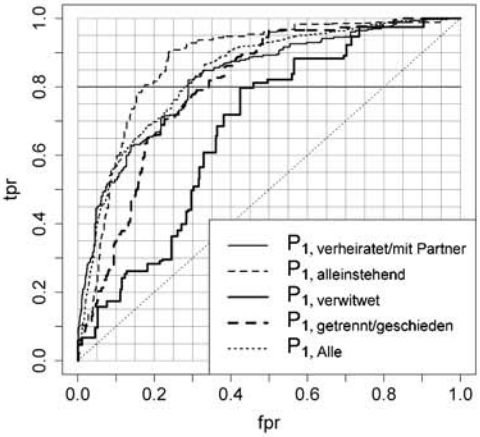
(b) ROC nach Geschlecht ($P_{6,G}$)



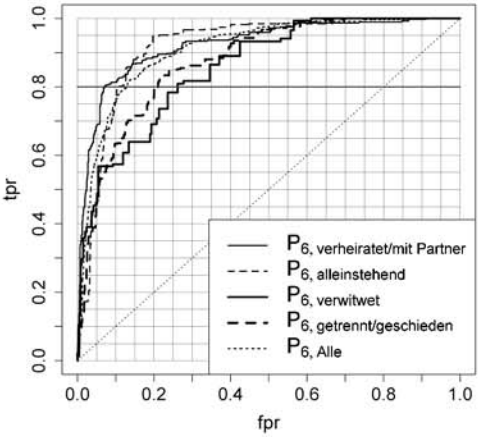
(c) ROC nach Alter ($P_{1,G}$)



(d) ROC nach Alter ($P_{6,G}$)



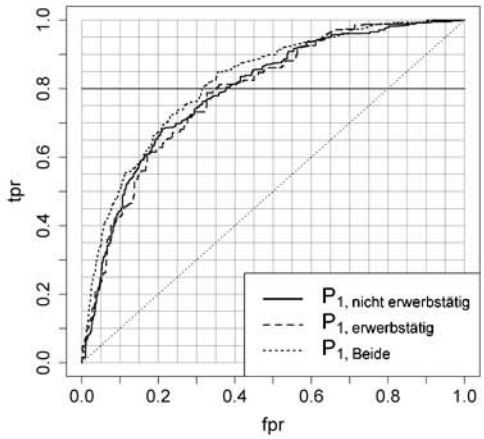
(e) ROC nach Familienstand ($P_{1,G}$)



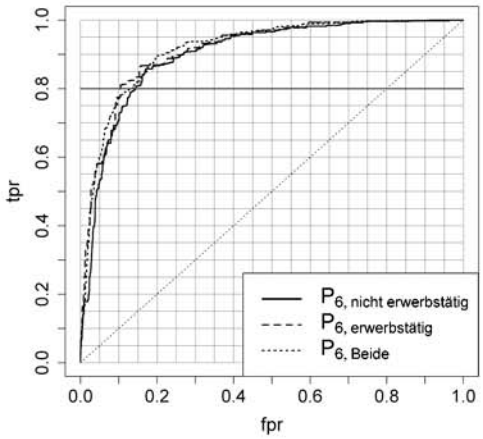
(f) ROC nach Familienstand ($P_{6,G}$)

Abbildung 4.43: *Genauigkeit der Schätzung der Vulnerabilität getrennt nach Merkmalen (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

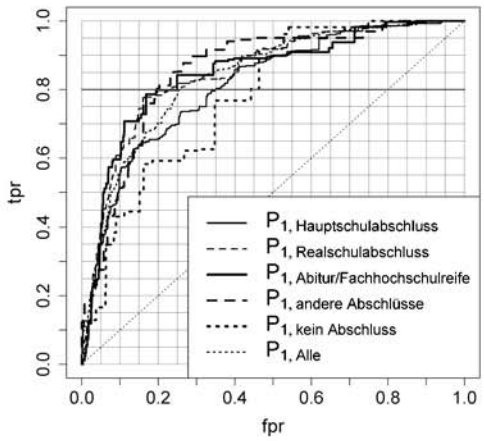
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



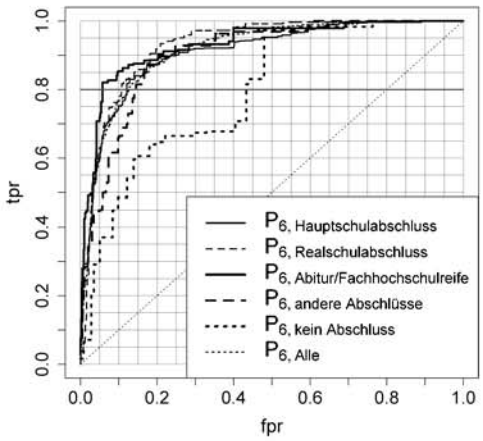
(a) ROC nach Status Erwerbstätigkeit ($P_{1,G}$)



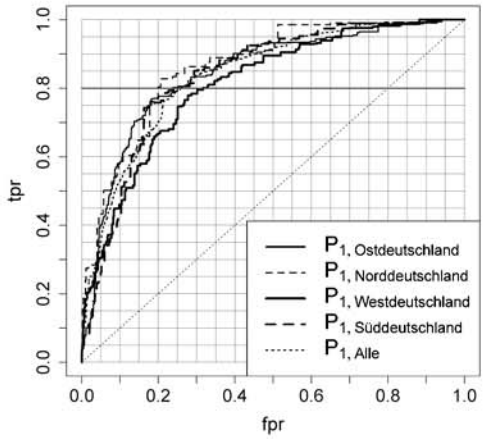
(b) ROC nach Status Erwerbstätigkeit ($P_{6,G}$)



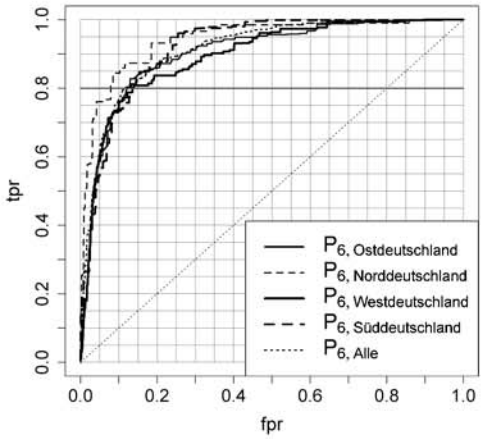
(c) ROC nach Bildung ($P_{1,G}$)



(d) ROC nach Bildung ($P_{6,G}$)



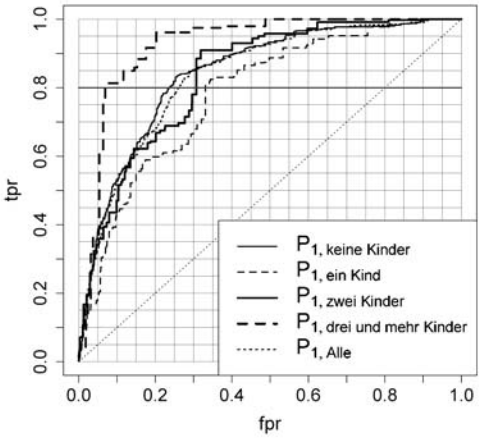
(e) ROC nach Region ($P_{1,G}$)



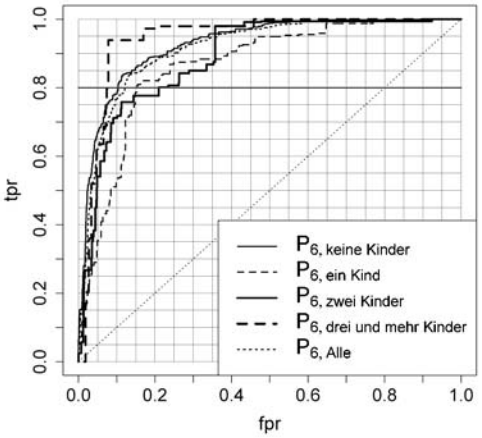
(f) ROC nach Region ($P_{6,G}$)

Abbildung 4.44: *Fortsetzung 1 Genauigkeit der Schätzung der Vulnerabilität getrennt nach Merkmalen (1994-2008) [eigene Darstellung auf Basis der in Kapitel 2 beschriebenen Daten]*

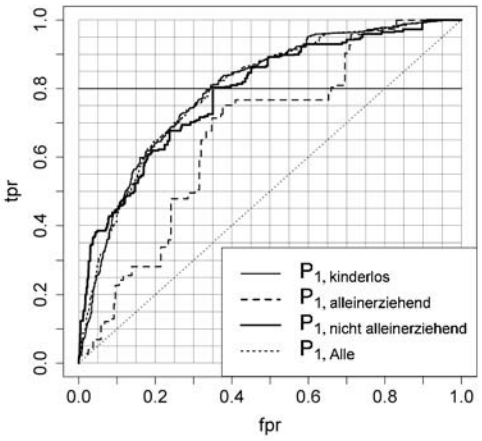
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



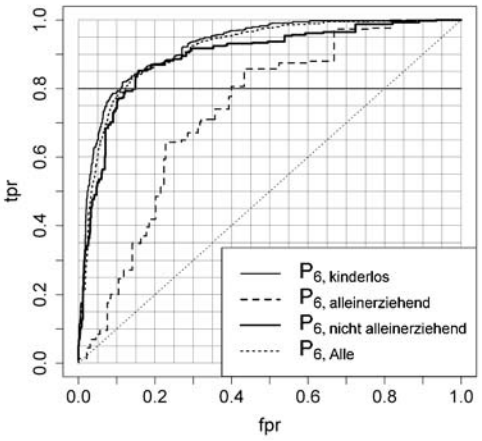
(a) ROC nach Anzahl der Kinder ($P_{1,G}$)



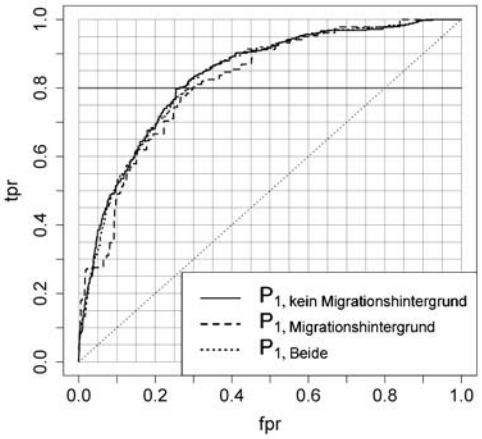
(b) ROC nach Anzahl der Kinder ($P_{6,G}$)



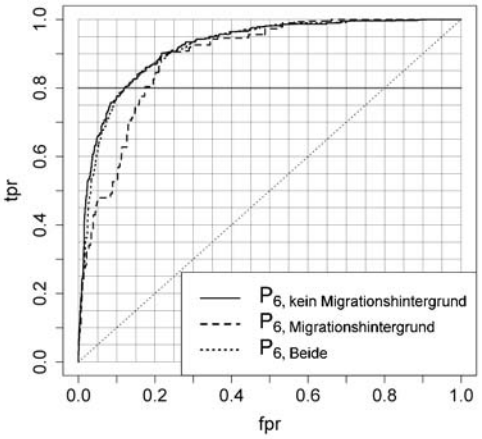
(c) ROC nach Erziehung ($P_{1,G}$)



(d) ROC nach Erziehung ($P_{6,G}$)



(e) ROC nach Migrationshintergrund ($P_{1,G}$)



(f) ROC nach Migrationshintergrund ($P_{6,G}$)

Abbildung 4.45: *Fortsetzung 2 Genauigkeit der Schätzung der Vulnerabilität getrennt nach Merkmalen (1994-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Es ist zu beobachten, dass sich für einige Merkmale die ROCs der Gruppen stark unterscheiden, (z.B. für die Merkmale Familienstand und Anzahl der Kinder im Haushalt), während sie für andere Merkmale relativ ähnlich sind (z.B. Erwerbstätigkeit und Migrationshintergrund). Die meisten Gruppen zeigen für die Schätzung unter Verwendung des (logarithmischen) Vorjahreseinkommens „bessere“, d.h. schneller in Richtung des y-Werts 1 verlaufende ROC-Kurven als nur mit Haushaltskovariablen. Eine Ausnahme bildet die Gruppe „drei oder mehr Kinder im Haushalt“, in der sich die Vulnerabilität mit Querschnittsdaten und Haushaltskovariablen sehr gut schätzen lässt. Weitere relativ gute Schätzer können mit dem Modell $P_{1,G}$ für folgende Gruppen erzielt werden: Haushalte mit alleinstehendem HV, Haushalte mit HV mit Realschul- oder anderen Schulabschlüssen und Haushalte, die in Norddeutschland ansässig sind. Für (bedingte) Vulnerabilitäts-Schätzer basierend auf $P_{6,G}$ ergeben sich dagegen eine ganze Reihe von genauen Vulnerabilitäts-Schätzern (z.B. für die Gruppe der Haushalte mit männlichem HV, mit HV älter als 65 und den kinderlosen Haushalten). Es ist jedoch häufig zu beobachten, dass die Vulnerabilität für andere Ausprägungen dieser Merkmale ungenauer geschätzt wird (z.B. für die Gruppe der Haushalte mit weiblichem HV, westdeutschen Haushalten oder der Haushalte mit einem Kind).

In den Tabellen 4.20-4.21 werden jeweils zwei verschiedene Maße für die Genauigkeit der Vulnerabilitäts-Schätzer der Gruppen (Methode 2) angegeben. Das erste Maß bezieht sich auf die Messung der Genauigkeit in den Gruppen und wird mit FPR(*Gruppenname*) bezeichnet. Es gibt an, wie viele nicht-arme Haushalte in einer bestimmten Gruppe als armutsgefährdet klassifiziert werden müssen, damit 80% der armen Haushalte in dieser Gruppe richtig identifiziert werden. Z.B. müssen 9% der nicht-armen Haushalte mit männlichem HV als armutsgefährdet geschätzt werden, wenn in dieser Gruppe 80% der armen Haushalte richtig identifiziert werden sollen und das Modell $P_{6,G}$ zur Schätzung verwendet wird. Ebenso müssen 16% der nicht-armen Haushalte mit weiblichem HV als armutsgefährdet geschätzt werden, wenn in dieser Gruppe 80% der armen Haushalte richtig identifiziert werden sollen.

Das zweite Maß gibt an, welcher Anteil der nicht-armen Haushalte einer Gruppe an den nicht-armen Haushalten der Gesamtpopulation als armutsgefährdet klassifiziert werden muss, wenn 80% der armen Haushalte in dieser Gruppe richtig identifiziert werden sollen. Das Maß wird mit $FPR_{gesamt}(Gruppenname)$ bezeichnet. Z.B. werden in der Gruppe der Haushalte mit männlichem HV etwa 5% aller nicht-armen Haushalte der Gesamtpopulation falsch als armutsgefährdet klassifiziert, während in der Gruppe der Haushalte mit weiblichem HV etwa 7% aller nicht-armen Haushalte falsch als armutsgefährdet klassifiziert werden, um in jeder

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Gruppe 80% der armen Haushalte richtig zu identifizieren. In der Gesamtbetrachtung ergibt sich für das Merkmal Geschlecht mit Methode 2 eine FPR von etwa 12%. Diese Summe der FPRs wird in Tabelle 4.22 mit dem FPRs aus Methode 1 verglichen.

Es zeigt sich, dass die beiden Methoden zu ähnlich genauen Vulnerabilitäts-Schätzern führen. Eine Ausnahme bildet die Schätzung für das Merkmal Erwerbstätigkeit mit Modell $P_{1,G}$. Methode 2 liefert in diesem Fall ungenauere Ergebnisse als die Methode 1. Ursache dafür ist, dass das prognostizierte Einkommen in der Gruppe der Haushalte mit erwerbstätigem HV höher ist als in der Gruppe mit nicht-erwerbstätigem HV. Das ist darauf zurückzuführen, dass in der Gruppe der Haushalte mit erwerbstätigem HV weniger arm sind. Da die Variable Gewerbe unberücksicht ist, wird auch das Einkommen weniger wohlhabender Haushalte höher geschätzt. Daher muss in der Gruppe mit erwerbstätigem HV die VPL deutlich höher gewählt werden, um 80% der Armen zu identifizieren als in der Gruppe mit nicht-erwerbstätigem HV. Wenn die Vulnerabilität bei 80% TPR in jeder Gruppe geschätzt wird, müssen mehr false positives in Kauf genommen werden als wenn eine VPL „dazwischen“ gewählt wird, die dazu führt, dass insgesamt (d.h. aus beiden Gruppen zusammen) 80% TPR erzielt werden. Vergleicht man die Genauigkeit der Vulnerabilitäts-Schätzer in Gruppen mit denen aller Haushalte bei 80% TPR für das Jahr 2005, so ist festzustellen, dass sich die Genauigkeit der Vulnerabilität durch die Schätzung in Gruppen für die meisten Merkmale und mit beiden Methoden verbessern lassen. Das gilt insbesondere für die Schätzung mit Querschnittsdaten und die Gruppen *Anzahl der Kinder im Haushalt*, *Bildung* und *Region*. Die Verbesserung der Schätzer von $P_{6,G}$ im Vergleich zu denen von P_6 ist relativ gering.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	mit Gruppen	
	P_1	P_6
FPR(männlich)	25	9
FPR(weiblich)	37	16
FPR_{gesamt}(männlich)	14	5
FPR_{gesamt}(weiblich)	16	7
FPR(Alter $(17, 25]$)	40	30
FPR(Alter $(26, 45]$)	26	13
FPR(Alter $(46, 65]$)	16	10
FPR(Alter > 65)	47	9
FPR_{gesamt}(Alter $(17, 25]$)	1	1
FPR_{gesamt}(Alter $(26, 45]$)	9	5
FPR_{gesamt}(Alter $(46, 65]$)	5	3
FPR_{gesamt}(Alter> 65)	13	3
FPR(verheiratet/mit Partner)	29	7
FPR(alleinstehend)	17	10
FPR(verwitwet)	46	26
FPR(geschieden/getrennt)	34	21
FPR_{gesamt}(verheiratet/mit Partner)	15	4
FPR_{gesamt}(alleinstehend)	4	2
FPR_{gesamt}(verwitwet)	7	4
FPR_{gesamt}(geschieden/getrennt)	5	3
FPR(nicht erwerbstätig)	38	14
FPR(erwerbstätig)	35	10
FPR_{gesamt}(nicht erwerbstätig)	17	6
FPR_{gesamt}(erwerbstätig)	20	6
FPR(Hauptschulabschluss)	35	12
FPR(Realschulabschluss)	23	11
FPR(Abitur/Fachhochschulreife)	25	6
FPR(andere Abschlüsse)	20	14
FPR(kein Abschluss)	44	43
FPR_{gesamt}(Hauptschulabschluss)	14	5
FPR_{gesamt}(Realschulabschluss)	6	3
FPR_{gesamt}(Abitur(Fachhochschulreife)	6	1
FPR_{gesamt}(andere Abschlüsse)	1	1
FPR_{gesamt}(kein Abschluss)	1	1

Tabelle 4.20: *Genauigkeit der Schätzung der Vulnerabilität mit Methode 2 für 2005 in % (FPR bei 80% TPR)[eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Je kleiner die FPR, desto genauer ist die Schätzung.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	mit Gruppen	
	P_1	P_6
FPR(Ostdeutschland)	25	12
FPR(Norddeutschland)	20	8
FPR(Westdeutschland)	32	14
FPR(Süddeutschland)	24	13
FPR_{gesamt}(Ostdeutschland)	5	3
FPR_{gesamt}(Norddeutschland)	3	1
FPR_{gesamt}(Westdeutschland)	9	4
FPR_{gesamt}(Süddeutschland)	8	6
FPR(keine Kinder)	24	10
FPR(ein Kind)	33	15
FPR(zwei Kinder)	31	21
FPR(drei und mehr Kinder)	7	8
FPR_{gesamt}(keine Kinder)	18	8
FPR_{gesamt}(ein Kind)	4	2
FPR_{gesamt}(zwei Kinder)	3	2
FPR_{gesamt}(drei und mehr Kinder)	0	0
FPR(keine Kinder)	34	11
FPR(alleinerziehend)	66	40
FPR(nicht allein erziehend)	35	15
FPR_{gesamt}(keine Kinder)	26	8
FPR_{gesamt}(alleinerziend)	2	1
FPR_{gesamt}(nicht alleinerziehend)	7	3
FPR(ohne Migrationshintergrund)	26	12
FPR(mit Migrationshintergrund)	30	17
FPR_{gesamt}(ohne Migrationshintergrund)	23	11
FPR_{gesamt}(mit Migrationshintergrund)	4	2

Tabelle 4.21: *Fortsetzung Genauigkeit der Schätzung der Vulnerabilität mit Methode 2 für 2005 in % (FPR bei 80% TPR)[eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Je kleiner die FPR, desto genauer ist die Schätzung.*

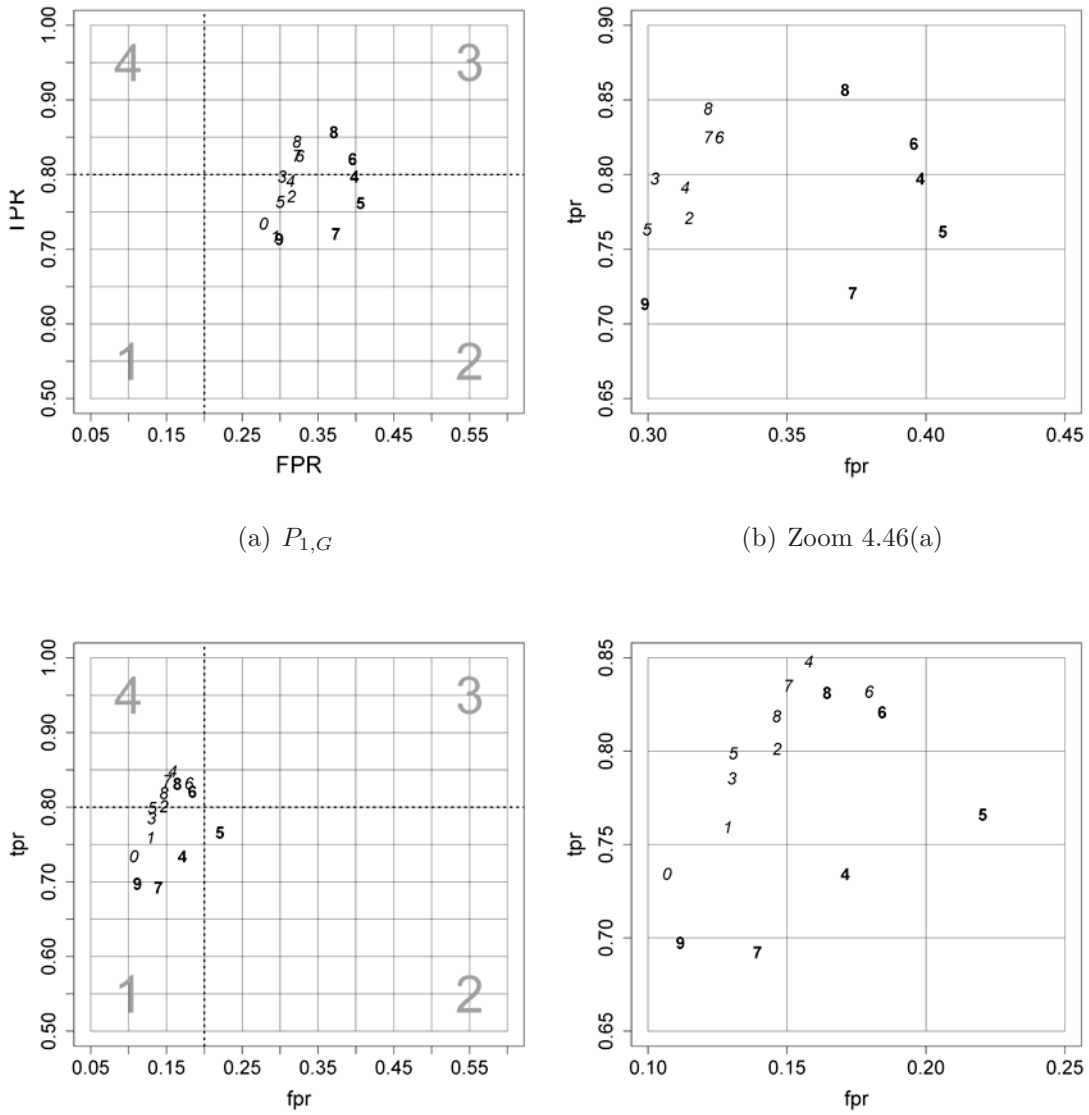
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Modell			
Merkmal	$P_{1,Gruppe}$		$P_{6,Gruppe}$	
	Methode 1	Methode 2	Methode 1	Methode 2
Geschlecht	29	30	12	12
Alter	29	28	13	12
Familienstand	28	30	13	13
Erwerbstätigkeit	31	37	13	12
Bildung	26	28	13	11
Region	26	25	12	13
Anzahl Kinder	26	25	12	12
Erziehung	35	35	12	12
Migrationshintergrund	29	27	12	13
alle Haushalte	**33**		**14**	

Tabelle 4.22: *Vergleich der Genauigkeit der Vulnerabilitäts-Schätzer von Methode 1, Methode 2 (Gesamtbetrachtung) sowie P_1 und P_6 für 2005 in % (FPR bei 80% TPR) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Je kleiner die FPR, desto genauer ist die Schätzung.*

Im Folgenden wird exemplarisch die Genauigkeit der Vulnerabilität mit festen VPLs für die Gruppen Erwerbstätigkeit (Methode 1) und Anzahl der Kinder (Methode 2) über den Zeitraum 1994-2008 bewertet. Die VPLs werden wieder aufgrund der historischen VPLs, die zu einer TPR (bei Methode 1 insgesamt, bei Methode 2 pro Gruppe) von 80% geführt hätten, festgelegt und sind im Anhang A.4.5 abgebildet.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) $P_{1,G}$

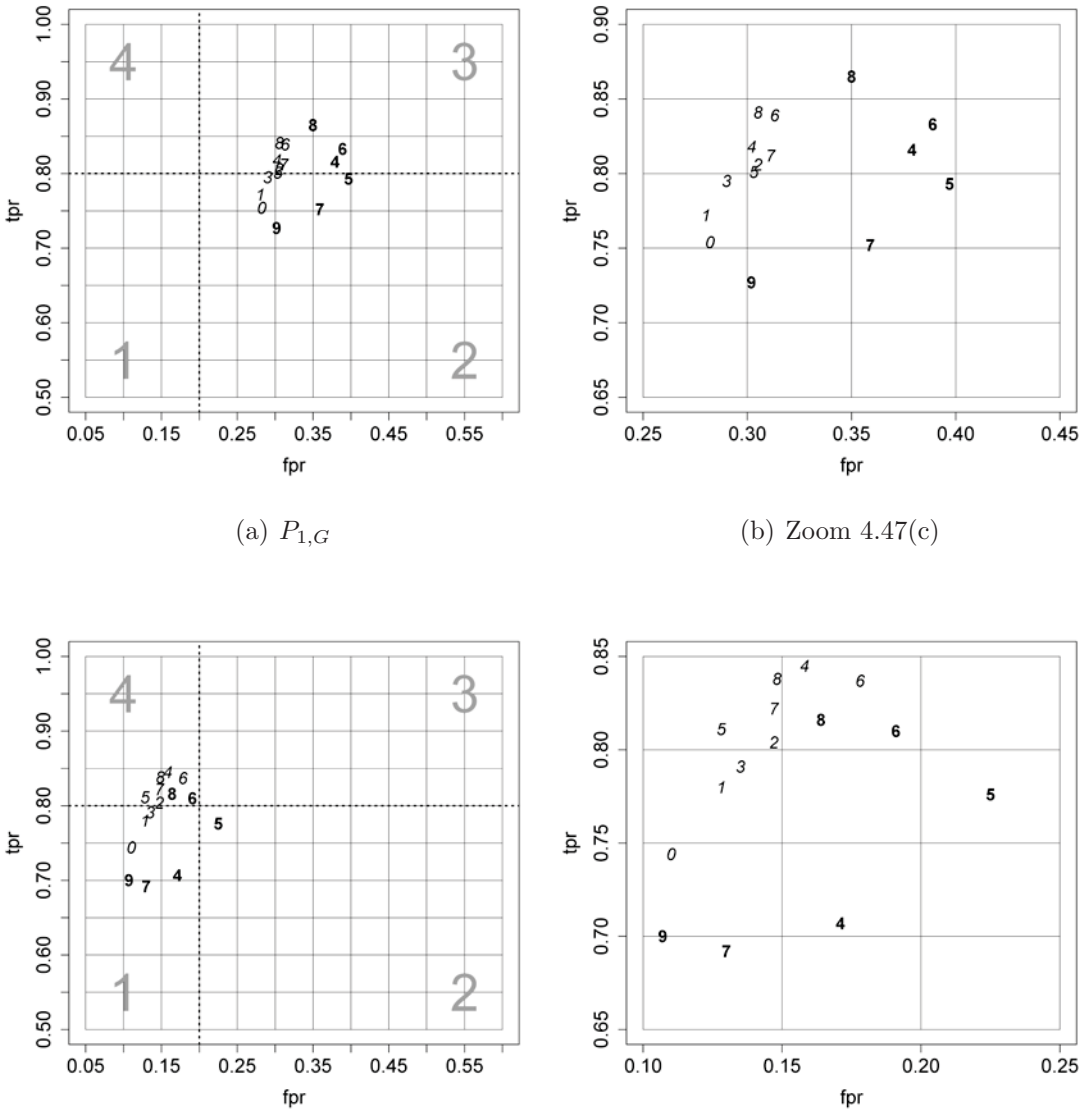
(b) Zoom 4.46(a)

(c) $P_{6,G}$

(d) Zoom 4.46(c)

Abbildung 4.46: *Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Erwerbstätigkeit für 1994-2008 (Methode 1) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) $P_{1,G}$

(b) Zoom 4.47(c)

(c) $P_{6,G}$

(d) Zoom 4.47(c)

Abbildung 4.47: *Gesamtbetrachtung der Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Anzahl der Kinder im Haushalt für 1994-2008 (Methode 2) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Für die Schätzung der Vulnerabilität nach dem Merkmal Erwerbstätigkeit werden die VPLs bei 9.7, entspricht 16 317 € (Schätzungen mit $P_{1,G}$), und 9.5, entspricht etwa 13 360 € (Schätzungen mit $P_{6,G}$), fixiert. Die Schätzer, die sich mit diesen VPLs im Zeitraum 1994-2008 ergeben, sind in Abbildung 4.46 dargestellt. Die

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Stabilität der Genauigkeit scheint mit durchgeführten Schätzungen für Gruppen „nicht-erwerbstätiger HV“ und „erwerbstätiger HV“ im Vergleich zu der Schätzung für alle Haushalte ähnlich zu sein. Die Schätzung für genannte Gruppen scheint bei Querschnittsdaten etwas besser zu funktionieren, während für Paneldaten die Schätzungen basierend auf der Stichprobe aller Haushalte etwas genauer sind. Wird die Vulnerabilität in Gruppen nach der Anzahl der Kinder im Haushalt geschätzt (Methode 2), so ergeben sich sogar mit Querschnittsdaten und mit Paneldaten genauere Schätzer als für alle Haushalte basierend auf P_1 und P_6. Die Gesamtbetrachtung von Methode 2 für die Gruppierung nach Kindern ist in Abbildung 4.47 dargestellt. Informationen zu den verwendeten VPLs sowie eine Einzeldarstellung der Genauigkeit der einzelnen Gruppen sind im Anhang in den Abbildungen D.14 und D.15 dargestellt. Es wird vermutet, dass die starken Schwankungen der VPL der Gruppe „3 Kinder im Haushalt“ auf die kleine Stichprobengröße zurückzuführen ist.[21]

Zusammenfassend kann gesagt werden, dass es durch die Schätzung der Vulnerabilität in Gruppen für viele Merkmale möglich ist, die Schätzer aus 4.1 etwas zu verbessern. Wenn es gelingt, die VPL stabiler zu wählen, so ist möglicherweise Potential für weitere Verbesserungen vorhanden.

4.3.5 Zwischenfazit

In diesem Abschnitt wurden mit dem SOEP, Makroindikatoren der Bundesagentur für Arbeit und des Statistischen Bundesamtes sowie den Modellen aus Abschnitt 2.4.3 Möglichkeiten untersucht, die Schätzer der Vulnerabilität aus Abschnitt 4.1 spezifisch für Deutschland zu verbessern. Vorausgegangen war eine detaillierte Diskussion über Ursachen der Entwicklung der Armutsquoten im Zeitraum 1992-2008 auf der Haushalts- und auf der Makroebene. Die Idee, die dahinter steckte, war, dass bisher explizit nicht berücksichtigte, volkswirtschaftliche Veränderungen (sowohl strukturelle Veränderungen als auch Schocks) und Veränderungen in der

[21] Um die Vulnerabilität in Gruppen nach dem Merkmal der Anzahl der Kinder über die Zeit schätzen zu können, mussten (aufgrund der kleinen Stichprobengröße der Gruppe „3 Kinder“) die Kovariablen im Vergleich zur Schätzung für 2005 leicht verändert werden. Der Faktor Eigentümer des Wohnsitzes wurde von drei auf zwei Stufen reduziert (Besitzer und Mieter); der Faktor Familienstand wurde auf drei Stufen (verheiratet, ledig, getrennt) reduziert, wobei getrennt die Stufen geschieden, getrennt und verwitwet zusammengefasst werden; bei dem Faktor Gewerbe des HVs wurde der Sektor Energie mit Dienstleistungen zusammengefasst. Dadurch ergaben sich nur geringfügige Veränderungen.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Haushaltsstruktur (demografischer Wandel, Wandel in der Haushaltsstruktur) die Vulnerabilität der Haushalte beeinflussen. Dazu wurde zum einen untersucht, ob sich das Einkommen durch einzelne makroökonomische Indikatoren erklären bzw. voraussagen lässt. Zum anderen wurde als Verbesserungsmöglichkeit analysiert, wie genau sich die Vulnerabilität in Gruppen als Ausprägungen von Merkmalen des Haushaltes schätzen lässt.

Die Ergebnisse der Analysen mit Makrovariablen zeigen, dass die verwendeten Ansätze nicht zur Verbesserung der Schätzer der Vulnerabilität genutzt werden können. Es war nicht zu beobachten, dass die Makroindikatoren Veränderungen in den unteren Einkommen erkären können. Dagegen gelang es, die Schätzer aus Abschnitt 4.1 durch die Schätzung der Variablen in Gruppen nach einer Reihe von Merkmalen etwas zu verbessern. Dies zeigte sich insbesondere bei der Betrachtung von Querschnittsdaten. Bei der Schätzung mit Paneldaten ist nach wie vor die Variable des (transformierten) Vorjahreseinkommens von sehr hoher Relevanz. Durch die Schätzung der Vulnerabilität in Gruppen bei 80% TPR konnten nur sehr kleine Verbesserungen erzielt werden. Eine entscheidende Rolle kommt der Wahl der VPL zu. Für die Gruppierung der Haushalte nach der Anzahl der Kinder gelang es, die VPL so zu fixieren, dass ein Großteil der Schätzer in Gruppen die Schätzer aus Abschnitt 4.1 etwas verbesserten.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

5 Zusammenfassung und Ausblick

Die Bekämpfung der Armut hat sowohl in Entwicklungsländern als auch in Industriestaaten in den letzten Jahren einen immer höheren Stellenwert in der Politik eingenommen. Die Armut und die Strategien zu ihrer Bekämpfung werden sowohl national durch die Regierungen als auch international u.a. durch die Weltbank und den IWF dokumentiert. Das Konzept des Weltentwicklungsberichts (World Bank, 2000/01) sieht drei Strategien zur Bekämpfung von Armut vor. Eine davon ist es, Armut von armutsgefährdeten Personen/Haushalten vorzubeugen. Dazu ist es natürlich von großer Bedeutung, die Haushalte zu identifizieren, die dem Risiko unterliegen, in der Zukunft in Armut zu fallen. Viele axiomatische und empirische Studien wurden im Bereich der Vulnerabilitäts-Forschung mit verschiedenen Konzepten durchgeführt. Häufig basieren diese Studien ihre Schätzungen lediglich auf Querschnittsdaten, da in Entwicklungsländern selten Paneldaten verfügbar sind. Die Qualität dieser Daten ist in den meisten Fällen nicht sehr verlässlich. Es ist anzunehmen, dass Schätzer basierend auf Querschnittsdaten niedriger Qualität nicht besonders genau sind. Die Genauigkeit der Schätzer ist jedoch von hoher Bedeutung, da diese die Grundlage für politische Maßnahmen sind.

In den meisten Studien zur Vulnerabilität wird bisher vernachlässigt, wie genau sich die Vulnerabilität schätzen lässt. Nur in wenigen Fällen wurde dieser Aspekt berücksichtigt, allerdings nach Kriterien, wie diese Arbeit darlegt, von nur geringem Nutzen. In dieser Arbeit wurde gezeigt, dass das Verfahren ROC zur Bewertung von Schätzern der Vulnerabilität gut geeignet ist. Es berücksichtigt nicht nur den Anteil der richtig identifizierten armen Haushalte sondern auch den Anteil der richtig identifizierten *nicht-armen Haushalte*. Für die Genauigkeit von Vulnerabilitäts-Schätzern ist es wichtig, dass eine Methode nicht zu viele *false positives* erzeugt. Diese treten auf, wenn ein Haushalt, der als armutsgefährdet klassifiziert wurde, nicht arm geworden ist. Andernfalls würden viele nicht-armutsgefährdete Haushalte von politischen Maßnahmen, wie z.B. Krediten, profitieren.

In dieser Arbeit wurde die ROC angewendet, um für Daten hoher Qualität die Genauigkeit von Vulnerabilitäts-Schätzern zu messen und zu bewerten. Als Daten-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

grundlage für die Analysen wurde das SOEP verwendet, da dieses die Vorteile eines relativ verständlichen und verlässlichen langen Paneldatensatz mit sich bringt. Insbesondere ermöglicht der Datensatz es, nachdem die Haushalte als armutsgefährdet oder nicht klassifiziert wurden, zu ermitteln, welche Haushalte tatsächlich arm geworden sind. Ebenso konnten die Analysen für eine ganze Reihe von Jahren wiederholt werden. Es wurden rollierende Panels von 1992-2009 und eine absolute Armutslinie von 9 000 € verwendet.

Wesentliche Bestandteile der Untersuchung waren

- Messung der Genauigkeit der Vulnerabilität in verschiedenen Situationen der Datenverfügbarkeit (Querschnittsdaten, Paneldaten mit stetigem Einkommen oder mit Einkommensklassen)
- Ermittlung relevanter Kovariablen (Haushaltskovariablen und Vorjahreseinkommen)
- Messung der Genauigkeit der Vulnerabilität über verschiedene Zeiträume
- Untersuchung von Möglichkeiten zur Verbesserung von erhaltenen Vulnerabilitäts-Schätzern für Deutschland unter Einbezug volkswirtschaftlicher Indikatoren oder der Haushaltsstruktur

Die Ergebnisse zeigen, dass die Vulnerabilität in Deutschland (und für Länder, in denen die Armutsquote niedrig ist) nur mit eingeschränkter Genauigkeit gemessen werden kann, sogar wenn idealerweise Paneldaten verfügbar sind und das beste Regressionsmodell mit vielen Kovariablen verwendet wird. Das ist insbesondere dann der Fall, wenn eine niedrige Armutsgrenze verwendet wurde, wie es in dieser Arbeit der Fall war. Im Gegensatz dazu könnte die Vulnerabilität der Armut unter Verwendung einer höheren Armutsgrenze, die zu einer höheren Armutsquote führen würde, genauer gemessen werden, wenn verlässliche Paneldaten für mindestens zwei Wellen verfügbar sind. Also wäre es für Entwicklungsländern mit hohen Armutsquoten möglich, die Vulnerabilität mit brauchbarer Genauigkeit vorherzusagen.

Es zeigte sich, dass Haushaltskovariablen wenig Einfluss auf die Genauigkeit von Vulnerabilitäts-Schätzern haben. Im Gegensatz dazu verbessert stetiges Vorjahreseinkommen die Schätzung wesentlich. Auch das Einkommen in Klassen verbessert die Schätzung im Vergleich zu Modellen, die lediglich auf Haushaltskovariablen basieren. Wenn die Einkommensklassen nach bestimmten Kriterien gewählt werden, ist es sogar möglich, Schätzer vergleichbarer Genauigkeit zu erzielen wie mit stetigem Einkommen. Wenn kleine oder mittlere Fehler bei der Zuteilung der Haushalte

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

auf die Einkommensklassen gemacht werden, führt dies nur zu geringen Verschlechterungen der Vulnerabilitäts-Schätzer. Es zeigt sich also, dass das Vorjahreseinkommen, auch wenn es nur grob und in wenigen Klassen vorhanden ist, zur Schätzung der Vulnerabilität verwendet werden sollte. Für Entwicklungsländer, in denen Einkommen schwierig zu messen und Daten daher selten vorhanden sind, empfiehlt es sich daher, die Haushalte grob Einkommensklassen zuzuteilen. Die Möglichkeit, die Schätzung der Vulnerabilität konkret für Deutschland mit Makroindikatoren zu verbessern, führte, zumindest mit den in dieser Arbeit verwendeten Methoden, zu keinem Ergebnis.

Es wurden Profile für einzelne Eigenschaften berechnet mit dem Ergebnis, dass einige Gruppen stärker armutsgefährdet sind als andere, z.B. getrennte oder geschiedene Haushalte im Gegensatz zu verheirateten; Haushalte mit erwerbstätigem HV in Banken oder Versicherungen gegenüber solchen mit nicht-erwerbstätigem HV sowie Ansässigkeit von Haushalten in ostdeutschen im Vergleich zu westdeutschen Bundesländern. Die Schätzung der Vulnerabilität in Gruppen nach verschiedenen Merkmalen des Haushalts führte für viele Merkmale zu einer leichten Verbesserung der Genauigkeit für Deutschland. Die Verbesserungen waren bei der Schätzung mit Querschnittsdaten etwas größer als bei der Schätzung mit Paneldaten. Bei der Schätzung mit Paneldaten ist die Kovariable Vorjahreseinkommen von hoher Relevanz. Es scheint, dass diese durch die Schätzung der Vulnerabilität in Gruppen nicht gesteigert werden kann. Ein wichtiger Aspekt für die Schätzung der Vulnerabilität liegt in der Festlegung der VPLs. In der Anpassung der VPL liegt möglicherweise noch Verbesserungspotential für die Genauigkeit der Vulnerabilitäts-Schätzung.

Wie erwartet, führte die Schätzung mit einem größeren Vorhersagehorizont zu ungenaueren Schätzern, je weiter dieser in der Zukunft liegt.

Der SOEP-Datensatz ist verständlicher und vollständiger als solche von anderen Haushaltsbefragungen. Es ist anzunehmen, dass die Schätzer der Vulnerabilität weniger genau sind, wenn sie auf Daten schlechterer Qualität basieren. Folglich müssen Schätzer der Vulnerabilität, die auf Querschnittsdaten basieren, mit Vorsicht betrachtet werden.

Aus politischer Perspektive betrachtet, zeigen die Ergebnisse, dass es durchaus nicht der beste Ansatz zur Prävention von Armut ist, statistisch die armutsgefährdeten Haushalte zu identifizieren und diese durch Kredite oder Hilfsprogramme zu unterstützen. Diese Maßnahme ist insbesondere dann wenig erfolgsversprechend, wenn die Armutsquoten niedrig sind. In diesen Fällen würden selbst

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

dann, wenn die armutsgefährdeten Haushalte perfekt identifiziert würden, zu viele *false positives* profitieren. Das bedeutet natürlich auch, dass Schocks, die nicht aus den Daten antizipiert werden können, sehr wichtig sind, um die armutsgefährdeten Haushalte zu identifizieren. Bisher mangelt es jedoch an Methoden, die externe Schocks (Wetter/Krisen/Katastrophen) berücksichtigen, aus dem Grund, da die Datenbasis nicht ausreicht, um externe Schocks zu modellieren und vorauszusagen.

Für die weitere Forschung könnte versucht werden, Szenario-Analysen für verschiedene Ereignisse durchzuführen (z.B. für Wetterschocks wie Dürre/extremer Niederschlag (Christiaensen und Subbarao, 2005), z.B. für die Finanzkrise das Szenario Austritt eines Landes aus dem Euro/extrem niedriger Wert des Euros und für Katastrophen z.B. Erdbeben/Tsunami/Bürgerkrieg) und zu analysieren, welche Auswirkungen diese Ereignisse auf die Vulnerabilität der Haushalte und deren Genauigkeit haben könnten.

Schließlich könnte auch auf der Mikroebene untersucht werden, welche Ereignisse dazu führen, dass Haushalte arm werden (z.B. Trennung des Haushaltes, Arbeitslosigkeit) und versucht werden, diese Veränderungen für die Haushalte vorherzusagen.

Von großem Interesse ist es natürlich auch weiterhin, Makrovariablen für die Verbesserung der Vulnerabilitäts-Schätzer zu verwenden. In Zukunft könnte versucht werden, Zeitreihenmodelle anzupassen. Eine weitere Möglichkeit besteht darin, die Makrovariablen interagiert mit den Haushaltseigenschaften in die verwendeten Modelle P_1-P_6 einzufügen (Ansätze dazu in Christiaensen und Subbarao (2005)), so dass die Makrovariable bei der Schätzung auf Gruppen von Haushalten unterschiedlich wirken.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A Anhang

A.1 Daten

A.1.1 Datei- und Variablenbezeichnungen im SOEP

Verwendet wurden Variablen aus

- der Masterdatei *ppfad*,
- der Datei der Haushaltsgewichte *hhrf*,
- den generierten und Statusvariablen auf der Personenebene *$pgen* und auf der Haushaltsebene *$hgen*,
- den harmonisierten Variablen aus dem CNEF *$pequiv*.

Ein $ Zeichen vor einer Datei oder einer Variablen gibt die Welle an, von der die Datei/Variable ist. Diese Welleneigenschaft wird mit Kleinbuchstaben $a-z$ notiert, wobei a für das Jahr 1984 und dementsprechend z für das Jahr 2009 steht. Z.B. ist *apgen* die Datei der personenbezogenen generierten und Statusvariablen für 1984. Bei den harmonisierten und einigen anderen Variablen werden die Wellen nicht durch einen Kleinbuchstaben am Anfang der Variablen gekennzeichnet sondern durch die letzten beiden Ziffern des Jahres am Ende der Variablen. Einige Variablen sind Vorjahres-Variablen; dies ist explizit gekennzeichnet (DTC (2005), Grabka (2005), Anger (2011), SOEP (2009)).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.1.2 Auswahl der Daten

Themen	Variablen	Datei	Name
Wohlfahrt	Brutto-Haushaltseinkommen (des Vorjahres)	$pequiv	i11101$$
	Netto-Haushaltseinkommen (des Vorjahres)	$pequiv	i11102$$
	Kalkulatorische Miete (des Vorjahres)	$pequiv	i11105$$
Demographie	Geschlecht	ppfad	sex
	Alter	$pequiv	d11101$$
	Familienstand	$pequiv	d11104$$
	Anzahl der Personen im Haushalt	$pequiv	d11106$$
	Anzahl der Kinder im Haushalt	$pequiv	d11107$$
	Anzahl Kinder 0-14 Jahren im Haushalt	$pequiv	h11101$$
	Haushaltsvorstand	$pequiv	d11105$$
	Bundesland	$pequiv	l11101$$
Bildung	Höchster Bildungsstand (bzgl. Schulbildung)	$pgen	$psbil
Arbeitsmarkt	Industrie	$pequiv	e11106$$
	Labour Force	$pgen	lfs$$
	Arbeitsmarktstatus	$pgen	emplst$$
	Öffentlicher Dienst	$pgen	oeffd$$
	Arbeitsstunden (des Vorjahres)	$pequiv	e11101$$
Wohnsitz	Status des Eigentums	$hgen	owner$$
	Wohnfläche	$hgen	size$$
	Miete	$hgen	rent$$
Haushalt	Art	ppfad	$pop
	Migrationshintergrund	ppfad	migback
Gesundheit	Selbsteinschätzung Gesundheit	$pequiv	m11126$$
	Krankenhausaufenthalte (im Vorjahr)	$pequiv	m11102$$
	Arztbesuche (im Vorjahr)	$pequiv	m11127$$
	Arbeitsunfähigkeit	$pequiv	m11125$$
Makrovariable	Verbraucherpreisindex (für das Vorjahreseinkommen)	$pequiv	y11101$$
	Haushalts-Querschnittsgewichte	hhrf	&hhrf
	Haushalts-Bleibefaktoren	hhrf	$hbleib

Tabelle A.1: *Bezeichnung ausgewählter Variablen und Speicherort im SOEP [Haisken-DeNew und Frick (2005), Soepinfo]*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.1.3 Darstellung der Einkommensmobilitäten als Markov-Kette

Im Folgenden werden im Rahmen der Untersuchungen von Abschnitt 3.3 die Übergangsmatrizen $\hat{T}_{t,t+1}$, $\hat{T}_{t,t+2}$ und $(\hat{T}_{t,t+1})^2$ für $t = 1992$ dargestellt. Die Matrix $\hat{T}_{92,93}$ (und damit auch $\hat{T}^2_{92,93}$) basiert auf einem Panel mit zwei und die Matrix $\hat{T}_{92,94}$ auf einem Panel bestehend aus drei Wellen Einkommen.

$$\hat{T}_{92,93} = \begin{pmatrix} 0.508 & 0.314 & 0.113 & 0.050 & 0.014 \\ 0.115 & 0.603 & 0.207 & 0.058 & 0.017 \\ 0.033 & 0.164 & 0.534 & 0.213 & 0.056 \\ 0.010 & 0.066 & 0.223 & 0.531 & 0.170 \\ 0.006 & 0.037 & 0.085 & 0.251 & 0.621 \end{pmatrix}$$

$$\hat{T}_{92,94} = \begin{pmatrix} 0.491 & 0.329 & 0.113 & 0.048 & 0.019 \\ 0.156 & 0.508 & 0.221 & 0.096 & 0.018 \\ 0.044 & 0.199 & 0.474 & 0.232 & 0.052 \\ 0.045 & 0.065 & 0.249 & 0.464 & 0.177 \\ 0.019 & 0.048 & 0.116 & 0.299 & 0.517 \end{pmatrix}$$

$$(\hat{T}_{92,93})^2 = \begin{pmatrix} 0.298 & 0.371 & 0.195 & 0.098 & 0.036 \\ 0.135 & 0.438 & 0.263 & 0.120 & 0.044 \\ 0.056 & 0.213 & 0.375 & 0.252 & 0.104 \\ 0.026 & 0.121 & 0.267 & 0.376 & 0.210 \\ 0.016 & 0.078 & 0.162 & 0.310 & 0.434 \end{pmatrix}$$

$\rightarrow (\hat{T}_{92,93})^2 \neq \hat{T}_{92,94}$

Es wird vermutet, dass es sich bei den Einkommensmobilitäten nicht um eine homogene Markov-Kette handelt. Insbesondere die Bleibewahrscheinlichkeiten von $(\hat{T}_{92,93})^2$ und $\hat{T}_{92,94}$ unterscheiden sich relativ stark.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.2 Experimente zur Modellwahl zur Schätzung der Vulnerabilität

Im Rahmen der Modellauswahl der ersten Studie (Abschnitt 4.1) wurden weitere Modellvarianten untersucht, die im Hauptteil lediglich skizziert wurden. Dieser Abschnitt des Anhangs stellt die Ergebnisse der Varianten dar. Diese sind:

- die Berücksichtigung der Armutshistorie als Kovariable im Modell,
- die Modellierung mit logarithmisch transformierten Einkommen und
- die Verwendung von heteroskedastischen Modellen unter Verwendung verschiedener Verteilungsannahmen.

A.2.1 Berücksichtigung der Armutshistorie

In diesem Abschnitt werden die Ergebnisse der Schätzung der Vulnerabilität unter Berücksichtigung der Armutshistorie (der vorangegangenen 5 Jahre $(t-2, ..., t-6)$) als Kovariable gezeigt. Es werden zwei verschiedene Varianten betrachtet. In der ersten Variante (Modell $P_{6,D}$) wird die Armutshistorie durch fünf binäre Dummy-Variablen (Haushalt nicht-arm, arm) modelliert (für jedes Jahr eine). In der zweiten Variante (Modell $P_{6,anz5}$) wird die Armutshistorie durch die Anzahl der Armutsperioden eines Haushalts zwischen den Jahren $t-2, ..., t-6$ abgebildet. Diese Betrachtungen bedürfen eines längeren Panels von neun Jahren. Die Ergebnisse werden im Folgenden für einen Panel von 1999-2007 gezeigt. Regressionstabellen zu $P_{6,D}$ und $P_{6,anz5}$ sind in B.3 und B.4 dargestellt. Zum Vergleich wird auch die Regressionstabelle von P_6 für den langen Panel gezeigt. Vergleiche der Genauigkeit der Vulnerabilitäts-Schätzer der drei Modelle basierend auf den ROC-Kurven sind in B.1 abgebildet. Im Hauptteil wird zu den Eregbnissen auf Seite 76 referiert.

$$\begin{aligned} y_{05} &= X_{04}\beta + y_{04}\gamma + d_{03}\delta_1 + d_{02}\delta_2 + d_{01}\delta_3 + d_{00}\delta_4 + d_{99}\delta_5 + e_{05} \\ \hat{y}_{06} &= X_{05}\hat{\beta} + y_{05}\hat{\gamma} + d_{04}\hat{\delta}_1 + d_{03}\hat{\delta}_2 + d_{02}\hat{\delta}_3 + d_{01}\hat{\delta}_4 + d_{00}\hat{\delta}_5 \end{aligned} \qquad (P_{6,D})$$

$$\begin{aligned} y_{05} &= X_{04}\beta + y_{04}\gamma + a_{03,99}\delta \\ \hat{y}_{06} &= X_{05}\hat{\beta} + y_{05}\hat{\gamma} + a_{04,00}\hat{\delta} \end{aligned} \qquad (P_{6,anz5})$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	5583.94	550.81	10.14	<0.001
Vorjahreseinkommen	0.72	0.02	48.07	<0.001
Geschlecht: weiblich	-70.52	161.59	-0.44	0.663
$n[0, 18)$	-391.91	103.45	-3.79	<0.001
$n[18, 34)$	-380.60	155.51	-2.45	0.014
$n[34, 59)$	-636.94	185.92	-3.43	0.001
$n[59-)$	-512.37	212.18	-2.41	0.016
Familienstand: alleinstehend	-562.31	284.22	-1.98	0.048
Familienstand: verwitwet	79.84	273.05	0.29	0.770
Familienstand: geschieden	-756.56	259.55	-2.91	0.004
Familienstand: getrennt	-1449.09	570.06	-2.54	0.011
Gesamtarbeitsstunden im Vorjahr	-0.03	0.11	-0.24	0.811
Gewerbe: Landwirtschaft	1668.77	756.81	2.21	0.028
Gewerbe: Energiesektor	650.42	1101.06	0.59	0.555
Gewerbe: Bergbau	-1049.25	1483.05	-0.71	0.479
Gewerbe: Produktion	1364.40	345.80	3.95	<0.001
Gewerbe: Baugewerbe	1.56	376.57	0.00	0.997
Gewerbe: Handel	19.86	347.69	0.06	0.954
Gewerbe: Transport	-309.44	450.66	-0.69	0.492
Gewerbe: Bank/Versicherung	1005.82	659.80	1.52	0.128
Gewerbe: Dienstleistungen	280.50	285.47	0.98	0.326
Wohnverhältnis: Hauptmieter	-495.90	172.84	-2.87	0.004
Wohnverhältnis: Untermieter	-457.77	480.82	-0.95	0.341
Größe des Wohnsitzes	15781.67	4837.89	3.26	0.001
(Größe des Wohnsitzes)^2	-4384.85	3622.71	-1.21	0.226
Schulabschluss: Realschulabschluss	490.98	167.41	2.93	0.003
Schulabschluss: Abitur/FHSR	1278.00	223.88	5.71	<0.001
Schulabschluss: andere Abschlüsse	223.61	276.60	0.81	0.419
Schulabschluss: kein Abschluss	-492.43	489.07	-1.01	0.314
n(Vollzeitbeschäftigung)	15355.84	9722.96	1.58	0.114
n(Vollzeitbeschäftigung)^2	-4223.04	4667.43	-0.90	0.366
n(Teilzeitbeschäftigung)	892.74	253.32	3.52	<0.001
adjustiertes Betimmtheitsmaß	0.647			
Anz. Beobachtungen (ungewichtet)	2534			
F-Statitsik	148.6	(df: 31,2435)		

Tabelle B.2: *Ergebnisse der Regression für Modell P_6 mit einem Panel von 1999-2007 [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	6356.79	563.76	11.28	<0.001
Vorjahreseinkommen	0.69	0.02	43.40	<0.001
arm1	-935.00	300.32	-3.11	0.002
arm2	152.07	323.49	0.47	0.638
arm3	-302.94	340.67	-0.89	0.374
arm4	-569.44	337.68	-1.69	0.092
arm5	-595.88	300.55	-1.98	0.048
Geschlecht: weiblich	-89.19	160.58	-0.56	0.579
$n[0,18)$	-368.06	102.94	-3.58	<0.001
$n[18,34)$	-333.30	154.94	-2.15	0.032
$n[34,59)$	-623.07	184.94	-3.37	0.001
$n[59-)$	-595.03	211.81	-2.81	0.005
Familienstand: alleinstehend	-395.41	283.80	-1.39	0.164
Familienstand: verwitwet	64.20	271.40	0.24	0.813
Familienstand: geschieden	-660.48	259.34	-2.55	0.011
Familienstand: getrennt	-1242.48	570.17	-2.18	0.029
Gesamtarbeitsstunden im Vorjahr	-0.06	0.11	-0.53	0.598
Gewerbe: Landwirtschaft	1607.12	752.79	2.13	0.033
Gewerbe: Energiesektor	595.33	1093.15	0.54	0.586
Gewerbe: Bergbau	-1149.01	1472.43	-0.78	0.435
Gewerbe: Produktion	1296.03	343.80	3.77	<0.001
Gewerbe: Baugewerbe	-40.76	375.25	-0.11	0.914
Gewerbe: Handel	-23.49	345.63	-0.07	0.946
Gewerbe: Transport	-313.87	447.48	-0.70	0.483
Gewerbe: Bank/Versicherung	976.89	655.13	1.49	0.136
Gewerbe: Dienstleistungen	200.08	284.03	0.70	0.481
Wohnverhältnis: Hauptmieter	-517.02	172.16	-3.00	0.003
Wohnverhältnis: Untermieter	-357.84	479.58	-0.75	0.456
Größe des Wohnsitzes	14341.01	4811.48	2.98	0.003
(Größe des Wohnsitzes)^2	-3663.42	3599.94	-1.02	0.309
Schulabschluss: Realschulabschluss	427.02	166.69	2.56	0.010
Schulabschluss: Abitur/FHSR	1305.15	222.53	5.87	<0.001
Schulabschluss: andere Abschlüsse	220.24	274.77	0.80	0.423
Schulabschluss: kein Abschluss	-278.95	487.25	-0.57	0.567
n(Vollzeitbeschäftigung)	14093.80	9674.21	1.46	0.145
n(Vollzeitbeschäftigung)^2	-1724.61	4656.21	-0.37	0.711
n(Teilzeitbeschäftigung)	920.02	252.01	3.65	<0.001
adjustiertes Betimmtheitsmaß	0.652			
Anz. Beobachtungen (ungewichtet)	2534			
F-Statitsik	148.6	(df: 36,2430)		

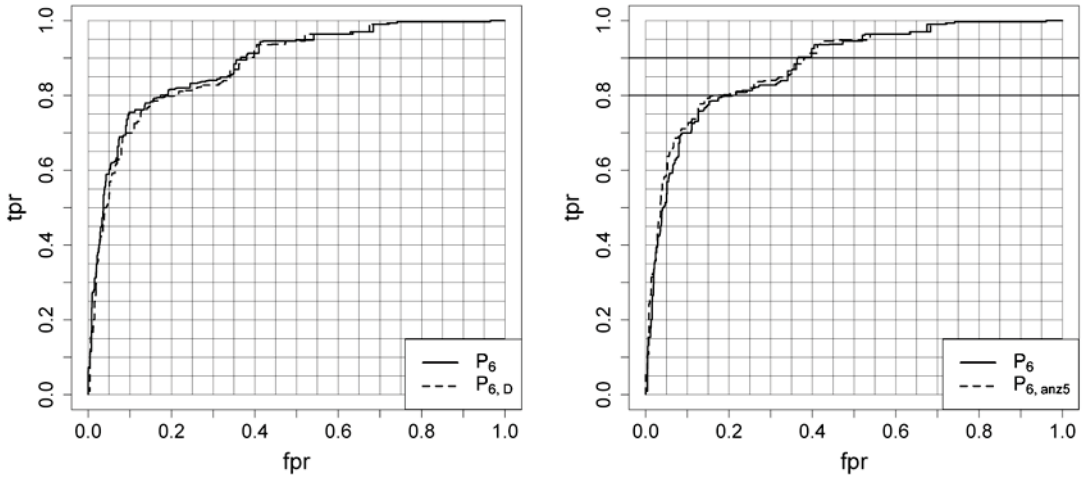
Tabelle B.3: *Ergebnisse der Regression für Modell $P_{6,D}$ mit einem Panel von 1999-2007 [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	6343.21	561.63	11.29	<0.001
Vorjahreseinkommen	0.69	0.02	43.46	<0.001
n(arm)	-442.43	74.32	-5.95	<0.001
Geschlecht: weiblich	-93.48	160.51	-0.58	0.560
$n[0, 18)$	-370.03	102.79	-3.60	<0.001
$n[18, 34)$	-320.84	154.75	-2.07	0.038
$n[34, 59)$	-614.84	184.66	-3.33	0.001
$n[59-)$	-589.55	211.09	-2.79	0.005
Familienstand: alleinstehend	-388.07	283.74	-1.37	0.172
Familienstand: verwitwet	74.86	271.14	0.28	0.782
Familienstand: geschieden	-653.82	258.31	-2.53	0.011
Familienstand: getrennt	-1239.22	567.17	-2.18	0.029
Gesamtarbeitsstunden im Vorjahr	-0.07	0.11	-0.59	0.552
Gewerbe: Landwirtschaft	1672.23	751.51	2.23	0.026
Gewerbe: Energiesektor	602.92	1093.38	0.55	0.581
Gewerbe: Bergbau	-1127.55	1472.73	-0.77	0.444
Gewerbe: Produktion	1309.14	343.51	3.81	<0.001
Gewerbe: Baugewerbe	-20.50	373.96	-0.05	0.956
Gewerbe: Handel	-33.45	345.38	-0.10	0.923
Gewerbe: Transport	-307.58	447.50	-0.69	0.492
Gewerbe: Bank/Versicherung	978.55	655.20	1.49	0.135
Gewerbe: Dienstleistungen	219.59	283.66	0.77	0.439
Wohnverhältnis: Hauptmieter	-506.46	171.64	-2.95	0.003
Wohnverhältnis: Untermieter	-418.60	477.50	-0.88	0.381
Größe des Wohnsitzes	14534.60	4808.61	3.02	0.003
(Größe des Wohnsitzes)^2	-3918.94	3598.21	-1.09	0.276
Schulabschluss: Realschulabschluss	429.18	166.56	2.58	0.010
Schulabschluss: Abitur/FHSR	1313.98	222.40	5.91	<0.001
Schulabschluss: andere Abschlüsse	205.19	274.68	0.75	0.455
Schulabschluss: kein Abschluss	-264.21	487.16	-0.54	0.588
n(Vollzeitbeschäftigung)	14289.79	9656.59	1.48	0.139
n(Vollzeitbeschäftigung)^2	-1847.23	4651.93	-0.40	0.691
n(Teilzeitbeschäftigung)	923.89	251.60	3.67	<0.001
adjustiertes Betimmtheitsmaß	0.652			
Anz. Beobachtungen (ungewichtet)	2534			
F-Statitsik	145.3	(df: 32,2435)		

Tabelle B.4: *Ergebnisse der Regression für Modell $P_{6,anz5}$ mit einem Panel von 1999-2007 [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Genauigkeit der Schätzung mit $P_{6,D}$ (b) Genauigkeit der Schätzung mit $P_{6,anz5}$

Abbildung B.1: *ROC-Kurven von Vulnerabilitäts-Schätzern mit Armutshistorie (1999-2007) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

A.2.2 Vulnerabilitäts-Schätzung mit transformiertem Einkommen

In diesem Abschnitt werden die im Hauptteil verwendeten Modelle zur Schätzung der Vulnerabilität modifiziert, indem das Einkommen logarithmisch transformiert wird. Dargestellt werden im Folgenden die Ergebnisse zu den Schätzern der Modelle P_1, P_4 und P_6 für das Jahr 1996, die im Folgenden mit $\mathrm{P}_{1log,96}$, $\mathrm{P}_{4log,96}$ und $\mathrm{P}_{6log,96}$ bezeichnet werden. Zu den Regressionen werden die Regressionstabellen in B.5, B.6 und B.7 dargestellt. Weiterhin werden die Residuenplots der drei Regressionen betrachtet und hinsichtlich der Residuenplots zu den zugehörigen Regressionen ohne Transformation des Einkommens verglichen (siehe Abbildung B.2). Abbildung B.3 zeigt die ROC-Kurven zu den Schätzern basierend auf den Modellen mit und ohne Transformation. Zu den Ergebnissen wird im Hauptteil in Abschnitt 4.1.1 auf Seite 72 referiert.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$$\begin{aligned} \text{Schätze } \beta \text{ mit} \quad & \log(y_{95}) = X_{95}\beta + e_{95} \\ \text{Vorhersage mit} \quad & E(\log y_{96}|X_{95}) = X_{95}\hat{\beta} \end{aligned} \tag{P$_{1log,96}$}$$

$$\begin{aligned} \text{Schätze } \beta \text{ und } \gamma \text{ mit} \quad & \log(y_{95}) = \gamma \log(y_{94}) + X_{95}\beta + e_{95} \\ \text{Vorhersage mit} \quad & E(\log y_{96}| \log y_{95}, X_{96}) = \hat{\gamma} \log(y_{95}) + X_{96}\hat{\beta} \end{aligned} \tag{P$_{4log,96}$}$$

$$\begin{aligned} \text{Schätze } \beta \text{ und } \gamma \text{ mit} \quad & \log(y_{95}) = \gamma \log(y_{94}) + X_{94}\beta + e_{95} \\ \text{Vorhersage mit} \quad & E(\log y_{96}| \log y_{95}, X_{95}) = \hat{\gamma} \log(y_{95}) + X_{95}\hat{\beta} \end{aligned} \tag{P$_{6log,96}$}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	9.73	0.04	239.88	<0.001
Geschlecht: weiblich	-0.04	0.02	-2.68	0.007
$n[0,18)$	-0.10	0.01	-11.39	<0.001
$n[18,34)$	-0.09	0.01	-6.15	<0.001
$n[34,59)$	-0.07	0.02	-4.40	<0.001
$n[59-)$	0.02	0.02	1.20	0.229
Familienstand: alleinstehend	0.00	0.02	0.07	0.944
Familienstand: verwitwet	0.07	0.02	2.79	0.005
Familienstand: geschieden	-0.11	0.02	-4.85	<0.001
Familienstand: getrennt	-0.11	0.05	-2.12	0.034
Gesamtarbeitsstunden im Vorjahr	0.00	0.00	1.84	0.065
Gewerbe: Landwirtschaft	-0.01	0.06	-0.19	0.847
Gewerbe: Energiesektor	0.27	0.07	3.85	<0.001
Gewerbe: Bergbau	0.04	0.12	0.35	0.725
Gewerbe: Produktion	0.12	0.03	4.49	<0.001
Gewerbe: Baugewerbe	0.11	0.03	3.76	<0.001
Gewerbe: Handel	0.04	0.03	1.27	0.205
Gewerbe: Transport	0.14	0.04	3.66	<0.001
Gewerbe: Bank/Versicherung	0.28	0.05	5.29	<0.001
Gewerbe: Dienstleistungen	0.11	0.03	4.27	<0.001
Wohnverhältnis: Hauptmieter	-0.06	0.02	-4.14	<0.001
Wohnverhältnis: Untermieter	-0.26	0.03	-7.75	<0.001
Größe des Wohnsitzes	4.11	0.51	8.08	<0.001
(Größe des Wohnsitzes)^2	-1.48	0.39	-3.78	<0.001
Schulabschluss: Realschulabschluss	0.09	0.02	5.64	<0.001
Schulabschluss: Abitur/FHSR	0.15	0.02	7.80	<0.001
Schulabschluss: andere Abschlüsse	-0.11	0.04	-2.87	0.004
Schulabschluss: kein Abschluss	-0.15	0.05	-3.36	0.001
n(Vollzeitbeschäftigung)	9.45	1.05	9.00	<0.001
n(Vollzeitbeschäftigung)^2	-2.99	0.56	-5.32	<0.001
n(Teilzeitbeschäftigung)	0.08	0.02	3.88	<0.001
adjustiertes Betimmtheitsmaß	0.246			
Anz. Beobachtungen (ungewichtet)	4493			
F-Statitsik	48.08	(df: 30,4301)		

Tabelle B.5: *Ergebnisse der Regression für Modell $P_{1log,96}$ [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	5.00	0.15	34.33	<0.001
log(Vorjahreseinkommen)	0.49	0.01	33.52	<0.001
Geschlecht: weiblich	-0.01	0.01	-0.64	0.525
$n[0, 18)$	-0.05	0.01	-6.08	<0.001
$n[18, 34)$	-0.05	0.01	-3.73	<0.001
$n[34, 59)$	-0.04	0.01	-3.15	0.002
$n[59-)$	-0.02	0.02	-1.15	0.252
Familienstand: alleinstehend	0.03	0.02	1.36	0.173
Familienstand: verwitwet	0.05	0.02	2.51	0.012
Familienstand: geschieden	-0.07	0.02	-3.13	0.002
Familienstand: getrennt	-0.09	0.05	-2.03	0.042
Gesamtarbeitsstunden im Vorjahr	-0.00	0.00	-2.69	0.007
Gewerbe: Landwirtschaft	0.03	0.05	0.61	0.542
Gewerbe: Energiesektor	0.18	0.06	2.97	0.003
Gewerbe: Bergbau	-0.02	0.10	-0.18	0.858
Gewerbe: Produktion	0.07	0.02	2.81	0.005
Gewerbe: Baugewerbe	0.08	0.03	3.16	0.002
Gewerbe: Handel	0.04	0.03	1.65	0.100
Gewerbe: Transport	0.09	0.03	2.66	0.008
Gewerbe: Bank/Versicherung	0.15	0.05	3.29	0.001
Gewerbe: Dienstleistungen	0.06	0.02	2.82	0.005
Wohnverhältnis: Hauptmieter	-0.06	0.01	-4.17	<0.001
Wohnverhältnis: Untermieter	-0.26	0.03	-8.63	<0.001
Größe des Wohnsitzes	2.23	0.46	4.87	<0.001
(Größe des Wohnsitzes)^2	-0.80	0.35	-2.28	0.023
Schulabschluss: Realschulabschluss	0.04	0.01	2.82	0.005
Schulabschluss: Abitur/FHSR	0.07	0.02	3.98	<0.001
Schulabschluss: andere Abschlüsse	-0.04	0.03	-1.26	0.208
Schulabschluss: kein Abschluss	-0.05	0.04	-1.16	0.245
n(Vollzeitbeschäftigung)	6.94	0.94	7.39	<0.001
n(Vollzeitbeschäftigung)^2	-1.00	0.50	-1.98	0.047
n(Teilzeitbeschäftigung)	0.07	0.02	3.69	<0.001
adjustiertes Betimmtheitsmaß	0.402			
Anz. Beobachtungen (ungewichtet)	4493			
F-Statitsik	94.93	(df: 31,4300)		

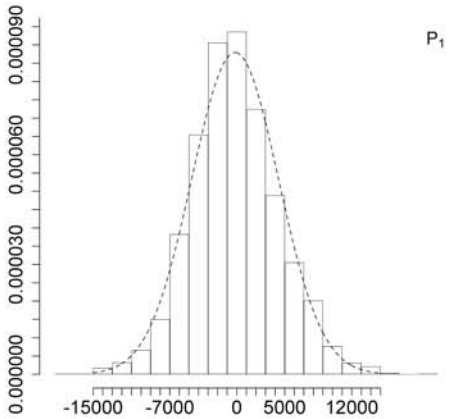
Tabelle B.6: *Ergebnisse der Regression für Modell $P_{4log,96}$ [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

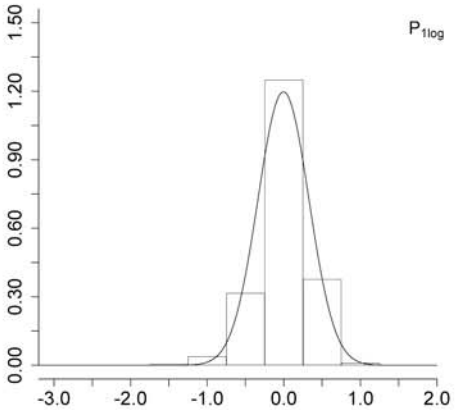
	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	4.74	0.15	31.65	<0.001
log(Vorjahreseinkommen)	0.50	0.01	33.79	<0.001
Geschlecht: weiblich	-0.01	0.01	-0.76	0.446
$n[0,18)$	-0.03	0.01	-3.65	<0.001
$n[18,34)$	-0.02	0.01	-1.42	0.156
$n[34,59)$	-0.01	0.01	-0.77	0.439
$n[59-)$	0.02	0.02	1.55	0.122
Familienstand: alleinstehend	0.07	0.02	3.38	0.001
Familienstand: verwitwet	0.05	0.02	2.28	0.023
Familienstand: geschieden	-0.05	0.02	-2.11	0.035
Familienstand: getrennt	0.05	0.04	1.16	0.247
Gesamtarbeitsstunden im Vorjahr	0.00	0.00	2.35	0.019
Gewerbe: Landwirtschaft	-0.03	0.06	-0.59	0.557
Gewerbe: Energiesektor	0.15	0.06	2.58	0.010
Gewerbe: Bergbau	0.01	0.09	0.10	0.921
Gewerbe: Produktion	0.06	0.02	2.50	0.012
Gewerbe: Baugewerbe	0.06	0.03	2.39	0.017
Gewerbe: Handel	0.03	0.03	0.98	0.326
Gewerbe: Transport	0.07	0.04	2.04	0.042
Gewerbe: Bank/Versicherung	0.14	0.05	3.09	0.002
Gewerbe: Dienstleistungen	0.05	0.02	2.17	0.030
Wohnverhältnis: Hauptmieter	-0.05	0.01	-3.88	<0.001
Wohnverhältnis: Untermieter	-0.25	0.03	-8.17	<0.001
Größe des Wohnsitzes	1.81	0.46	3.90	<0.001
(Größe des Wohnsitzes)^2	-0.68	0.36	-1.88	0.060
Schulabschluss: Realschulabschluss	0.03	0.01	2.38	0.017
Schulabschluss: Abitur/FHSR	0.07	0.02	4.07	<0.001
Schulabschluss: andere Abschlüsse	-0.05	0.04	-1.50	0.132
Schulabschluss: kein Abschluss	-0.08	0.04	-1.92	0.055
n(Vollzeitbeschäftigung)	1.77	0.95	1.86	0.063
n(Vollzeitbeschäftigung)^2	-1.25	0.55	-2.27	0.023
n(Teilzeitbeschäftigung)	0.02	0.02	1.10	0.271
adjustiertes Betimmtheitsmaß	0.387			
Anz. Beobachtungen (ungewichtet)	4493			
F-Statitsik	89.24	(df: 31,4300)		

Tabelle B.7: *Ergebnisse der Regression für Modell $P_{6log,96}$ [Eigene Berechnungen auf Basis der in Kapitel 3 beschriebenen Daten].*

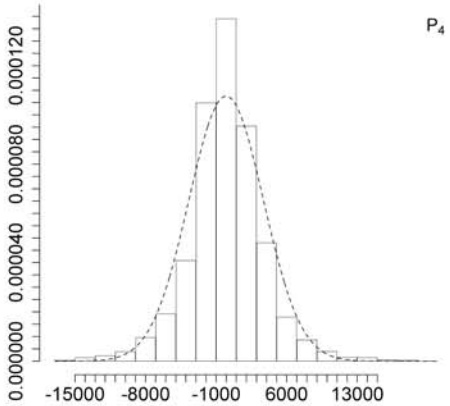
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



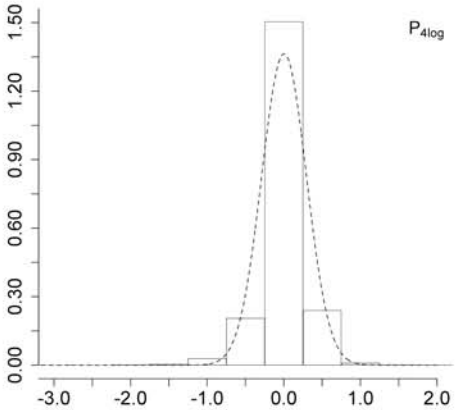
(a) Residuen der Regression von $P_{1,96}$



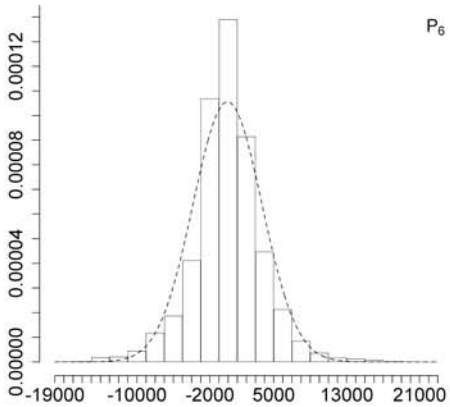
(b) Residuen der Regression von $P_{1log,96}$



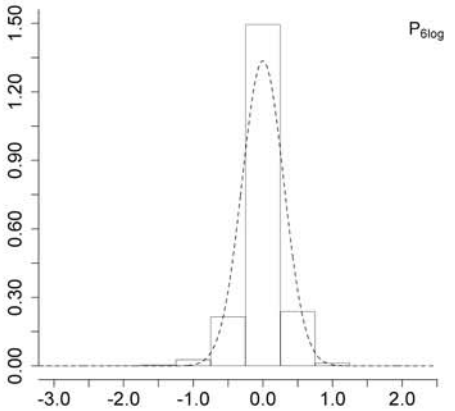
(c) Residuen der Regression von $P_{4,96}$



(d) Residuen der Regression von $P_{4log,96}$



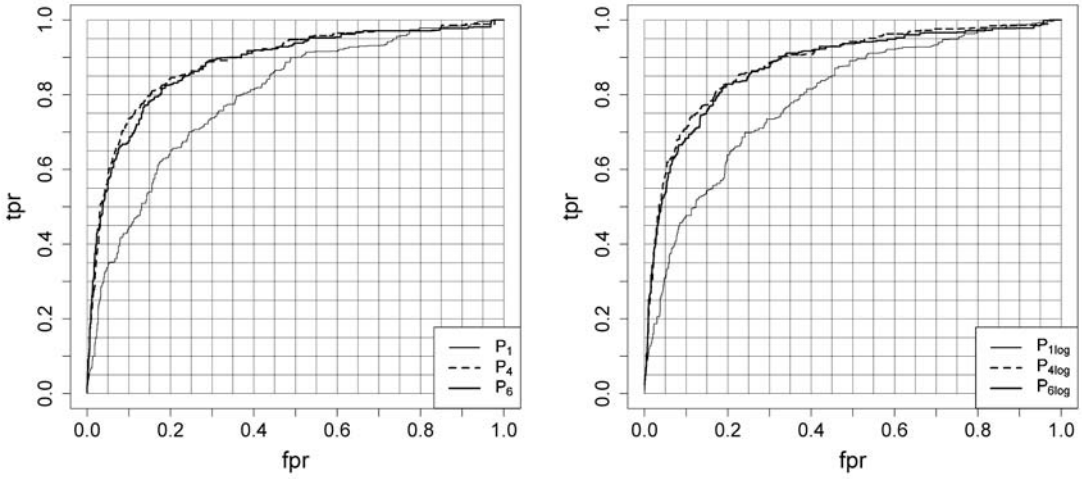
(e) Residuen der Regression von $P_{6,96}$



(f) Residuen der Regression von $P_{6log,96}$

Abbildung B.2: *Residuen der Regressionen mit und ohne log Transformation [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) ROC-Kurven der Ausgangsmodelle (b) ROC mit logarithmischem Einkommen

Abbildung B.3: *ROC-Kurven der Schätzungen mit und ohne log Transformation des Einkommens (1996) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

A.2.3 Vulnerabilitäts-Schätzung mit heteroskedastischen Methoden

In diesem Abschnitt wird die Genauigkeit mit heteroskedastischen Modellen geschätzt und die Genauigkeit mit derer von homoskedastischen Modellen verglichen. Die bedingte Verteilung $y|x$ wird durch den Erwartungswert und die Varianz (ggfs. durch Schiefe und Kurtosis) einer gewählten Verteilung als Funktionen in Abhängigkeit von den Kovariablen x dargestellt. g_1 und g_2 sind Linkfunktionen für den Erwartungswert und die Varianz. Ist g_1 ein Link der Identität, ergibt sich die Vorhersage des Response aus der Schätzung des ersten Momentes der Verteilung.

$$y|x \sim V\left(g_1(x), g_2(x)\right)$$
$$E(y|x) = \hat{g}_1(x)$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$$g_1(\mu_{96}) = X_{1,95}\beta$$

$$g_2(\sigma_{96}) = X_{2,95}\theta \qquad (\mathrm{P}_{1het,96})$$

$$g_1(\mu_{96}) = \gamma y_{95} + X_{1,95}\beta$$

$$g_2(\sigma_{96}) = X_{2,95}\beta \qquad (\mathrm{P}_{6het,96})$$

Die Vulnerabilität wird unter der Verteilungsannahme eine Normalverteilung mit Erwartungswert μ und Varianz σ (Familie NO2() mit `gamlss`) geschätzt. Die Ergebnisse der Regressionen der Modelle $\mathrm{P}_{1het,96}$ (korrespondierend zu dem homoskedastischen Modell P_1) und $\mathrm{P}_{6het,96}$ (korrespondierend zu P_6) sind in den Tabellen B.8-B.11 dargestellt. Ein Vergleich der Genauigkeit der Schätzer auf Basis der ROC-Kurven ist in Abbildung B.4 zu sehen. In Abbildung B.2 war zu sehen, dass weder die Unterstellung einer Normalverteilung noch die einer Lognormalverteilung zu normalverteilten Residuen führte. Aus diesem Grunde wurde mit weiteren Verteilungen, nämlich der t-Verteilung, der Gamma-Verteilung und der Weibull-Verteilung, experimentiert. Die Quantilsresiduen der Regressionen sind in Abbildung C.7 dargestellt. Die Erläuterung der Ergebnisse befindet sich im Hauptteil im Abschnitt 4.1.1 auf Seite 72.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

mu link		Schätzer	Std.fehler	t-Wert	P-Wert
Identität	(Achsenabschnitt)	18970.00	6.27	3025.93	<0.001
	Geschlecht: weiblich	-605.50	2.36	-256.66	<0.001
	$n[0, 18)$	-1599.00	1.21	-1318.11	<0.001
	$n[18, 34)$	-1514.00	2.21	-686.24	<0.001
	$n[34, 59)$	-1210.00	2.35	-514.29	<0.001
	$n[59-)$	-309.20	2.74	-112.95	<0.001
	Familienstand: alleinstehend	-6.53	3.68	-1.77	0.08
	Familienstand: verwitwet	219.50	3.63	60.44	<0.001
	Familienstand: geschieden	-1139.00	3.67	-310.38	<0.001
	Familienstand: getrennt	-871.00	8.46	-102.92	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.23	0.00	186.35	<0.001
	Gewerbe: Landwirtschaft	-555.70	10.58	-52.51	<0.001
	Gewerbe: Energiesektor	3441.00	8.44	407.69	<0.001
	Gewerbe: Bergbau	518.60	15.71	33.02	<0.001
	Gewerbe: Produktion	1140.00	4.14	275.23	<0.001
	Gewerbe: Baugewerbe	952.90	4.38	217.50	<0.001
	Gewerbe: Handel	-5.08	4.69	-1.08	0.28
	Gewerbe: Transport	1516.00	5.87	258.34	<0.001
	Gewerbe: Bank/Versicherung	4091.00	8.37	488.56	<0.001
	Gewerbe: Dienstleistungen	974.40	3.90	249.60	<0.001
	Wohnverhältnis: Hauptmieter	-1026.00	2.34	-439.30	<0.001
	Wohnverhältnis: Untermieter	-1458.00	5.89	-247.58	<0.001
	Größe des Wohnsitzes	69870.00	83.84	833.37	<0.001
	(Größe des Wohnsitzes)^2	-24100.00	71.15	-338.74	<0.001
	Schulabschluss: Realschulabschluss	950.00	2.36	402.08	<0.001
	Schulabschluss: Abitur/FHSR	2065.00	2.93	705.10	<0.001
	Schulabschluss: andere Abschlüsse	-1388.00	5.70	-243.80	<0.001
	Schulabschluss: kein Abschluss	-1884.00	6.62	-284.52	<0.001
	n(Vollzeitbeschäftigung)	143900.00	160.60	896.28	<0.001
	n(Vollzeitbeschäftigung)^2	-31810.00	88.70	-358.60	<0.001
	n(Teilzeitbeschäftigung)	1106.00	3.22	343.69	0.00

Tabelle B.8: *Regression für Modell $P_{1,heteroskedastisch}$ mit heteroskedastischer Varianz (Erwartungswert) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sigma link		Schätzer	Std.fehler	t-Wert	P-Wert
log	(Achsenabschnitt)	16.59	0.00	10317.59	<0.001
	Geschlecht: weiblich	-0.04	0.00	-58.55	<0.001
	$n[0, 18)$	-0.12	0.00	-327.28	<0.001
	$n[18, 34)$	-0.07	0.00	-116.78	<0.001
	$n[34, 59)$	-0.02	0.00	-24.88	<0.001
	$n[59-)$	0.02	0.00	29.30	<0.001
	Familienstand: alleinstehend	0.21	0.00	212.27	<0.001
	Familienstand: verwitwet	0.10	0.00	99.82	<0.001
	Familienstand: geschieden	0.16	0.00	158.15	<0.001
	Familienstand: getrennt	0.36	0.00	182.40	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.00	0.00	52.12	<0.001
	Gewerbe: Landwirtschaft	0.20	0.00	89.85	<0.001
	Gewerbe: Energiesektor	-0.59	0.00	-182.30	<0.001
	Gewerbe: Bergbau	-0.15	0.01	-28.09	<0.001
	Gewerbe: Produktion	-0.05	0.00	-40.28	<0.001
	Gewerbe: Baugewerbe	-0.21	0.00	-162.15	<0.001
	Gewerbe: Handel	0.10	0.00	82.56	<0.001
	Gewerbe: Transport	-0.06	0.00	-34.30	<0.001
	Gewerbe: Bank/Versicherung	0.08	0.00	36.03	<0.001
	Gewerbe: Dienstleistungen	-0.04	0.00	-34.12	<0.001
	Wohnverhältnis: Hauptmieter	-0.12	0.00	-183.73	<0.001
	Wohnverhältnis: Untermieter	0.24	0.00	187.42	<0.001
	Größe des Wohnsitzes	0.00	0.00	485.43	<0.001
	Schulabschluss: Realschulabschluss	0.10	0.00	156.96	<0.001
	Schulabschluss: Abitur/FHSR	0.10	0.00	129.42	<0.001
	Schulabschluss: andere Abschlüsse	0.07	0.00	37.49	<0.001
	Schulabschluss: kein Abschluss	0.01	0.00	5.43	<0.001
	n(Vollzeitbeschäftigung)	0.05	0.00	52.02	<0.001
	n(Teilzeitbeschäftigung)	-0.02	0.00	-24.28	<0.001
				N	4493
				N_{gew}	25841690
				logL	-254812456
				aic	509625153

Tabelle B.9: *Regression für Modell $P_{1,heteroskedastisch}$ mit heteroskedastischer Varianz (Varianz) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

mu link		Schätzer	Std.fehler	t-Wert	P-Wert
Identität	(Achsenabschnitt)	6073.78	6.06	1002.28	<0.001
	Vorjahreseinkommen	0.62	0.00	3838.04	<0.001
	Geschlecht: weiblich	-42.19	1.95	-21.67	<0.001
	$n[0, 18)$	-317.92	1.05	-303.47	<0.001
	$n[18, 34)$	-103.17	1.89	-54.51	<0.001
	$n[34, 59)$	-3.96	2.00	-1.98	0.05
	$n[59-)$	145.84	2.25	64.73	<0.001
	Familienstand: alleinstehend	789.01	3.04	259.17	<0.001
	Familienstand: verwitwet	236.61	2.97	79.75	<0.001
	Familienstand: geschieden	-159.42	3.06	-52.15	<0.001
	Familienstand: getrennt	583.43	7.01	83.26	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.12	0.00	116.35	<0.001
	Gewerbe: Landwirtschaft	-355.67	7.46	-47.71	<0.001
	Gewerbe: Energiesektor	1489.68	7.20	206.94	<0.001
	Gewerbe: Bergbau	57.46	10.54	5.45	<0.001
	Gewerbe: Produktion	496.55	3.59	138.41	<0.001
	Gewerbe: Baugewerbe	325.63	3.72	87.53	<0.001
	Gewerbe: Handel	249.61	3.96	63.08	<0.001
	Gewerbe: Transport	604.77	4.78	126.47	<0.001
	Gewerbe: Bank/Versicherung	1496.52	6.64	225.24	<0.001
	Gewerbe: Dienstleistungen	300.14	3.37	89.14	<0.001
	Wohnverhältnis: Hauptmieter	-630.73	1.91	-330.89	<0.001
	Wohnverhältnis: Untermieter	-1152.71	4.91	-234.92	<0.001
	Größe des Wohnsitzes	23232.61	68.94	336.98	<0.001
	(Größe des Wohnsitzes)^2	-5038.07	59.68	-84.41	<0.001
	Schulabschluss: Realschulabschluss	222.21	1.85	119.96	<0.001
	Schulabschluss: Abitur/FHSR	639.56	2.55	250.52	<0.001
	Schulabschluss: andere Abschlüsse	-570.91	5.01	-113.91	<0.001
	Schulabschluss: kein Abschluss	-830.30	5.89	-141.01	<0.001
	n(Vollzeitbeschäftigung)	17876.42	140.00	127.72	<0.001
	n(Vollzeitbeschäftigung)^2	-8795.25	86.21	-102.02	<0.001
	n(Teilzeitbeschäftigung)	91.94	2.72	33.78	0.00

Tabelle B.10: *Regression für Modell $P_{6,heteroskedastisch}$ mit heteroskedastischer Varianz (Erwartungswert) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

sigma link		Schätzer	Std.fehler	t-Wert	P-Wert
log	(Achsenabschnitt)	16.38	0.00	9262.43	<0.001
	Geschlecht: weiblich	-0.07	0.00	-101.20	<0.001
	$n[0,18)$	-0.16	0.00	-392.17	<0.001
	$n[18,34)$	0.17	0.00	269.97	<0.001
	$n[34,59)$	-0.03	0.00	-42.98	<0.001
	$n[59-)$	-0.05	0.00	-70.18	<0.001
	Familienstand: alleinstehend	-0.02	0.00	-19.67	<0.001
	Familienstand: verwitwet	-0.10	0.00	-92.27	<0.001
	Familienstand: geschieden	0.05	0.00	42.97	<0.001
	Familienstand: getrennt	0.50	0.00	262.07	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.00	0.00	15.33	<0.001
	Gewerbe: Landwirtschaft	-0.25	0.00	-85.33	<0.001
	Gewerbe: Energiesektor	-0.45	0.00	-152.61	<0.001
	Gewerbe: Bergbau	-0.36	0.00	-78.58	<0.001
	Gewerbe: Produktion	-0.01	0.00	-6.74	<0.001
	Gewerbe: Baugewerbe	-0.25	0.00	-183.45	<0.001
	Gewerbe: Handel	0.04	0.00	32.77	<0.001
	Gewerbe: Transport	-0.30	0.00	-162.29	<0.001
	Gewerbe: Bank/Versicherung	0.07	0.00	27.85	<0.001
	Gewerbe: Dienstleistungen	-0.06	0.00	-50.87	<0.001
	Wohnverhältnis: Hauptmieter	-0.04	0.00	-60.68	<0.001
	Wohnverhältnis: Untermieter	0.30	0.00	199.84	<0.001
	Größe des Wohnsitzes	0.00	0.00	306.14	<0.001
	Schulabschluss: Realschulabschluss	-0.14	0.00	-191.89	<0.001
	Schulabschluss: Abitur/FHSR	0.14	0.00	160.58	<0.001
	Schulabschluss: andere Abschlüsse	0.06	0.00	35.00	<0.001
	Schulabschluss: kein Abschluss	0.02	0.00	10.71	<0.001
	n(Vollzeitbeschäftigung)	0.03	0.00	31.54	<0.001
	n(Teilzeitbeschäftigung)	-0.13	0.00	-130.94	<0.001
				N	4493
				N_{gew}	25841690
				logL	-249726288
				AIC	499452697

Tabelle B.11: *Regression für Modell $P_{6,heteroskedastisch}$ mit heteroskedastischer Varianz (Varianz) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

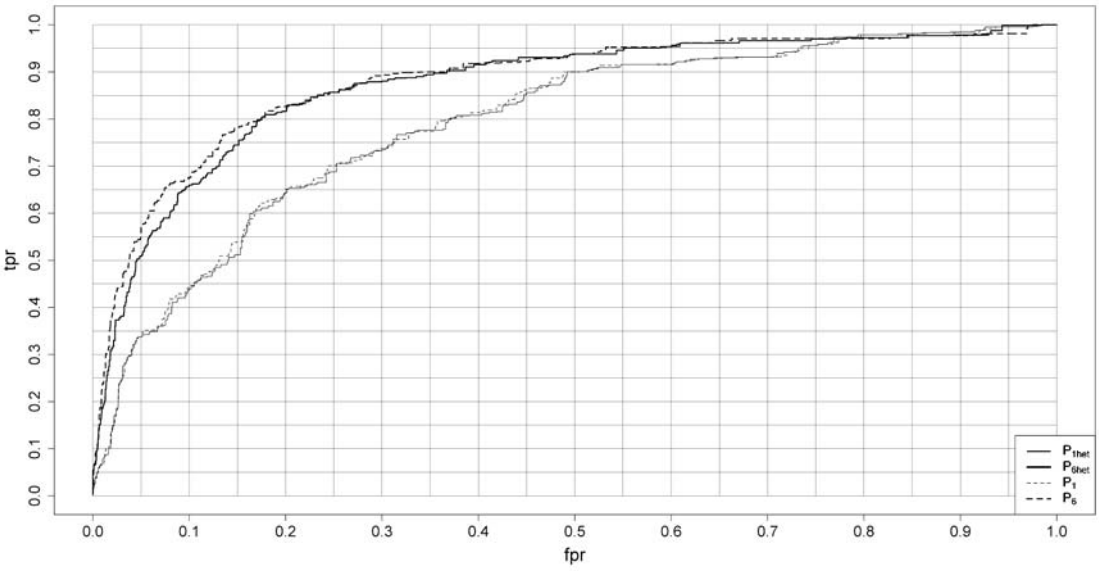
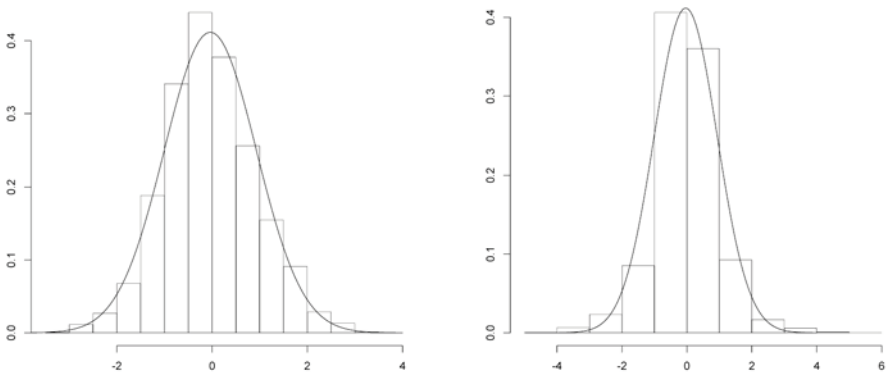


Abbildung B.4: *ROC für Schätzungen von Modellen mit heteroskedastischer und homoskedastischer Varianz [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

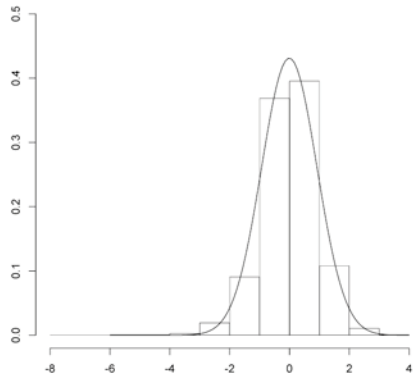
A.2.4 Quantilsresiduen weiterer berücksichtigter Verteilungsannahmen



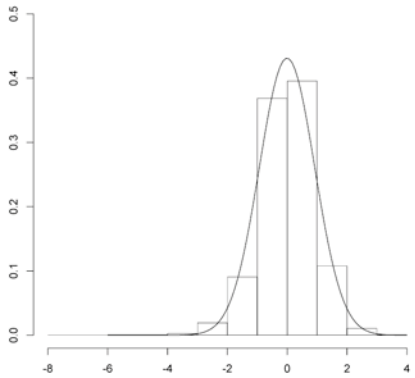
(a) Annahme t-Verteilung P_1 (b) Annahme t-Verteilung P_6

Abbildung B.5: *Quantilsresiduen von Regressionen unter Annahme einer t-Verteilung [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

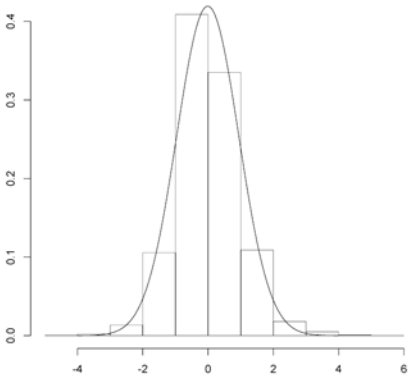
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



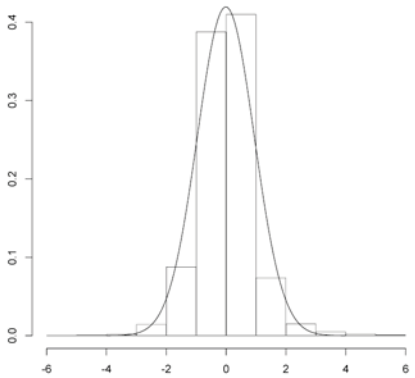
(a) Annahme einer Gamma-Verteilung P_1



(b) Annahme einer Gamma-Verteilung P_6



(c) Annahme einer Weibull-Verteilung P_1



(d) Annahme einer Weibull-Verteilung P_6

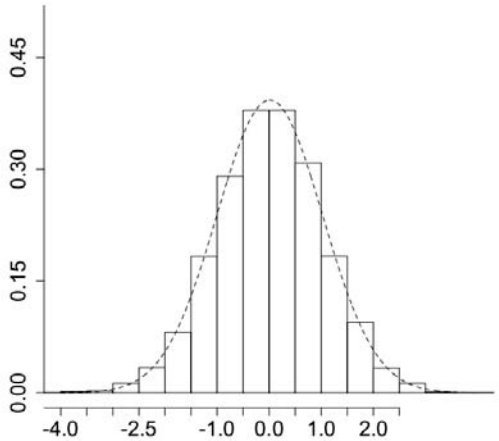
Abbildung B.6: *Quantilsresiduen von Regressionen mit verschiedenen Verteilungsannahmen [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

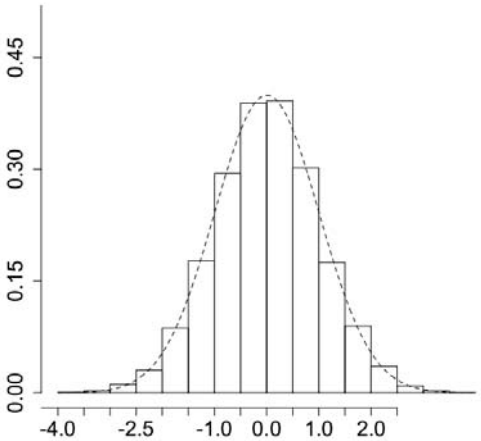
A.3 Ergänzungen zu den Modellen mit Klassen

A.3.1 Residuenplots für Modelle mit zwei Einkommensklassen

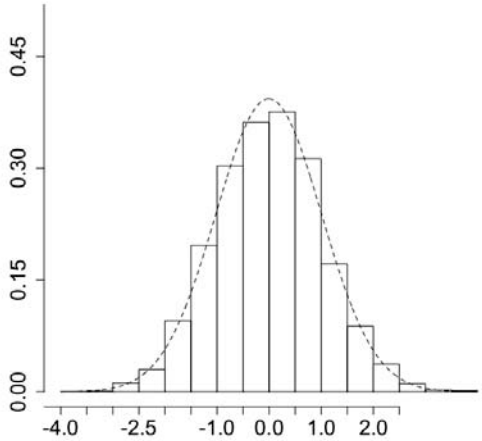
Aussagen über die Abbildungen C.7 werden im Hauptteil auf Seite 97 gemacht.



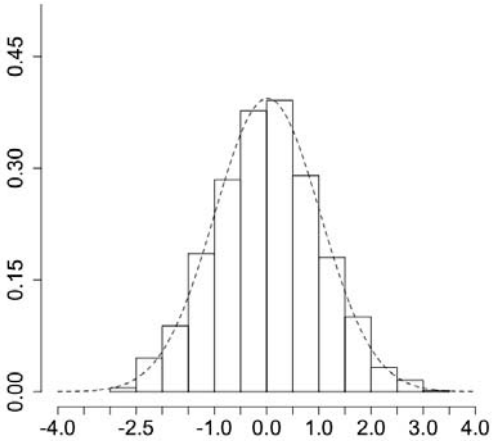
(a) M_{12} mit $L = 1.3z$



(b) M_{12} mit $L = 1.7z$



(c) M_{22} mit $L = 1.3z$ (obere Einkommensklasse)



(d) M_{22} mit $L = 1.3z$ (untere Einkommensklasse)

Abbildung C.7: *Quantilsresiduen für Regressionen mit zwei Einkommensklassen (Abschnitt 4.2.1) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.3.2 Ergebnisse der Regressionen mit drei Einkommensklassen

Link		Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-5.57	0.01	-843.46	<0.001
	Einkommen $[0, L_1)$	3.70	0.00	1062.74	<0.001
	Einkommen $[L_1, L_2)$	1.42	0.00	426.55	<0.001
	Geschlecht: weiblich	-0.15	0.00	-76.99	<0.001
	$n[0, 18)$	0.32	0.00	259.26	<0.001
	$n[18, 34)$	0.56	0.00	312.55	<0.001
	$n[34, 59)$	0.57	0.00	260.92	<0.001
	$n[59-)$	0.29	0.00	111.11	<0.001
	Familienstand: alleinstehend	0.94	0.00	264.87	<0.001
	Familienstand: verwitwet	0.44	0.00	119.81	<0.001
	Familienstand: geschieden	0.97	0.00	305.80	<0.001
	Familienstand: getrennt	0.49	0.00	101.46	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.00	0.00	-104.40	<0.001
	Gewerbe: Landwirtschaft	-0.62	0.01	-67.84	<0.001
	Gewerbe: Energiesektor	1.85	0.01	214.33	<0.001
	Gewerbe: Bergbau	-16.68	59.64	-0.28	0.78
	Gewerbe: Produktion	-0.47	0.00	-94.69	<0.001
	Gewerbe: Baugewerbe	-0.31	0.01	-56.88	<0.001
	Gewerbe: Handel	-0.11	0.00	-30.26	<0.001
	Gewerbe: Transport	-0.80	0.01	-125.23	<0.001
	Gewerbe: Bank/Versicherung	-0.61	0.01	-78.37	<0.001
	Gewerbe: Dienstleistungen	-0.28	0.00	-88.19	<0.001
	Wohnverhältnis: Hauptmieter	0.22	0.00	92.35	<0.001
	Wohnverhältnis: Untermieter	-0.50	0.01	-95.37	<0.001
	Größe des Wohnsitzes	-20.26	0.11	-179.93	<0.001
	(Größe des Wohnsitzes)^2	12.16	0.09	140.97	<0.001
	Schulabschluss: Realschulabschluss	0.17	0.00	78.76	<0.001
	Schulabschluss: Abitur/FHSR	0.19	0.00	71.03	<0.001
	Schulabschluss: andere Abschlüsse	0.43	0.00	136.69	<0.001
	Schulabschluss: kein Abschluss	0.71	0.00	186.72	<0.001
	n(Vollzeitbeschäftigung)	-28.66	0.18	-159.12	<0.001
	n(Vollzeitbeschäftigung)^2	4.75	0.13	36.59	<0.001
	n(Teilzeitbeschäftigung)	-0.16	0.00	-50.04	<0.001
				n (ungewichtet)	6095
				N (gewichtet)	25058614
				logL	-5325058
	Modellfamilie: binomial			aic	10650182

Tabelle C.12: *Regression für Modell $M_{13,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link		Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-0.21	0.01	-21.81	<0.001
	Geschlecht: weiblich	-0.39	0.00	-127.09	<0.001
	$n[0,18)$	0.40	0.00	206.76	<0.001
	$n[18,34)$	0.30	0.00	108.18	<0.001
	$n[34,59)$	0.54	0.00	151.38	<0.001
	$n[59-)$	0.27	0.00	61.52	<0.001
	Familienstand: alleinstehend	0.50	0.01	82.43	<0.001
	Familienstand: verwitwet	-0.06	0.01	-9.09	<0.001
	Familienstand: geschieden	0.32	0.01	59.50	<0.001
	Familienstand: getrennt	-0.37	0.01	-51.96	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.00	0.00	-141.94	<0.001
	Gewerbe: Landwirtschaft	-1.88	0.02	-121.01	<0.001
	Gewerbe: Energiesektor	18.25	103.90	0.18	0.86
	Gewerbe: Bergbau	-18.94	229.20	-0.08	0.93
	Gewerbe: Produktion	0.38	0.01	31.63	<0.001
	Gewerbe: Baugewerbe	0.98	0.01	73.45	<0.001
	Gewerbe: Handel	-0.09	0.01	-15.29	<0.001
	Gewerbe: Transport	-1.32	0.01	-136.29	<0.001
	Gewerbe: Bank/Versicherung	-1.14	0.01	-93.44	<0.001
	Gewerbe: Dienstleistungen	-0.21	0.00	-41.16	<0.001
	Wohnverhältnis: Hauptmieter	-0.24	0.00	-57.61	<0.001
	Wohnverhältnis: Untermieter	-0.93	0.01	-123.71	<0.001
	Größe des Wohnsitzes	-15.16	0.05	-279.79	<0.001
	(Größe des Wohnsitzes)^2	-0.47	0.05	-9.27	<0.001
	Schulabschluss: Realschulabschluss	-0.18	0.00	-53.05	<0.001
	Schulabschluss: Abitur/FHSR	0.37	0.00	81.07	<0.001
	Schulabschluss: andere Abschlüsse	0.16	0.01	30.69	<0.001
	Schulabschluss: kein Abschluss	0.30	0.01	58.37	<0.001
	n(Vollzeitbeschäftigung)	-8.15	0.07	-125.15	<0.001
	n(Vollzeitbeschäftigung)^2	-1.84	0.06	-31.32	<0.001
	n(Teilzeitbeschäftigung)	-0.50	0.01	-92.44	<0.001
				n (ungewichtet)	685
				N (gewichtet)	3153437
				logL	-1906426
	Modellfamilie: binomial			aic	3812895

Tabelle C.13: *Regression für Modell $M_{23,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ (untere Einkommensgruppe) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link		Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-4.45	0.01	-522.06	<0.001
	Geschlecht: weiblich	-0.10	0.00	-34.65	<0.001
	$n[0,18)$	0.37	0.00	209.87	<0.001
	$n[18,34)$	0.61	0.00	233.24	<0.001
	$n[34,59)$	0.40	0.00	127.30	<0.001
	$n[59-)$	0.11	0.00	28.32	<0.001
	Familienstand: alleinstehend	1.26	0.00	253.33	<0.001
	Familienstand: verwitwet	0.77	0.01	151.27	<0.001
	Familienstand: geschieden	1.08	0.00	236.63	<0.001
	Familienstand: getrennt	1.11	0.01	162.54	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.00	0.00	-30.24	<0.001
	Gewerbe: Landwirtschaft	1.67	0.01	148.33	<0.001
	Gewerbe: Energiesektor	2.67	0.01	212.50	<0.001
	Gewerbe: Bergbau	-16.43	75.58	-0.22	0.83
	Gewerbe: Produktion	-0.58	0.01	-89.45	<0.001
	Gewerbe: Baugewerbe	-0.51	0.01	-66.61	<0.001
	Gewerbe: Handel	0.10	0.01	20.28	<0.001
	Gewerbe: Transport	-0.72	0.01	-68.30	<0.001
	Gewerbe: Bank/Versicherung	-0.51	0.01	-35.01	<0.001
	Gewerbe: Dienstleistungen	-0.34	0.00	-73.96	<0.001
	Wohnverhältnis: Hauptmieter	0.23	0.00	67.73	<0.001
	Wohnverhältnis: Untermieter	-0.64	0.01	-77.82	<0.001
	Größe des Wohnsitzes	-11.61	0.10	-117.33	<0.001
	(Größe des Wohnsitzes)^2	6.41	0.08	83.30	<0.001
	Schulabschluss: Realschulabschluss	0.52	0.00	166.03	<0.001
	Schulabschluss: Abitur/FHSR	0.12	0.00	30.05	<0.001
	Schulabschluss: andere Abschlüsse	0.89	0.00	198.53	<0.001
	Schulabschluss: kein Abschluss	1.17	0.01	208.91	<0.001
	n(Vollzeitbeschäftigung)	-22.64	0.15	-146.70	<0.001
	n(Vollzeitbeschäftigung)^2	7.77	0.07	103.66	<0.001
	n(Teilzeitbeschäftigung)	0.13	0.00	29.35	<0.001
				n (ungewichtet)	2851
				N (gewichtet)	11791389
				logL	-2624240
	Modellfamilie: binomial			aic	5248542

Tabelle C.14: *Regression für Modell $M_{23,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ (mittlere Einkommensgruppe) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

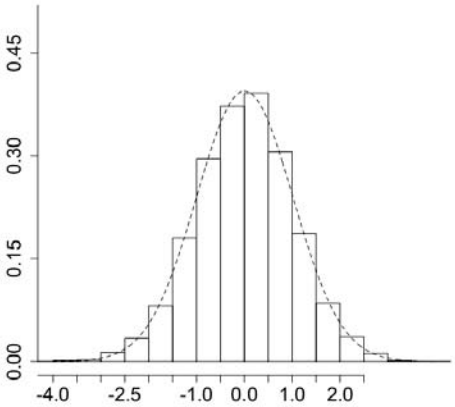
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link		Schätzer	Std.fehler	t-Wert	P-Wert
logit	(Achsenabschnitt)	-8.53	0.02	-381.66	<0.001
	Geschlecht: weiblich	0.85	0.01	113.08	<0.001
	$n[0,18)$	0.32	0.01	60.80	<0.001
	$n[18,34)$	1.11	0.01	169.03	<0.001
	$n[34,59)$	0.80	0.01	109.39	<0.001
	$n[59-)$	0.63	0.01	68.88	<0.001
	Familienstand: alleinstehend	0.38	0.01	29.25	<0.001
	Familienstand: verwitwet	-0.14	0.01	-9.76	<0.001
	Familienstand: geschieden	1.81	0.01	181.35	<0.001
	Familienstand: getrennt	-1.57	0.04	-37.69	<0.001
	Gesamtarbeitsstunden im Vorjahr	0.00	0.00	55.22	<0.001
	Gewerbe: Landwirtschaft	-18.99	137.00	-0.14	0.89
	Gewerbe: Energiesektor	-17.47	158.90	-0.11	0.91
	Gewerbe: Bergbau	-16.76	490.30	-0.03	0.97
	Gewerbe: Produktion	-0.61	0.02	-40.27	<0.001
	Gewerbe: Baugewerbe	-0.11	0.02	-6.70	<0.001
	Gewerbe: Handel	-1.51	0.02	-78.57	<0.001
	Gewerbe: Transport	0.83	0.01	56.14	<0.001
	Gewerbe: Bank/Versicherung	1.57	0.02	101.39	<0.001
	Gewerbe: Dienstleistungen	-0.03	0.01	-2.73	0.01
	Wohnverhältnis: Hauptmieter	2.20	0.01	236.34	<0.001
	Wohnverhältnis: Untermieter	-16.29	82.78	-0.20	0.84
	Größe des Wohnsitzes	18.96	0.20	95.14	<0.001
	(Größe des Wohnsitzes)^2	-0.76	0.16	-4.67	<0.001
	Schulabschluss: Realschulabschluss	-0.78	0.01	-93.10	<0.001
	Schulabschluss: Abitur/FHSR	-0.97	0.01	-104.78	<0.001
	Schulabschluss: andere Abschlüsse	-0.61	0.02	-31.92	<0.001
	Schulabschluss: kein Abschluss	0.57	0.03	22.30	<0.001
	n(Vollzeitbeschäftigung)	-69.51	0.53	-131.70	<0.001
	n(Vollzeitbeschäftigung)^2	-22.30	0.40	-55.62	<0.001
	n(Teilzeitbeschäftigung)	-1.01	0.01	-86.34	<0.001
				n (ungewichtet)	2559
				N (gewichtet)	10113788
				logL	-489506.4
	Modellfamilie: binomial			aic	979074.8

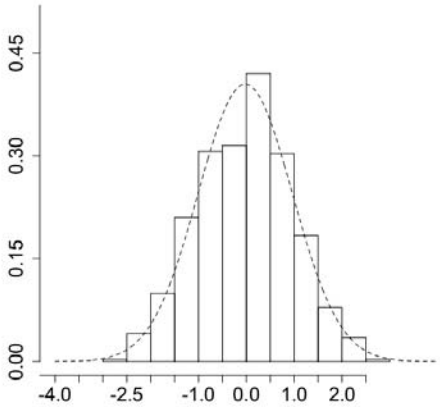
Tabelle C.15: *Regression für Modell $M_{23,06}$ mit $L_1 = 1.1z$ und $L_2 = 2z$ (obere Einkommensgruppe) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Erläuterungen zu den Regressionstabellen C.12-C.15 mit drei Einkommensklassen befinden sich im Hauptteil auf Seite 99.

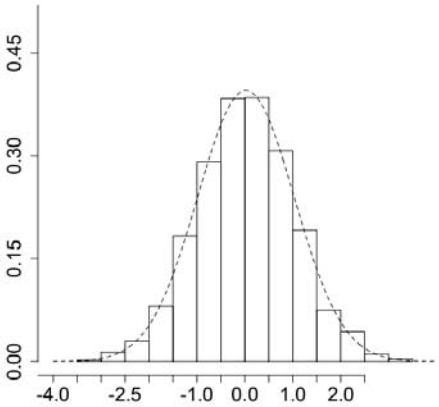
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



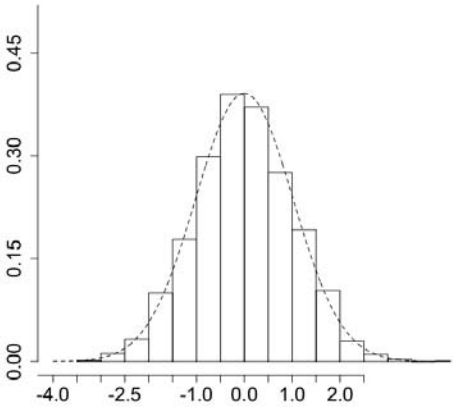
(a) M_{13} mit $L_1 = 1.1z$ und $L_2 = 2z$



(b) M_{21} mit $L_1 = 1.1z$ und $L_2 = 2z$ (niedriges Einkommen)



(c) M_{22} mit $L_1 = 1.1z$ und $L_2 = 2z$ (mittleres Einkommen)



(d) M_{23} mit $L_1 = 1.1z$ und $L_2 = 2z$ (hohes Einkommen)

Abbildung C.8: *Quantilsresiduen für Regressionen mit drei Einkommensklassen (Tabellen C.12-C.15) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.4 Armut und Vulnerabilität in Gruppen

A.4.1 Entwicklung von Anteilen und Armut nach Gewerbe des Haushaltsvorstands

Jahr	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
nicht-erwerbstätig	43.3	43.6	44.7	44.6	45.0	45.9	44.6	42.9	41.3	42.5	43.7	43.5	44.0	44.6	45.1	44.2	44.3
Landwirtschaft	1.1	0.9	0.8	0.9	0.8	0.8	0.7	0.7	0.9	0.8	0.9	0.7	0.7	0.7	0.6	0.5	0.5
Energie	0.9	1.0	0.9	0.8	0.8	0.8	0.8	0.7	0.5	0.5	0.5	0.5	0.7	0.6	0.6	0.5	0.4
Bergbau	0.4	0.3	0.3	0.2	0.2	0.2	0.2	0.3	0.3	0.2	0.2	0.1	0.1	0.1	0.1	0.0	0.0
Produktion	13.7	13.1	11.9	12.0	11.7	11.0	11.9	10.9	10.8	10.2	10.4	9.6	9.4	9.0	8.4	8.9	9.0
Baugewerbe	8.7	8.7	8.2	8.9	8.2	7.9	7.3	6.9	9.3	8.3	7.6	6.8	6.8	6.1	5.8	6.4	6.0
Handel	7.1	6.9	6.9	7.0	6.9	7.6	7.5	7.9	8.3	8.1	7.4	7.0	8.0	7.5	6.9	7.9	7.6
Transport	3.1	3.0	3.0	3.2	3.0	3.0	2.9	2.9	3.2	3.2	3.0	2.6	3.0	2.7	2.5	2.8	2.8
Bank,Versicherung	2.1	2.2	2.1	2.4	2.4	2.5	2.4	2.4	2.6	2.6	2.4	2.4	2.5	2.4	2.3	2.3	2.5
Dienstleistungen	16.2	17.2	16.3	17.5	16.5	18.1	19.6	19.2	21.1	20.1	20.6	19.9	20.5	19.6	20.0	22.3	21.6
keine Antwort	3.5	3.1	4.9	2.4	4.5	2.2	2.1	5.2	1.7	3.4	3.3	7.0	4.2	6.9	7.9	4.0	5.2

Tabelle D.16: *Anteile der Haushalte nach Beruf des Haushaltsvorstands in % (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Sektor/Jahr	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
nicht-erwerbstätig	17.2	14.2	16.2	14.5	13.2	13.5	12.5	10.8	12.1	12.9	12.8	13.5	13.0	13.9	15.1	15.7	15.6
Landwirtschaft	7.8	12.7	7.9	3.3	5.2	4.9	9.5	11.3	6.8	7.7	6.7	7.6	9.7	6.1	2.0	3.0	12.4
Energie	4.0	1.7	3.3	0.0	0.0	0.0	0.0	0.0	0.0	3.8	4.4	0.0	1.1	3.2	9.0	0.0	1.5
Bergbau	1.1	0.3	3.1	10.2	8.1	6.6	0.0	0.0	3.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Produktion	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Baugewerbe	1.6	1.7	5.3	3.9	3.3	1.8	3.8	3.9	3.3	2.6	4.8	4.2	3.4	5.3	5.4	1.8	3.9
Handel	6.2	4.2	8.5	6.5	5.7	6.8	5.1	3.9	6.4	8.2	6.2	7.2	8.9	7.3	7.7	8.4	8.7
Transport	2.9	2.3	1.5	3.5	3.8	1.7	1.6	4.3	4.8	3.6	1.1	2.7	5.0	2.9	3.7	2.7	7.0
Bank, Versicherung	0.0	0.9	2.8	1.0	3.9	1.2	0.0	1.9	1.9	1.8	0.2	0.3	5.1	1.2	0.3	0.2	2.4
Dienstleistungen	1.9	2.5	1.6	3.3	2.3	2.0	3.7	2.1	3.3	3.7	3.5	3.9	4.6	4.9	4.5	5.7	4.6

Tabelle D.17: *Bedingte Armutsquoten der Haushalte nach Berufsgruppe des Haushaltsvorstands in % (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Sektor/Jahr	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08
nicht-erwerbstätig	7.4	6.2	7.2	6.5	5.9	6.2	5.6	4.6	5.0	5.5	5.6	5.9	5.7	6.2	6.8	6.9	6.9
Landwirtschaft	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1
Energie	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Bergbau	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Produktion	0.5	0.3	0.3	0.3	0.3	0.2	0.4	0.1	0.4	0.3	0.3	0.2	0.1	0.2	0.1	0.2	0.1
Baugewerbe	0.1	0.1	0.4	0.3	0.3	0.1	0.3	0.3	0.3	0.2	0.4	0.3	0.2	0.3	0.3	0.1	0.2
Handel	0.4	0.3	0.6	0.5	0.4	0.5	0.4	0.3	0.5	0.7	0.5	0.5	0.7	0.5	0.5	0.7	0.7
Transport	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.2	0.1	0.0	0.1	0.2	0.1	0.1	0.1	0.2
Bank, Versicherung	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
Dienstleistungen	0.3	0.4	0.3	0.6	0.4	0.4	0.7	0.4	0.7	0.7	0.7	0.8	0.9	1.0	0.9	1.3	1.0

Tabelle D.18: *Anteil der Armut von Berufsgruppen in % (1992-2008) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

A.4.2 Kovariablen zur Vulnerabilitäts-Schätzung in Gruppen für 2005

Dokumentiert werden nur die Veränderungen in der Wahl der Kovariablen zu P_1 bzw. P_6. Das Merkmal, nach dem die Gruppen gebildet werden, wird nicht in die Regressionsmodelle aufgenommen. Migrationshintergrund des HVs wird trotz der Betrachtung als Merkmal in keine der Regressionen aufgenommen, da diese keine Verbesserungen der Modellanpassung und der ROC ergaben. Tabelle D.19 listet die Besonderheiten der Kovariablen in den Regressionen der Gruppen auf. Für das Merkmal Geschlecht wurde im Gegensatz zu den übrigen Gruppen auch die Variable Region in die Modelle aufgenommen. Der Grund dafür ist, dass die Region in den Gruppen des Merkmals Geschlecht eine Wirkung hat. Der Faktor Gewerbe beinhaltet viele Faktorstufen, bei denen einige schon in der Grundgesamtheit sehr dünn besetzt sind, so dass dies bei der Betrachtung von einzelnen Gruppen ein noch größeres Problem darstellt. Aus diesem Grund werden für den Faktor Gewerbe dünn besetzte Stufen zusammengelegt. Da die Datenlage für die Gruppierungen nach verschiedenen Merkmalen unterschiedlich ist, werden unterschiedliche Zusammenlegungen für verschiedene Merkmale gewählt. Im Falle der Gruppierung nach Alter des HVs wird nicht mehr das Gewerbe sondern nur noch die Erwerbstätigkeit betrachtet, da in der ältesten Altersgruppe ein Großteil der HVs bereits in Rente ist und so alle Stufen mit Ausnahme der Stufe „nicht- erwerbstätig“ dünn besetzt wären. Außerdem wird bei der Gruppierung nach Alter die Variable Anzahl der Haushaltsmitglieder älter als 65 nicht in das Modell aufgenommen, da diese Variable hoch mit der Variable der Erwerbstätigkeit korreliert ist. Bei der Gruppierung

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

nach Erwerbstätigkeit wurden die Variablen Geschlecht und Eigentümer des Wohnsitzes nicht ins Modell aufgenommen, da diese Variablen keine große Wirkung auf das Einkommen haben. Weiterhin wurden die Variablen Anzahl der Haushaltsmitglieder in Voll- und Teilzeitbeschäftigung nicht berücksichtigt. Bei der Gruppierung nach Erziehung konnten, aufgrund der Gruppengröße, Faktoren mit vielen Faktorstufen nicht berücksichtigt werden (Gewerbe, Status des Wohneigentums, Bildung). Ebenfalls weggelassen wurde die Kovariable Anzahl der Vollzeitbeschäftigten, da es im Haushalt maximal einen Erwerbstätigen geben kann, allerdings häufig wegen der Kinderbetreuung nicht in Vollzeit. Weiterhin wurden nur die Variable Anzahl der Kinder aufgenommen, da diese Information von besonderem Interesse für die Schätzung der Vulnerabilität ist. Da für alle Ausprägungen eines Merkmals dieselben Kovariablen verwendet werden, werden die Kovariablen der *Merkmale* notiert. Es soll noch einmal betont werden, dass das Einkommen für die Ausprägungen der Merkmale getrennt geschätzt wird.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

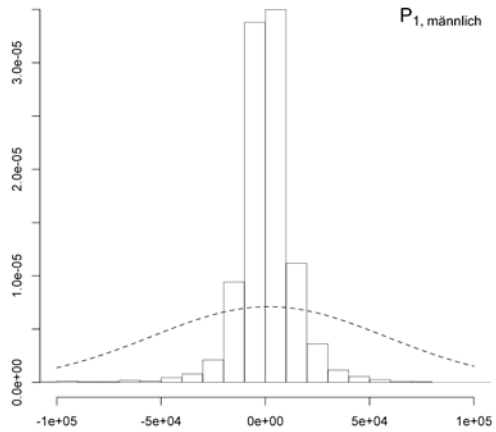
Gruppe	**Besonderheit**
Geschlecht des HVs	mit Kovariable Bundesland; Gewerbe Landwirtschaft und Bergbau zu einer Faktorstufe zusammengefasst; *ohne Variable Geschlecht*
Alter des HVs	Variable Erwerbstätigkeit (Nein, Ja) anstelle von Gewerbe; ohne Variable Alterstruktur älter als 65
Familienstand des HVs	Gewerbe Landwirtschaft, Energie, Bergbau und Produktion zur Faktorstufe andere Gewerbe zusammengefasst; *ohne Kovariable Familienstand*
Erwerbstätigkeit (Nein/Ja)	ohne Kovariablen Geschlecht und Eigentümer des Wohnsitzes, *ohne Kovariablen Gewerbe, Anzahl der Voll- und Teilzeitbeschäftigten*
Bildung	*ohne Kovariable Bildung*
Region	Stufen von Landwirtschaft und Bergbau zu einer Stufe zusammengelegt
Kinder	*ohne die Gruppe der Kinder*
Erziehung: allein	ohne Kovariablen Gewerbe, n(Vollzeit)^2, Bildung, Status des Wohnsitzes, nur Altersklasse [0,18)
Migrationshintergrund (Nein/Ja)	-

Tabelle D.19: *Besonderheiten der Kovariablen in den Regressionen für einzelne Merkmale*

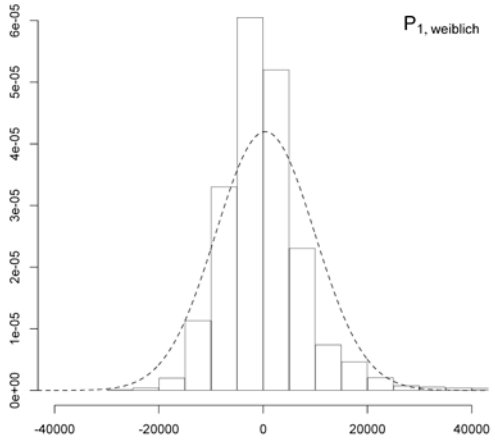
A.4.3 Modellwahl für die Schätzung in Gruppen

In den Abbildungen werden exemplarisch die Histogramme der Residuen von den Modellen $P_{1,G}$ und $P_{6,G}$ mit/ohne Logarithmustransformation des Einkommens für die Gruppen Männer und Frauen dargestellt. Der Hauptteil referiert zu diesen Abbildungen auf Seite 142.

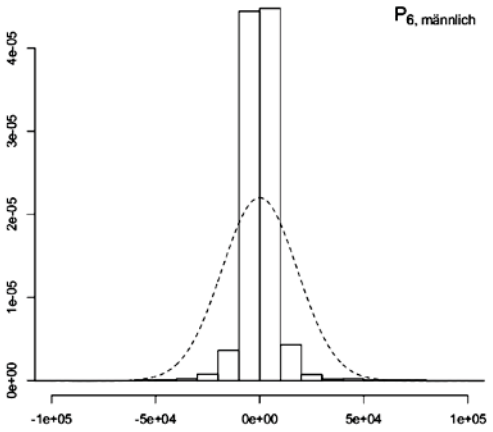
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



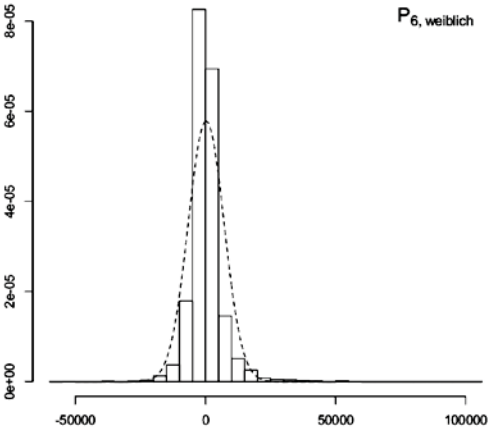
(a) Histogramm der Residuen für $P_{1,\text{männlich}}$



(b) Histogramm der Residuen für $P_{1,\text{weiblich}}$



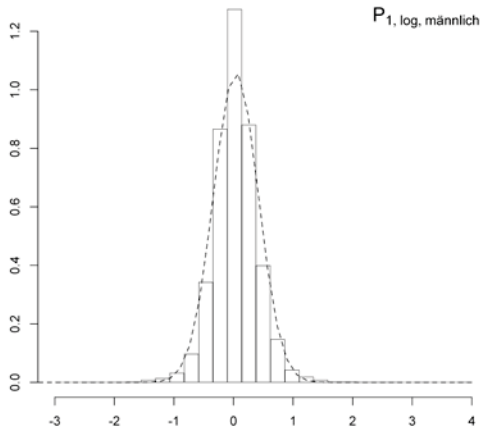
(c) Histogramm der Residuen für $P_{6,\text{männlich}}$



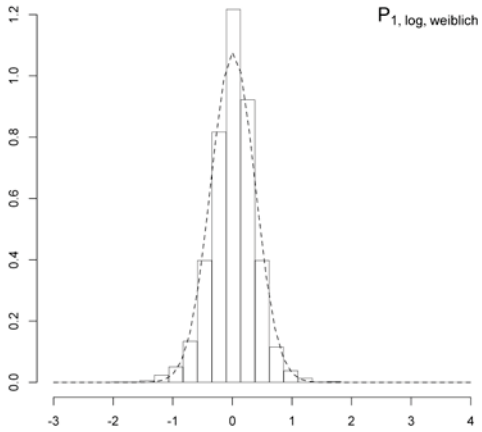
(d) Histogramm der Residuen für $P_{6,\text{weiblich}}$

Abbildung D.9: *Residuen der Regressionen für das Merkmal Geschlecht [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

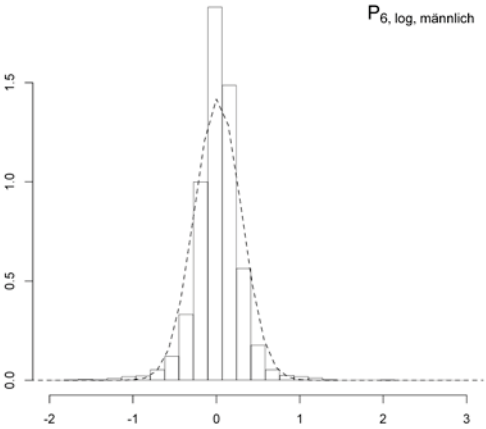
Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



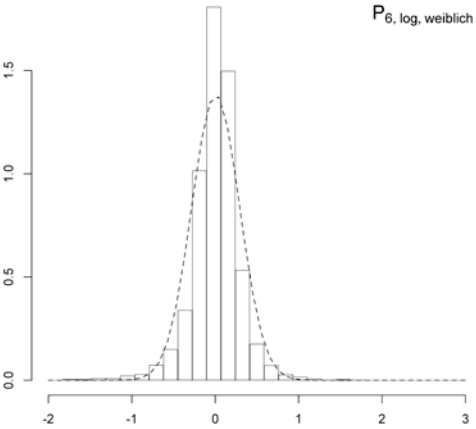
(a) Histogramm der Residuen für $P_{1,log,\text{männlich}}$



(b) Histogramm der Residuen für $P_{1,log,\text{weiblich}}$



(c) Histogramm der Residuen für $P_{6,log,\text{männlich}}$



(d) Histogramm der Residuen für $P_{6,log,\text{weiblich}}$

Abbildung D.10: *Residuen der Regressionen mit transformiertem Einkommen für das Merkmal Geschlecht [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A.4.4 Schätzung der Vulnerabilität für das Geschlecht für 2005

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	10.08	0.05	193.82	<0.001
Region: Schleswig Holstein	0.02	0.04	0.46	0.646
Region: Hamburg	0.05	0.05	0.92	0.357
Region: Niedersachsen	-0.03	0.04	-0.86	0.388
Region: Bremen	0.11	0.05	2.01	0.045
Region: Nordrhein-Westfalen	0.02	0.04	0.48	0.632
Region: Hessen	0.00	0.04	0.11	0.911
Region: Rheinland Pfalz/Saarland	0.01	0.04	0.37	0.715
Region: Baden-Wuerttemberg	0.05	0.04	1.32	0.188
Region: Bayern	0.02	0.04	0.57	0.566
Region: Berlin Ost	-0.08	0.05	-1.49	0.137
Region: Mecklenburg-Vorpommern	-0.14	0.05	-2.66	0.008
Region: Brandenburg	-0.15	0.04	-3.36	0.001
Region: Sachsen-Anhalt	-0.20	0.05	-4.32	<0.001
Region: Thüringen	-0.18	0.04	-3.97	<0.001
Region: Sachsen	-0.22	0.04	-5.48	<0.001
$n[0,18)$	-0.14	0.01	-17.64	<0.001
$n[18,34)$	-0.22	0.01	-18.98	<0.001
$n[34,59)$	-0.16	0.01	-11.10	<0.001
$n[59-)$	-0.04	0.02	-2.53	0.011
Familienstand: alleinstehend	-0.03	0.02	-1.30	0.194
Familienstand: verwitwet	0.17	0.03	6.10	<0.001
Familienstand: geschieden	-0.02	0.02	-0.97	0.330
Familienstand: getrennt	-0.00	0.04	-0.10	0.918
Gesamtarbeitsstunden im Vorjahr	0.00	0.00	8.71	<0.001
Gewerbe: Landwirtschaft	-0.15	0.05	-3.23	0.001
Gewerbe: Energiesektor	0.02	0.05	0.33	0.738
Gewerbe: Produktion	0.01	0.02	0.34	0.735
Gewerbe: Baugewerbe	0.02	0.03	0.89	0.375
Gewerbe: Handel	-0.05	0.03	-1.70	0.089
Gewerbe: Transport	-0.04	0.03	-1.10	0.273
Gewerbe: Bank/Versicherung	0.02	0.04	0.58	0.562
Gewerbe: Dienstleistungen	0.04	0.02	1.75	0.081
Wohnverhältnis: Hauptmieter	-0.12	0.01	-8.90	<0.001
Wohnverhältnis: Untermieter	-0.20	0.04	-5.04	<0.001
Größe des Wohnsitzes	9.54	0.54	17.74	<0.001
(Größe des Wohnsitzes)^2	-0.18	0.43	-0.41	0.681
Schulabschluss: Realschulabschluss	0.15	0.01	10.41	<0.001
Schulabschluss: Abitur/FHSR	0.31	0.01	21.28	<0.001
Schulabschluss: andere Abschlüsse	-0.03	0.02	-1.37	0.170
Schulabschluss: kein Abschluss	-0.14	0.04	-3.78	<0.001
n(Vollzeitbeschäftigung)	10.87	1.03	10.52	<0.001
n(Vollzeitbeschäftigung)^2	-5.63	0.52	-10.72	<0.001
n(Teilzeitbeschäftigung)	0.03	0.02	1.89	0.059
adjustiertes Betimmtheitsmaß	0.467			
Anz. Beobachtungen (ungewichtet)	5324			
F-Statitsik	106.7	(df: 43,5155)		

Tabelle D.20: *Ergebnisse der Regression für Modell $P_{1,männlich}$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	9.79	0.06	163.09	<0.001
Region: Schleswig Holstein	-0.04	0.05	-0.85	0.396
Region: Hamburg	-0.10	0.06	-1.69	0.091
Region: Niedersachsen	-0.13	0.05	-2.76	0.006
Region: Bremen	0.00	0.08	0.01	0.993
Region: Nordrhein-Westfalen	0.00	0.04	0.03	0.979
Region: Hessen	-0.06	0.05	-1.28	0.200
Region: Rheinland Pfalz/Saarland	-0.10	0.05	-2.19	0.028
Region: Baden-Wuerttemberg	0.01	0.04	0.29	0.769
Region: Bayern	-0.03	0.04	-0.61	0.544
Region: Berlin Ost	-0.16	0.06	-2.67	0.008
Region: Mecklenburg-Vorpommern	-0.35	0.06	-6.27	<0.001
Region: Brandenburg	-0.18	0.05	-3.40	0.001
Region: Sachsen-Anhalt	-0.23	0.05	-4.27	<0.001
Region: Thüringen	-0.15	0.05	-2.84	0.004
Region: Sachsen	-0.17	0.05	-3.55	<0.001
$n[0, 18)$	-0.11	0.01	-10.21	<0.001
$n[18, 34)$	-0.15	0.02	-9.37	<0.001
$n[34, 59)$	-0.05	0.02	-3.26	0.001
$n[59-)$	0.06	0.02	2.65	0.008
Familienstand: alleinstehend	-0.06	0.02	-2.43	0.015
Familienstand: verwitwet	0.08	0.02	3.12	0.002
Familienstand: geschieden	-0.15	0.02	-6.21	<0.001
Familienstand: getrennt	-0.12	0.04	-2.72	0.007
Gesamtarbeitsstunden im Vorjahr	0.00	0.00	5.29	<0.001
Gewerbe: Landwirtschaft	-0.02	0.14	-0.14	0.888
Gewerbe: Energiesektor	0.12	0.13	0.94	0.350
Gewerbe: Produktion	0.18	0.04	4.86	<0.001
Gewerbe: Baugewerbe	0.24	0.05	4.78	<0.001
Gewerbe: Handel	0.14	0.03	4.56	<0.001
Gewerbe: Transport	0.11	0.06	1.82	0.070
Gewerbe: Bank/Versicherung	0.26	0.05	5.40	<0.001
Gewerbe: Dienstleistungen	0.17	0.03	6.81	<0.001
Wohnverhältnis: Hauptmieter	-0.13	0.02	-7.69	<0.001
Wohnverhältnis: Untermieter	-0.13	0.04	-3.32	0.001
Größe des Wohnsitzes	6.70	0.57	11.73	<0.001
(Größe des Wohnsitzes)^2	-0.74	0.50	-1.48	0.139
Schulabschluss: Realschulabschluss	0.16	0.02	8.97	<0.001
Schulabschluss: Abitur/FHSR	0.36	0.02	17.65	<0.001
Schulabschluss: andere Abschlüsse	-0.04	0.03	-1.03	0.301
Schulabschluss: kein Abschluss	-0.02	0.05	-0.35	0.726
n(Vollzeitbeschäftigung)	4.72	1.14	4.14	<0.001
n(Vollzeitbeschäftigung)^2	-1.23	0.51	-2.40	0.017
n(Teilzeitbeschäftigung)	-0.06	0.03	-2.07	0.039
adjustiertes Betimmtheitsmaß	0.442			
Anz. Beobachtungen (ungewichtet)	3269			
F-Statitsik	60.48	(df: 43,3192)		

Tabelle D.21: *Ergebnisse der Regression für Modell $P_{1,weiblich}$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	4.07	0.11	38.38	<0.001
log(Vorjahreseinkommen)	0.59	0.01	60.52	<0.001
Region: Schleswig Holstein	0.04	0.03	1.15	0.250
Region: Hamburg	0.05	0.04	1.23	0.219
Region: Niedersachsen	0.01	0.03	0.18	0.856
Region: Bremen	0.08	0.04	1.81	0.070
Region: Nordrhein-Westfalen	0.02	0.03	0.75	0.454
Region: Hessen	-0.01	0.03	-0.34	0.733
Region: Rheinland Pfalz/Saarland	0.05	0.03	1.68	0.094
Region: Baden-Wuerttemberg	0.05	0.03	1.73	0.084
Region: Bayern	0.03	0.03	0.95	0.344
Region: Berlin Ost	-0.06	0.04	-1.59	0.111
Region: Mecklenburg-Vorpommern	-0.02	0.04	-0.54	0.586
Region: Brandenburg	-0.05	0.03	-1.52	0.128
Region: Sachsen-Anhalt	-0.11	0.04	-2.96	0.003
Region: Thüringen	-0.08	0.04	-2.23	0.026
Region: Sachsen	-0.08	0.03	-2.41	0.016
$n[0, 18)$	-0.05	0.01	-7.47	<0.001
$n[18, 34)$	-0.08	0.01	-8.99	<0.001
$n[34, 59)$	-0.06	0.01	-5.57	<0.001
$n[59-)$	-0.03	0.01	-2.34	0.019
Familienstand: alleinstehend	0.00	0.02	0.20	0.845
Familienstand: verwitwet	0.08	0.02	3.39	0.001
Familienstand: geschieden	0.00	0.02	0.13	0.895
Familienstand: getrennt	-0.03	0.03	-0.92	0.357
Gesamtarbeitsstunden im Vorjahr	0.00	0.00	6.71	<0.001
Gewerbe: Landwirtschaft	-0.03	0.04	-0.67	0.505
Gewerbe: Energiesektor	-0.05	0.05	-1.03	0.304
Gewerbe: Produktion	0.01	0.02	0.27	0.788
Gewerbe: Baugewerbe	0.01	0.02	0.47	0.640
Gewerbe: Handel	-0.06	0.02	-2.41	0.016
Gewerbe: Transport	-0.01	0.03	-0.47	0.639
Gewerbe: Bank/Versicherung	-0.06	0.03	-1.86	0.063
Gewerbe: Dienstleistungen	0.04	0.02	1.92	0.055
Wohnverhältnis: Hauptmieter	-0.04	0.01	-3.68	<0.001
Wohnverhältnis: Untermieter	-0.06	0.03	-1.99	0.046
Größe des Wohnsitzes	3.42	0.44	7.80	<0.001
(Größe des Wohnsitzes)^2	0.00	0.32	0.00	0.996
Schulabschluss: Realschulabschluss	0.07	0.01	6.17	<0.001
Schulabschluss: Abitur/FHSR	0.15	0.01	12.87	<0.001
Schulabschluss: andere Abschlüsse	-0.02	0.02	-1.05	0.296
Schulabschluss: kein Abschluss	-0.07	0.03	-2.19	0.029
n(Vollzeitbeschäftigung)	1.71	0.81	2.10	0.036
n(Vollzeitbeschäftigung)^2	-1.61	0.47	-3.40	0.001
n(Teilzeitbeschäftigung)	-0.00	0.01	-0.10	0.918
adjustiertes Betimmtheitsmaß	0.672			
Anz. Beobachtungen (ungewichtet)	5324			
F-Statitsik	345.1	(df: 44,5206)		

Tabelle D.22: *Ergebnisse der Regression für Modell $P_{6,männlich}$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

	Schätzer	Std.fehler	t-Wert	P-Wert
(Achsenabschnitt)	3.95	0.13	31.22	<0.001
log(Vorjahreseinkommen)	0.61	0.01	48.07	<0.001
Region: Schleswig Holstein	-0.06	0.04	-1.51	0.131
Region: Hamburg	-0.19	0.04	-4.45	<0.001
Region: Niedersachsen	-0.10	0.03	-2.88	0.004
Region: Bremen	-0.00	0.07	-0.04	0.964
Region: Nordrhein-Westfalen	-0.05	0.03	-1.75	0.080
Region: Hessen	-0.03	0.04	-0.85	0.396
Region: Rheinland Pfalz/Saarland	-0.07	0.04	-2.05	0.041
Region: Baden-Wuerttemberg	-0.07	0.03	-2.02	0.044
Region: Bayern	-0.05	0.03	-1.56	0.118
Region: Berlin Ost	-0.12	0.05	-2.57	0.010
Region: Mecklenburg-Vorpommern	-0.25	0.04	-6.12	<0.001
Region: Brandenburg	-0.17	0.04	-4.18	<0.001
Region: Sachsen-Anhalt	-0.16	0.04	-3.97	<0.001
Region: Thüringen	-0.14	0.04	-3.40	0.001
Region: Sachsen	-0.11	0.04	-2.95	0.003
$n[0, 18)$	-0.05	0.01	-6.19	<0.001
$n[18, 34)$	-0.06	0.01	-4.49	<0.001
$n[34, 59)$	-0.05	0.01	-3.92	<0.001
$n[59-)$	-0.03	0.02	-1.78	0.075
Familienstand: alleinstehend	-0.06	0.02	-3.09	0.002
Familienstand: verwitwet	-0.01	0.02	-0.53	0.598
Familienstand: geschieden	-0.09	0.02	-4.74	<0.001
Familienstand: getrennt	-0.02	0.03	-0.56	0.577
Gesamtarbeitsstunden im Vorjahr	0.00	0.00	3.43	0.001
Gewerbe: Landwirtschaft	-0.10	0.10	-1.04	0.299
Gewerbe: Energiesektor	0.01	0.10	0.08	0.934
Gewerbe: Produktion	0.00	0.03	0.00	0.998
Gewerbe: Baugewerbe	0.07	0.04	1.70	0.090
Gewerbe: Handel	-0.03	0.02	-1.20	0.230
Gewerbe: Transport	-0.09	0.05	-2.01	0.044
Gewerbe: Bank/Versicherung	0.04	0.04	1.01	0.311
Gewerbe: Dienstleistungen	0.00	0.02	0.20	0.845
Wohnverhältnis: Hauptmieter	-0.03	0.01	-2.18	0.029
Wohnverhältnis: Untermieter	0.10	0.03	3.39	0.001
Größe des Wohnsitzes	2.24	0.44	5.04	<0.001
(Größe des Wohnsitzes)^2	0.17	0.37	0.45	0.654
Schulabschluss: Realschulabschluss	0.03	0.01	2.00	0.046
Schulabschluss: Abitur/FHSR	0.14	0.02	8.61	<0.001
Schulabschluss: andere Abschlüsse	-0.08	0.03	-3.01	0.003
Schulabschluss: kein Abschluss	0.01	0.04	0.37	0.711
n(Vollzeitbeschäftigung)	1.71	0.85	2.01	0.044
n(Vollzeitbeschäftigung)^2	0.11	0.41	0.26	0.798
n(Teilzeitbeschäftigung)	0.02	0.02	1.11	0.265
adjustiertes Betimmtheitsmaß	0.666			
Anz. Beobachtungen (ungewichtet)	3215			
F-Statitsik	145.2	(df: 44,3139)		

Tabelle D.23: *Ergebnisse der Regression für Modell $P_{6,weiblich}$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Die Ergebnisse der Regressionstabellen werden im Hauptteil auf den Seiten 143 f. näher erläutert.

A.4.5 Historische VPLs für die Gruppen Erwerbstätigkeit und Anzahl der Kinder im Haushalt

Die VPLs für die Gruppierung der Haushalte nach Erwerbstätigkeit des HVs wurden bei 9.7, entspricht 16 317 €, und 9.5, entspricht etwa 13 360 €, (Schätzung mit $P_{6,G}$) fixiert. Es ist zu beobachten, dass der Verlauf der historischen VPLs zu denen in der ersten Studie relativ ähnlich ist (siehe Abbildungen 4.9 und 4.10). Es scheint so, als ob die getrennte Schätzung der beiden Gruppen keine großen Auswirkungen auf die VPLs bei 80% haben. In dieser Analyse werden die VPLs etwas höher gewählt als für die Modelle P_1 (15 500 €) und P_6 (13 000 €). Die Abbildungen und die Bewertung der Genauigkeit mit den fixierten VPLs für Vulnerabilitäts-Schätzer der Modelle für das Merkmal Erwerbstätigkeit befinden sich auf Seite 152 ff.

Abbildung D.11: *Historische VPLs für Schätzungen nach Merkmal Erwerbstätigkeit für den Zeitraum 1994-2008 (Methode 1) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Die VPLs für die Gruppierung der Haushalte nach Anzahl der Kinder wurden mit 9.7; 9.6; 9.6 und 9.4 (Schätzung mit $P_{1,keinKind}$, $P_{1,1Kind}$, $P_{1,2Kinder}$, $P_{1,>3Kinder}$) und 9.5; 9.5; 9.5 und 9.3 (Schätzung mit $P_{6,keinKind}$,$P_{6,1Kind}$, $P_{6,2Kinder}$, $P_{6,>3Kinder}$) fixiert. Es zeigt sich, dass die historischen VPLs für die einzelnen Gruppen (Metho-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

de 2) deutlich volatiler sind als die von Methode 1 bzw. die der Studie 1. Dies gilt insbesondere für die Gruppe *3 oder mehr Kinder*. Eine Ursache dafür könnte darin liegen, dass diese Gruppe sehr klein ist und daher Veränderungen in der Schätzung zu größeren Veränderungen bei den historischen VPLs führen. Mögliche weitere Ursachen könnten politisch bedingt sein (z.B. die Höhe des Kindergeldes).

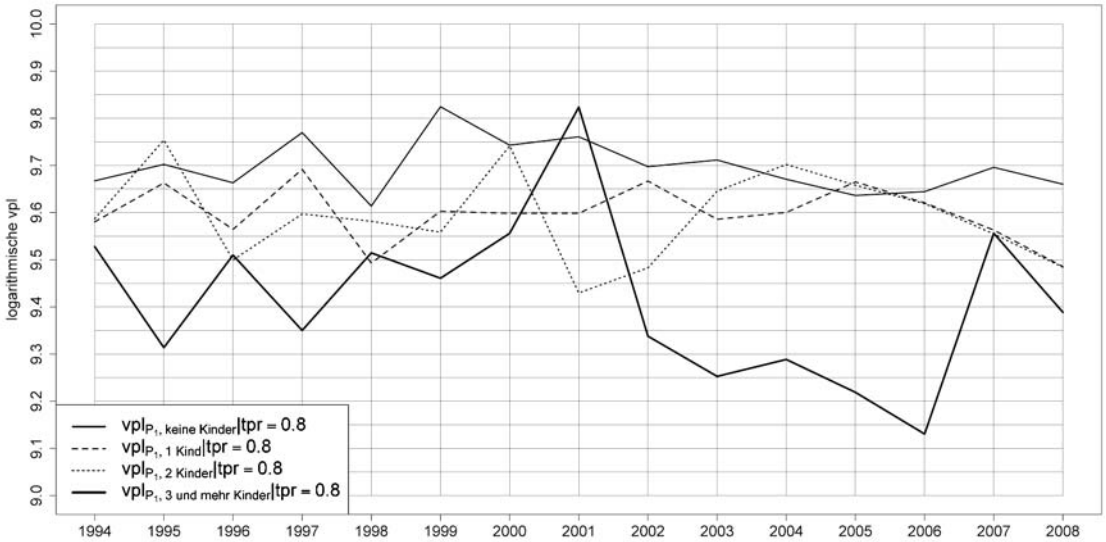


Abbildung D.12: *Historische VPLs für Schätzungen der Vulnerabilität nach Merkmal Anzahl der Kinder mit $P_{1,G}$ für den Zeitraum 1994-2008 (Methode 2) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

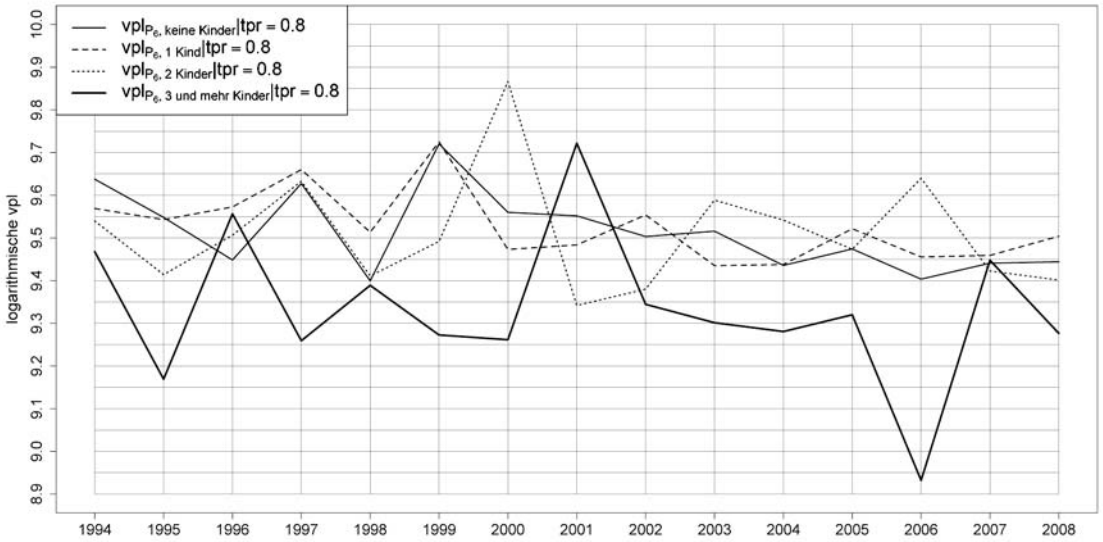
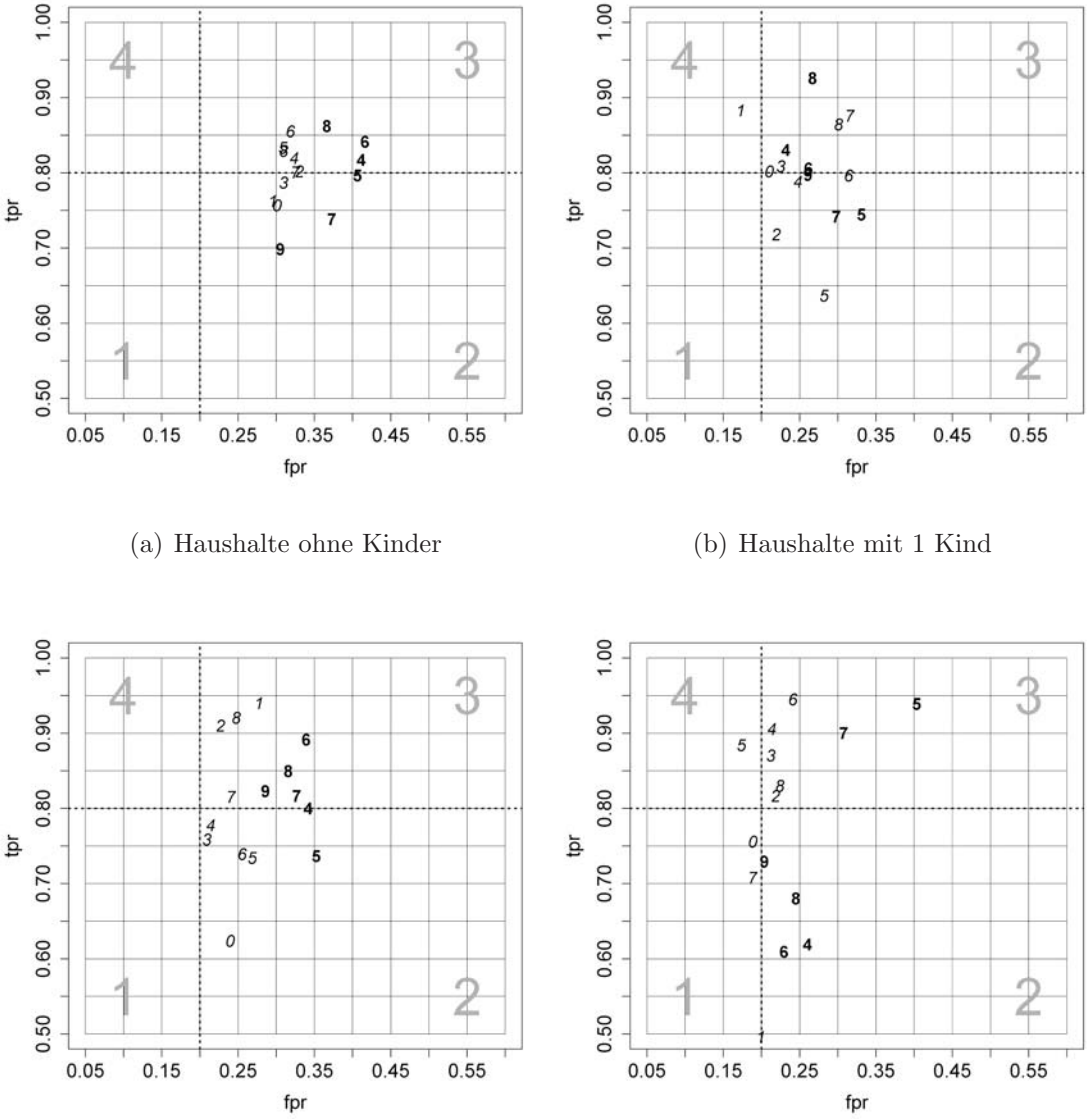


Abbildung D.13: *Historische VPLs für Schätzungen der Vulnerabilität nach Merkmal Anzahl der Kinder mit $P_{6,G}$ für den Zeitraum 1994-2008 (Methode 2) [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten].*

A.4.6 Stabilität der Vulnerabilitäts-Schätzer in den Gruppen Anzahl der Kinder

In den Abbildungen D.14 und D.15 ist die Stabilität der Vulnerabilitäts-Schätzer der Modelle $P_{1,G}$ und $P_{6,G}$ für die Gruppen nach Anzahl der Kinder für zuvor festgelegte VPLs dargestellt. Sowohl für das Model $P_{1,G}$ als auch für das Modell $P_{6,G}$ zeigt sich, dass die Genauigkeit der Schätzer für Haushalte ohne Kinder relativ stabil und vergleichbar zu denen der Schätzer von P_1 und P_6 ist. Dagegen sind die Schätzer der Gruppen mit Kindern nicht sehr stabil. Dies gilt insbesondere für die Gruppe der kinderreichen Haushalte. Dies kann durch die hohe Schwankung der VPLs erklärt werden. Im Hauptteil ist weiterhin die Genauigkeit der Vulnerabilitäts-Schätzer der einzelnen Gruppen, hochgerechnet auf die Grundgesamtheit, abgebildet (siehe Seite 154).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Haushalte ohne Kinder

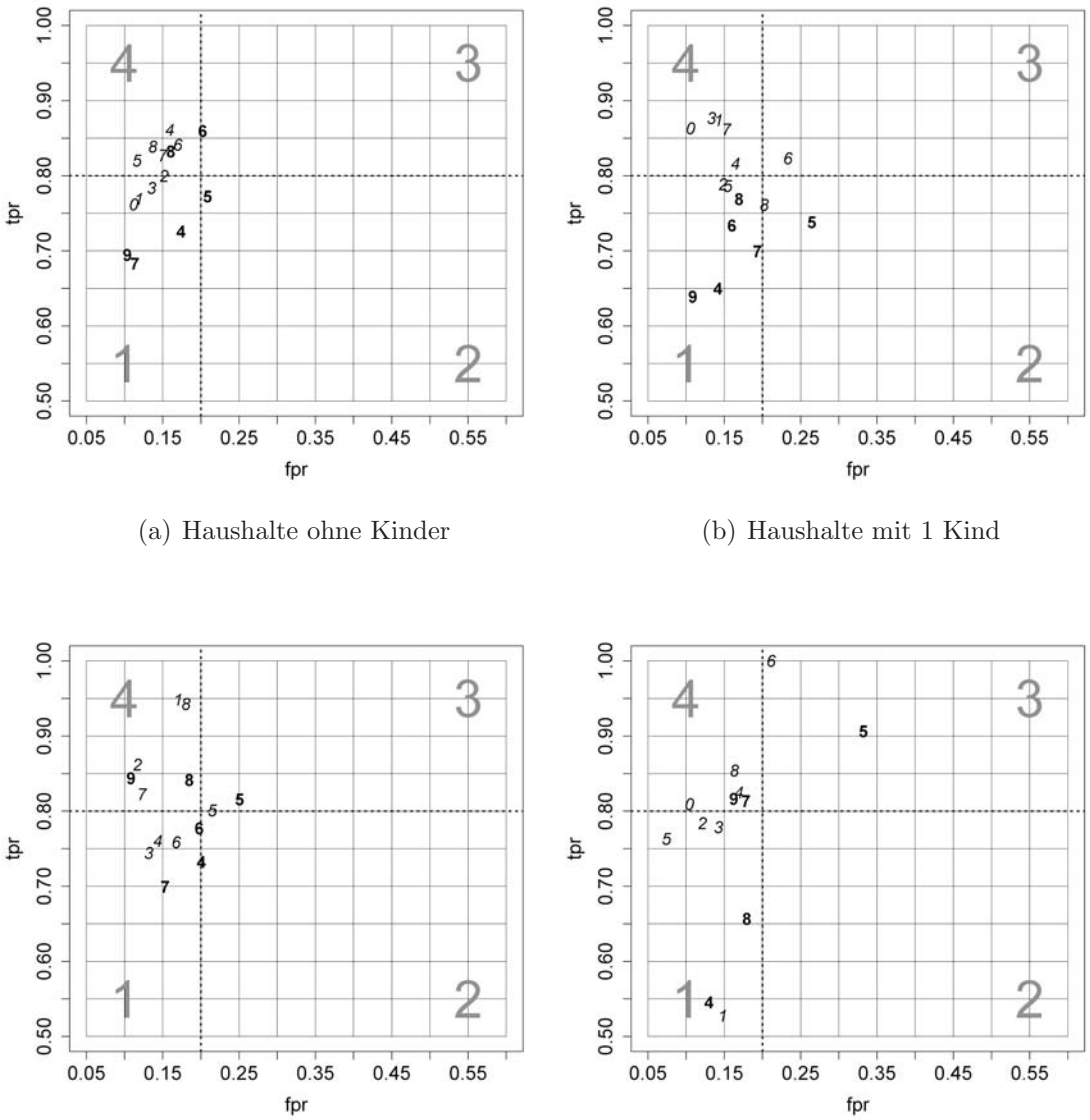
(b) Haushalte mit 1 Kind

(c) Haushalte mit 2 Kinder

(d) Haushalte mit 3 Kinder

Abbildung D.14: *Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Anzahl der Kinder im Haushalt mit Methode 2 und $P_{1,G}$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



(a) Haushalte ohne Kinder

(b) Haushalte mit 1 Kind

(c) Gruppe der Haushalte mit 2 Kinder

(d) Haushalte mit 3 Kinder

Abbildung D.15: *Stabilität der Vulnerabilitäts-Schätzer gruppiert nach Anzahl der Kinder im Haushalt mit Methode 2 und $P_{6,G}$ [eigene Darstellung auf Basis der in Kapitel 3 beschriebenen Daten]. Fett gedruckte Zahlen beziehen sich auf die 90er Jahre, während kursive Zahlen für die Jahre von 2000-2008 stehen.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Literaturverzeichnis

1. Amemiya, Takeshi (1977): The maximum likelihood estimator and the nonlinear three stage least squares estimator in the general nonlinear simultaneous equation model. Econometrica, 45, 955-968.

2. Anger, Silke (2011): Documentation PGEN. SOEP, DIW Berlin.
   Link: $http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.60055.de/pgen.pdf$ (29.03.2012)

3. Atkinson, Tony; Cantillon, Bea; Marlier, Eric und Nolan, Brian (2002): Social Indicators - The EU and Social Inclusion. Oxford University Press.

4. Bach, Hans-Uwe; Gaggermeier, Christian und Klinger, Sabine (2005): Sozialversicherungspflichtige Beschäftigung-Woher kommt die Talfahrt? IAB Kurzbericht, Nr. 26.
   Link: $http://doku.iab.de/kurzber/2005/kb2605.pdf$ (10.04.2012)

5. Biewen, Martin und Juhasz, Andos (2011): Income Inequality - Can Employment Changes Explain Rising Income Inequality in Germany? Schmollers Jahrbuch 132, 349-357.

6. Bispinck, Reinhard (2007): Löhne, Tarifverhandlungen und Tarifsystem in Deutschland 1995 - 2005. Diskussionspapier Nr. 150, Wirtschafts- und Sozialwissenschaftliches Institut (WSI) in der Hans-Böckler-Stiftung, Düsseldorf.

7. Böhmer, Michael und Heimer, Andreas (2008): Armutsrisiken von Kindern und Jugendlichen in Deutschland. Kompetenzzentrum für familienbezogene Leistungen im BMFSFJ.
   Link: $http://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/Dossier-Kinderarmut,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf$ (01.02.2012)

8. Boldin, Michael D. (1994): Data Turning Points in the Business Cycle. Journal of Business, Vol 67, Issue 1, 97-131.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

9. Box, George E.P. und Cox, David R. (1964): An Analysis of Transformations, Journal of the Royal Statistical Society. Series B, Vol. 26, No. 2, 211-252.

10. Botteril, Linda und Mazur, Nicole (2004): Risk and Risk Perception - A Literature Review. Report for the Rural Industries Research and Development Corporation, Australia.

11. brand eins und Statista (2009): Die Welt in Zahlen 2010. Tredition (1. Auflage).

12. Brenke, Karl (2009): Reallöhne über mehrere Jahre rückläufig. Wochenbericht des DIW Berlin Nr. 33/2009. 550-560.
Link: $http://www.diw.de/documents/publikationen/73/diw_02.c.289465.de/09-33-1.pdf$ (10.04.2012)

13. Brenke, Karl und Grabka, Markus (2011): Schwache Lohnentwicklung im letzten Jahrzehnt. Wochenbericht des DIW Berlin Nr. 45/2011, 3-15.
Link: $http://www.diw.de/documents/publikationen/73/diw_01.c.388567.de/11-45-1.pdf$ (10.04.2012)

14. Bundesagentur für Arbeit (2009): Der Arbeitsmarkt in Deutschland: Kurzarbeit-Aktuelle Entwicklungen.
Link: $http://statistik.arbeitsagentur.de/Statischer-Content/Arbeitsmarktberichte/Berichte-Broschueren/Arbeitsmarkt/Generische-Publikationen/Arbeitsmarkt-Deutschland-Kurzarbeit-Aktuelle-Entwicklungen.pdf$ (10.04.2012)

15. Bundesagentur für Arbeit (2011): Analytikreport der Statistik - Arbeitsmarkt in Deutschland Zeitreihen bis 2010.
$http://statistik.arbeitsagentur.de/Statischer-Content/Statistische-Analysen/Analytikreports/Zentrale-Analytikreports/Jaehrliche-Analytikreports/Generische-Publikationen/Arbeitsmarkt-Deutschland-Zeitreihen/Analyse-Arbeitsmarkt-Deutschland-Zeitreihen-2010.pdf$ (10.04.2012)

16. Bundesagentur für Arbeit (2012): Der Arbeitsmarkt in Deutschland: Zeitarbeit-Aktuelle Entwicklungen.
Link: $http://statistik.arbeitsagentur.de/Statischer-Content/Arbeitsmarktberichte/Berichte-Broschueren/Arbeitsmarkt/Generische-Publikationen/Arbeitsmarkt-Deutschland-Zeitarbeit-Aktuelle-Entwicklung-1HJ2011.pdf$ (10.04.2012)

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

17. Bundesfinanzministerium (BMF) (2007): Vergleich der Konjunkturzyklen für Deutschland: Ist der aktuelle Aufschwung anders? Monatsbericht des BMFs, Juni 2007.
Link: $http://www.bundesfinanzministerium.de/nn_17844/DE/BMF_Startseite/Publikationen/Monatsbericht_des_BMF/2007/06/070618agmb005.html$ (11.04.2012)

18. Bundesfinanzministerium (BMF) (2011): Entwicklung der Steuer- und Abgabenquoten (12.01.2011).
Link: $http://www.bundesfinanzministerium.de/nn_4316/DE/BMF_Startseite/Service/Downloads/Abt_I/Steuer_Abgabenquoten_24022012,templateId=raw,property=publicationFile.pdf$ (11.04.2012)

19. Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) (2004): Frauen in Deutschland.
Link: $http://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/Frauen-in-Deutschland,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf$ (30.01.2012)

20. Butterwegge, Christoph (2005): Krise und Zukunft des Sozialstaates. VS Verlag für Sozialwissenschaften, 2. Auflage.

21. Butterwegge, Christoph (2009): Armut in einem reichen Land - Wie das Problem verharmlost und verdrängt wird. Campus, 2. Auflage.

22. Calvo, Cesar und Dercon, Stefan (2003): Vulnerability: An axiomatic approach (the v_γ familiy of vulnerability measures). Unpublished Manuscript.

23. Calvo, Cesar und Dercon, Stefan (2005): Measuring Individual Vulnerability, Economics Series Working Papers 229, University of Oxford, Department of Economics.

24. Canberra Group (2001): Final Report and Recommendations. Ottawa.
Link: $www.lisproject.org/links/canberra/finalreport.pdf$ (25.03.2011)

25. Chambers, Robert (2006): Vulnerability: How the Poor Cope. IDS Bulletin 37 (4), 33-40. (First published 1989 in IDS Bulletin 20 (2), 1-7.)

26. Chaudhuri, Shubham (2001): Empirical Methods for assessing household vulnerability to poverty. Unpublished Manuscript.

27. Chaudhuri, Shubham; Jalan, Jyotsna und Suryahadi, Asep (2002): Assessing Household Vulnerability to Poverty from Cross-Sectional Data: A Methodo-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

logy und Estimates from Indonesia. Discussion Paper No. 010252, New York, Columbia University.

28. Chaudhuri, Shubham (2003): Empirical methods for assessing household vulnerability to poverty, Mimeo, Department of Economics, New York, Columbia University.

29. Christiaensen, Luc und Boisvert, Richard N. (2000): On Measuring Household Food Vulnerability: Case Evidence From Northern Mali. Working Paper, Department of Applied Economics and Management, Cornell University, Ithaca, New York.

30. Christiaensen, Luc und Subbarao, Kalanidhi (2005): Towards an Understanding of Household Vulnerability in Rural Kenya. Journal of African Economies 14 (4), 520-558.

31. Creedy, John (1998): The dynamics of inequality and poverty: comparing income distributions. Elgar Verlag, Cheltenham.

32. Cunningham, Wendy und Maloney, William F. (2000): Measuring Vulnerability: Who Suffered in the 1995 Mexiko Crisis. Mimeo, World Bank. Washington D.C.

33. Debels, Annelies und Vandecasteele, Leen (2008): The time lag in annual household-based income measures: assessing and correcting the bias. Review of Income and Wealth 54 (1), 71-88.

34. Dietz, Kristina (2006): Vulnerabilität und Anpassung gegenüber Klimawandel aus Sozial-ökologischer Perspektive. Diskussionspapier 01/06 des Projektes 'Global Governance und Klimawandel', Berlin.
Link: $http://www.sozial-oekologische-forschung.org/pot/download.php/Mi\%3A561\%20Vulnerabilit\%E4t\%20und\%20Anpassung\%20gegen\%FCber\%20Klimawundel\%20aus\%20sozial-\%F6kologischer\%20Perspektive/\tilde{}/intern/upload/literatur/Dietz1.pdf$ (24.03.2011).

35. DiNardo, J.; Fortin, N. und Lemieux, T. (1996): Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, Econometrica 64, 1001-44.

36. Efron, Bradley und Tishirani, Robert J. (1994): Introduction to the Bootstrap. New York: Chapman & Hall/CRC.

37. Egan, James P. (1975): Signal Detection Theory and ROC Analysis. Academic Press, New York.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

38. Egle, Christoph; Ostheim, Tobias und Zohlnhöfer, Reimut (2003): Das Rot-Grüne Projekt - Eine Bilanz der Regierung Schröder 1998-2002. Westdeutscher Verlag.

39. Egle, Christoph und Zohlnhöfer, Reimut (2007): Ende des Rot-Grünen Projekts - Eine Bilanz der Regierung Schröder 2002-2005. Westdeutscher Verlag.

40. Egle, Christoph und Zohlnhöfer Reimut (2010): Die zweite Große Koaltion - Eine Bilanz der Regierung Merkel 2005 - 2009. VS Verlag.

41. Faraway, Julien (2002): Practical Regression and Anova Using R.
Link: $http://ftp.udc.es/CRAN/doc/contrib/Faraway-PRA.pdf$
(17.04.2012)

42. Fahrmeir, Ludwig; Künstler, Rita; Pigeot, Iris und Tutz, Gerhard (2007): Statistik - Der Weg zur Datenanalyse. Springer-Verlag Berlin, Heidelberg, 6. Auflage.

43. Fawcett, Tom (2006): An introduction to ROC analysis. Pattern Recognition Letters, Vol.7 Issue 8, 861-874.

44. Fienberg, Stephen E. und Stern, Paul C.(2005): In Search of the Magic Lasso: The Truth About the Polygraph. Statistical Science Vol.20, No 3, 249-260.

45. Fishburn, Peter C. (1977): Mean-Risk Analysis with Risk Associated with Below Target Returns. The American Economic Review, Vol. 67, No. 2, 116-126.

46. Foster, Jame; Greer, Joel und Thorbecke Erik (1984): A Class of Decomposable Poverty Measures. Econometrics, Vol. 52, No. 3, 761-766.

47. Frick, Joachim R. und Grabka, Markus M. (2004): Missing Income Data in the German SOEP: Incidence, Imputation and its Impact on the Income distribution. Discussion Papers 376.
Link: $http://www.diw.de/documents/publikationen/73/diw_01.c.40900.de/dp376.pdf$ (31.05.2011)

48. Frick, Joachim R. und Grabka, Markus M. (2005): Zur Entwicklung der Einkommen privater Haushalte in Deutschland bis 2004. Wochenbericht des DIW Nr.28/2005, 429-436.

49. Frick, Joachim R.; Goebel, Jan; Grabka Markus M.; Krause Peter; Schäfer Andrea; Tucci Ingrid und Wagner, Gert.G. (2005): Zur langfristigen Entwicklung von Einkommen und Armut in Deutschland. Wochenbericht des DIW Nr. 4/2005, 59-68.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

50. Frick, Joachim R. und Grabka, Markus M. (2007): Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. Discussion Papers 3043.
Link: $ftp://repec.iza.org/RePEc/Discussionpaper/dp3043.pdf$
(31.05.2011)

51. Frick, Joachim R.; Jenkins, Stephen P.; Lillard, Dean R. ;Lipps, Oliver und Wooden, Mark (2008): Die internationale Einbettung des Sozio-oekonomischen Panels (SOEP) im Rahmen des Cross-National Equivalent File (CNEF). Vierteljahreshefte zur Wirtschaftsforschung 77 (3), 110-129.

52. Gaiha, Raghav und Imai, Katsushi (2008): Measuring Vulnerability and Poverty. UNU-Wider Research Paper No 2008/40.

53. Galler, Heinz P. (1987): Zur Längsschnittgewichtung des Sozio-oekonomischen Panels. In: Krupp, H.-J., und Hanefeld, U. (Hrsg.): Lebenslagen im Wandel: Analysen 1987, Band 2 der Reihe: Sozio-oekonomische Daten und Analysen für die Bundesrepublik Deutschland. New York, Frankfurt a. M., Campus, 295-317.

54. Goebel, Jan; Grabka, Markus M.; Krause, Peter; Kroh, Martin; Pischner, Rainer; Sieber, Ingo und Spieß Martin (2008): Mikrodaten, Gewichtung und Datenstruktur der Längsschnittstudie Sozio-oekonomischer Panel (SOEP). Vierteljahreshefte zur Wirtschaftsforschung, 77, 3, 77-109.

55. Goebel, Jan und Grabka, Markus M. (2011): Zur Entwicklung der Altersarmut in Deutschland. Wochenbericht 25/2011 des DIW Berlin, 3-16.
Link: $http://www.diw.de/documents/publikationen/73/diw_01.c.375488.de/11-25-1.pdf$ (10.04.2012)

56. Grabka, Markus M. (2005): Data Documentation 7. Codebook for the $PEQUIV File 1984-2005. DIW Berlin.
Link: $http://www.diw.de/documents/publikationen/73/43414/diw_datadoc_2005-007.pdf$ (25.11.2009)

57. Griffin, James P. (1986): Well-being: Its Meaning, Measurement and Moral Importance. Oxford: Clarendon Press.

58. Günther, Isabel und Harttgen, Kenneth (2009): Estimating Households Vulnerability to Idiosyncratic und Covariate Shocks: A Novel Method Applied in Madagascar. World Development, 37 (7), 1222-1234.

59. Haisken-DeNew, John P. und Frick, Joachim R. (2005): DTC- Desktop Companion to the German Socio-Economic Panel.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Link: $http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.38951.de/dtc.409713.pdf$ (28.03.2012)

60. Hoddinott, John und Quisumbing, Agnes (2003): Methods for Microeconometric Risk und Vulnerability Assessments. Social Protection Discussion Papers 0324. The World Bank.
Link: $http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2004/06/01/000011823_20040601153122/Rendered/PDF/29138.pdf$ (25.03.2011)

61. Horvitz, Daniel G. und Thompson, David J. (1952): A Generalization of Sampling Without Replacement Form a Finite Universe, Journal of The American Statistical Association, Vol. 47, No. 260, 663-685.

62. Institut für Arbeitsmarkt- und Berufsforschung (IAB) (2009): Kurzbericht: „Working poor“ in Deutschland und den USA-Arbeit und Armut im transatlantischen Vergleich.
Link: $http://doku.iab.de/kurzber/2009/kb0109.pdf$ (10.04.2012)

63. Jalan, Jyotsna und Ravallion, Martin (1999): Are the poor less well insured? Evidence on vulnerability to income risk in rural China. Journal of Development Economics 58, 61-81.

64. Jha, Raghbendra und Dang, Tu (2010): Vulnerability to Poverty in Papua New Guinea in 1996. Asian Economic Journal 24 (3), 235-251.

65. Just, Richard E. und Pope, Rulon D. (2003): Agricultural Risk Analysis: Adequacy of Models, Data and Issues. American Journal of Agricultural Economics, Vol. 85, No. 5, 1249-1256.

66. Kahnemann, Daniel; Slovic, Paul und Tversky, Amos (1982): Judgement Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

67. Kamanou, Giesele und Modurch, Jonathan (2002): Measuring Vulnerability to Poverty. NYU Wagner Working Paper No. WP1012. New York: New York University.

68. Kanbur, Ravi (1987): The Standard of Living: Uncertainty, Inequality and Opportunity. In The Standard of Living, edited by G. Hawthorn. New York: Cambridge University Press.

69. Krause, Peter und Ritz, Daniel (2006): EU-Indikatoren zur sozialen Inklusion in Deutschland. Vierteljahresheft zur Wirtschaftsforschung 75, 1, 152-173.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

70. Landau, Katja; Klasen, Stephan und Zucchini, Walter (2012): Measuring Vulnerability to Poverty Using Long-Term Panel Data. CRC Working Paper 118.
Link: $http://www2.vwl.wiso.uni-goettingen.de/courant-papers/CRC-PEG_DP_118.pdf$ (31.07.2012)

71. Ligon, Ethan und Schechter, Laura (2003): Measuring Vulnerability. Economic Journal 113(486), C95-C102.

72. Ligon, Ethan und Schechter, Laura (2004): Evaluating Different Approaches to Estimating Vulnerability, Social Protection Discussion Papers 30159, World Bank.

73. Little, R. J. A. und Su, H.-L. (1989): Item Non-Response in Panel Surveys. In D. Kasprzyk, G. Duncan und M. P. Singh (Eds.), Panel Surveys. New York: John Wiley, 400-431.

74. McCullagh, Peter und Nelder, John A. (1989): Generalized Linear Models, 2nd Edition. Chapman & Hall, London.

75. Organisation for Economic Co-operation and Development (2008): Growing Unequal? Income Distribution and Poverty in OECD Countries.
Link: $http://www.oecd-ilibrary.org/growing-unequal_5kzrb41dm3kd.pdf?contentType=/ns/Book\&itemId=/content/book/9789264044197-en\&containerItemId=/content/book/9789264044197-en\&accessItemIds=\&mimeType=application/pdf$ (27.01.2012)

76. Organisation for Economic Co-operation and Development (2011): Divided We Stand: Why Unequality Keeps Rising.
Link: $http://www.oecd-ilibrary.org/divided-we-stand_5kg8ghtxttd7.pdf;jsessionid=259j9ulpdfhrb.epsilon?contentType=/ns/Book\&itemId=/content/book/9789264119536-en\&containerItemId=/content/book/9789264119536-en\&accessItemIds=\&mimeType=application/pdf$ (27.01.2012)

77. Povel, Felix (2010): Perceived Vulnerability to Downside Risk. Courant Research Centre: Poverty, Equity and Growth - Discussion Papers No 43.

78. Pritchett, Lant; Suryahadi, Asep und Sumarto, Sudarno (2000). Quantifying vulnerability to poverty - a proposed measure, applied to Indonesia. Policy Research Working Paper Series 2437, The World Bank.

79. Ray, Debraj (1998): Development Economics. Princetown University Press. Princetown, New Jersey.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

80. Rendtel, Ulrich (1995): Lebenslagen im Wandel: Panelausfälle und Panelrepresentativität. Frankfurt-New York: Campus.

81. Robins, James M.; Andrea Rotnitzky und Lue Ping Zhao (1994): Estimation of Regression Coefficients when some Regressors are not always observed. Journal of the American Statistical Association, 89 (427), 846-866.

82. Robins, James M.; Andrea Rotnitzky und Lue Ping Zhao (1995): Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. Journal of the American Statistical Association, 90 (429), 106-121.

83. Rothschild, Michael und Stiglitz, Joseph Eugene (1970): Increasing Risk: A definition. Journal of Economic Theory, Vol. 2, 225-243.

84. Särndal, Carl-Rik; Swensson, Bengt und Wretman, Jan (2003): Model Assisted Survey Sampling. Springer New York Berlin Heidelberg, first softcover print.

85. Schirwitz, Beate (2007): Eine Chronik klassischer Konjunkturzyklen für Deutschland. Institut für Wirtschaftsforschung (ifo) Dresden Bericht 5/2007. Link: $http://www.cesifo-group.de/DocDL/ifo_DD_ber_0705-gesamt.pdf$ (11.04.2012)

86. Schwarze, Johannes (1995): Simulating German Income and Social Security Tax Payments Using The SOEP. Cross-National Studie in Aging. Programm Project Paper No. 19. All-University Gerontology Center Maxwell School of Citizenship and Public Affairs, Syracuse University, New York.

87. Selten, Reinhard (1991): Properties of a Measure of Predictive Success. Mathematical Social Sciences 21, 153-167.

88. Selten, Reinhard und Krischker, Wilhelm (1983): Comparison of Two Theories for Characteristic Function Experiments. In Aspiration Levels in Bargaining und Economic Decision Making, ed. R. Tietz, 259-264. Berlin: Springer.

89. Sen, Amartya (1976): Poverty: An Ordinal Approach to Measurement. Econometrica 44, 219-231.

90. Sen, Amartya (1981): Poverty und Famine: An Essay on Entitlement und Deprivation. Oxford, New York: Oxford University Press.

91. Shackle, George L.S. (1965): Comment on Two Papers on Time in Economics. Artha Vijnana.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

92. Socio-Economic Panel (SOEP) (2009): Documentation HGEN. DIW Berlin. Link: $http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.60053.de/hpfad.467493.pdf$ (18.04.2012)

93. Socio-Economic Panel (SOEP) (2010): Data for years 1984-2009, version 26, SOEP.

94. Spackman, Kent A. (1989): Signal detection theory: Valuable tools for evaluating inductive learning. Proceedings of the Sixth International Workshop on Machine Learning, 160-163.

95. Stasinopoulos,Mikis D. und Rigby, Robert A. (2007): Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R. Journal of Statistical Software, 23(7), 1-46.

96. Statistisches Bundesamt (2010): Volkswirtschaftliche Gesamtrechnungen. Fachserie 18, Reihe 1.4.

97. Statistische Ämter des Bundes und der Länder (2011): Demografischer Wandel in Deutschland, Heft 1, 2011.
Link: $http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungsHaushaltsentwicklung5871101119004,property=file.pdf$ (27.01.2012)

98. Stauder, Johannes und Hüning, Wolfgang (2004): Die Messung von Äquivalenzeinkommen und Armutsquoten auf der Basis des Mikrozensus. Statistische Analysen und Studien NRW 13, 9-30.

99. Stirzaker, David (2005): Stochastic Processes & Models. Oxford University Press, USA.

100. Stocker, Herbert (2012): Einführung in die angewandte Ökonometrie. Skript WS 2012/2013, Universität Innsbruck.
Links: $http://www.uibk.ac.at/econometrics/einf/kap08.pdf$ (26.06.2012)

101. Suryahadi, Asep und Sumatro, Sudarno (2003): Poverty and Vulnerability in Indonesia Before and After the Economic Crisis. Asian Economic Journal 2003, Vol. 17 No. 1, 45-64.

102. Swets, John A.; Dawes, Robin und Monahan, John (2000): Better Decisions through Science. Scientific American, October, 82-87.

103. Thompson, Mary Lou und Zucchini, Walter (1989): On the Statistical Analysis of ROC curves. Statistics in Medicine 8, 1277-1290.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

104. United Nations (1995): Report of the world summit for social development, Kopenhagen.
Link: $http://www.un.org/Docs/journal/asp/ws.asp?m = A/CONF.166/9$ (05.04.2012)

105. United Nations General Assembly (2000): Millenniums-Erklärung der Vereinten Nationen.
Link: $http://www.unric.org/html/german/millennium/millennium erklaerung.pdf$ (05.04.2012)

106. Wagner, Gert G.; Goebel, Jan; Krause, Peter; Pischner, Rainer und Siber, Ingo (2008): Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland - Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender). AStA Wirtschafts- und Sozialstatistisches Archiv 2, No. 4, S.301-328.

107. Weikard, Hans-Peter (2004): Poverty Measurement under Income Risk. Jahrbücher fur Nationalökonomie und Statistik, Vol. 224, No. 3, 337-350.

108. Wewer, Göttrik (1998): Bilanz der Ära Kohl. Leske + Budrich, Opladen.

109. Wooldridge, Jeffrey M. (2002a): Econometric Analysis of Cross Section and Panel Data. Cambridge, MA, The MIT Press.

110. Wooldridge, Jeffrey M. (2002b): Inverse Probability Weighted M-Estimators for Sample Selection, Attrition and Stratification. Portugese Economic Journal, 1 (2), 117-139.

111. Wooldridge, Jeffrey M. (2004): Inverse Probability Weighted Estimation for General Missing Data Problems. CeMMAP Working Papers. CWP05/04.

112. World Bank (2001): World-Development Report 2000/01- Attacking Poverty, Oxford University Press.
Link: $http://wdronline.worldbank.org//worldbank/bookpdfdownload/24$ (25.03.2011)

113. Yamin, Farhana; Rahman, Atiq und Huq, Saleemul (2005): Vulnerability, adaptation und climate disasters: a conceptual overview. Institute of Development Studies (IDS) 36 (4), 1-14.

114. Zhang, Yuan und Wan, Guanghua (2009): How Precisely Can We Estimate Vulnerability to Poverty? Oxford Development Studies, Taylor und Francis Journals 37(3), 277-287.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.