

Hans-Friedrich Eckey
Reinhold Kosfeld / Christian Dreger

Ökonometrie

Grundlagen – Methoden – Beispiele

4., durchgesehene Auflage



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.

4., durchgesehene Auflage 2011

Alle Rechte vorbehalten

© Gabler Verlag | Springer Fachmedien Wiesbaden GmbH 2011

Lektorat: Irene Buttkus

Gabler Verlag ist eine Marke von Springer Fachmedien.

Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.

www.gabler.de



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in Germany

ISBN 978-3-8349-3352-2

Vorwort zur 4. Auflage

Empirische Methoden werden in den Wirtschaftswissenschaften zunehmend wichtiger. Dazu zählt insbesondere die Ökonometrie, die in den letzten Jahren zahlreiche Neuerungen erlebt hat. Daher ist es notwendig, das Rüstzeug in verständlicher Form und anhand von Beispielen erläutert präsentiert zu bekommen. Die Grundlagen des Faches werden hier in einer neuen Auflage vorgelegt. Im Vergleich zur dritten Auflage sind zahlreiche Druckfehler ausgemerzt. Wir hoffen, dass das Buch, das sich inzwischen in einer Vielzahl von Vorlesungen bewährt hat, von Dozenten, Studierenden und Praktikern mit Gewinn verwendet werden kann.

Hans-Friedrich Eckey
Reinhold Kosfeld
Christian Dreger

Vorwort zur 3. Auflage

In der 3. Auflage haben wir Druckfehler korrigiert und einige Passagen besser aufeinander abgestimmt. Gründlich überarbeitet worden sind vor allem die Abschnitte 2.3 „Multikollinearität“ und 2.4 „Autokorrelation und Heteroskedastizität“.

Zusätzlich haben wir wiederum neuere ökonometrische Methoden aufgenommen, die z.T. in ökonometrischer Software, z.B. EViews, verfügbar sind bzw. für die Programmcodes (z.B. Rats, Gauss) existieren. So sind in Abschnitt 2.9 „Nichtstationäre Variablen und Kointegration“ die Unterabschnitte „Integration und Kointegration bei Strukturbruch“ und „Nichtlineare Einheitswurzeltests und Kointegration“ eingebaut worden. Der Abschnitt 2.12 „Panelökonometrische Modelle“ ist um den Unterabschnitt „Einheitswurzeltests für Paneldaten“ erweitert worden.

Wir hoffen, dass die Neuauflage bei Dozenten und Studierenden erneut eine positive Aufnahme finden kann.

Hans- Friedrich Eckey
Reinhold Kosfeld
Christian Dreger

Vorwort zur 2. Auflage

Zu der 2. Auflage der „Ökonometrie“ haben wir neben Korrekturen von Druckfehlern und einer Überarbeitung methodische Neuerungen aufgegriffen, die inzwischen eine Anwendungsreife erhalten haben. Die Grundkonzeption hat sich hierdurch nicht verändert. Ausgehend von den klassischen ökonometrischen Verfahren werden der Student der Wirtschaftswissenschaften und der empirische Forscher mit neuen methodischen Ansätzen vertraut gemacht, die mehr und mehr zum Standardrepertoire der angewandten Ökonometrie zählen. Die eingefügten Aufgaben zielen darauf ab, eine aktive Auseinandersetzung mit der behandelten Thematik zu fördern.

Neuere ökonometrische Verfahren sind in das zweite Kapitel –Ökonometrische Eingleichungsmodelle– integriert worden. Neben dem Breusch-Pagan-Test wird der White-Test in Abschnitt 2.4 als Alternative zum Goldfeld-Quandt-Test bei der Überprüfung existierender Heteroskedastizität vorgestellt. Als Test auf Autokorrelation wird neben dem Durbin-Watson- und dem Ljung-Box-Test der Breusch-Goldfrey-Test erörtert. Neu aufgenommen worden sind außerdem Tests auf Parameterinstabilität (Abschnitt 2.8), worunter wir neben den CUSUM- und CUSUMQ-Tests den RESET-Test und den Harvey-Collier-Test subsumiert haben. Im selben Abschnitt wird zusätzlich der Jarque-Bera-Test skizziert.

Bei der Thematik der nichtstationären Variablen und Kointegration (Abschnitt 2.9) ist angesichts der Anwendungserfahrungen eine Überarbeitung erforderlich gewesen. So wird z.B. der IDW-Test (Integration-Durbin-Watson-Test) aufgrund seiner Unzulänglichkeiten nicht mehr von uns als Test auf Integration empfohlen. An seiner Stelle stellen wir den KPSS-Test als Alternative zum ADF-Test vor. Bei der Kointegrationstechnik steht aus didaktischen Gründen nach wie vor das Engle-Granger-Verfahren im Vordergrund.

Neu aufgenommen worden ist außerdem die Thematik der bedingten Heteroskedastizität (Abschnitt 2.10). Die hier vorgestellten GARCH-Modelle gehören inzwischen zum Standardrepertoire bei der Modellierung von Finanzmarktdaten. Die gleichzeitige Berücksichtigung der Querschnitts- und Zeitdimension in Form panelökonometrischer Modelle (Abschnitt 2.12) schließt das Kapitel 2 ab.

Aus einer Reihe von Zuschriften und Gesprächen haben wir ebenso wie aus Lehrerfahrungen vielfältige Anregungen für die Neuauflage der „Ökonometrie“ erhalten. Unser Dank gilt hierfür den Fachleuten und Studenten gleichermaßen. Der druckfertige Text

ist mit großer Sorgfalt von Frau Iris Röttger erstellt worden, wofür wir ihr herzlich danken. Schließlich gilt unser Dank Frau Eckert, Gabler-Verlag, für die gute und vertrauensvolle Zusammenarbeit.

Hans- Friedrich Eckey
Reinhold Kosfeld
Christian Dreger

Vorwort zur 1. Auflage

Die Ökonometrie hat ihren Stellenwert in der empirischen Wirtschaftsforschung nicht nur behaupten, sondern ausbauen können. Ökonometrische Modelle werden nicht nur von Wirtschaftsforschungsinstituten verwendet, sondern ebenfalls von staatlichen Behörden und Institutionen bei gesamtwirtschaftlichen und strukturpolitischen Problemstellungen eingesetzt. Was den Unternehmensbereich anbelangt, so wird vor allem im Finanzsektor von ökonometrischen Methoden Gebrauch gemacht. Die Entwicklung neuer und verbesserter ökonometrischer Methoden bietet nicht nur eine Erweiterung des Anwendungsbereichs der Ökonometrie, sondern auch eine bessere Fundierung empirisch-ökonomischer Analysen. Erfahrungsgemäß ist ein Praxistransfer wissenschaftlicher Methoden stark von ihrer Einbeziehung in die akademische Lehre abhängig. Um erfolgreich von der empirischen Wirtschaftsforschung genutzt werden zu können, ist daher eine ausgewogene Integration neuerer Entwicklungen in das bekannte ökonometrische Methodenspektrum erforderlich.

Während sich diese Vorstellung bereits in einigen angloamerikanischen Lehrbüchern zur Ökonometrie widerspiegelt, sind die deutschsprachigen Lehrbücher praktisch noch ausschließlich den traditionellen ökonometrischen Verfahren verpflichtet. Sicherlich ist eine sorgfältige Prüfung neuerer Methoden erforderlich, um nicht Gefahr zu laufen, allein mit der Mode zu gehen, ohne dass eine substantielle Verbesserung erfolgt. Doch denken wir, dass es nun an der Zeit ist, zumindest einige Neuerungen in ein Ökonometrie-Lehrbuch aufzunehmen, wenn es sich nicht dem Vorwurf aussetzen will, nicht up to date zu sein. Wir haben hierbei vor allem die Gebiete der robusten Verfahren, der Kointegration und der Vektorautoregression ins Auge gefasst. Aus diesen drei Bereichen liegen inzwischen bereits vielversprechende ökonomische Anwendungen vor.

Robuste Verfahren tragen der Tatsache Rechnung, dass ökonomische Daten nicht als ideal im Sinne der Modellannahmen betrachtet werden können, sondern mit den verschiedenartigsten Kontaminationen durchsetzt sind. Aus der Wirtschaftsstatistik ist bekannt, welche Arten von Störungen im Rahmen einer Erhebung auftreten können. Robuste Verfahren zielen hauptsächlich darauf ab, verzerrende Einflüsse deviante Beobachtungen unter Kontrolle zu halten. Gerade bei wirtschaftstheoretischen Modellen wird es im Allgemeinen nicht zulässig sein, alle Beobachtungen ohne weiteres als Gleichgewichtsdaten zu betrachten. Eine ökonometrische Schätzung auf der Grundlage robuster Verfahren läßt sich anschaulich dahingehend interpretieren, dass die Gleichgewichte der ökonomischen Theorie in Form von Gleichgewichtsbereichen operationalisiert werden.

Gleichgewichtsbeziehungen werden auch im Rahmen von Kointegrationsmodellen betrachtet. Jedoch stehen hier keine kurzfristigen Gleichgewichte im Zentrum der ökonometrischen Analyse, sondern es werden vielmehr langfristige Gleichgewichtsrelationen zwischen ökonomischen Variablen im Sinne eines "steady states" untersucht. Die Vektorautoregression kommt als empirisches Analyseinstrument in Betracht, wenn Probleme bezüglich der Endogenität und Exogenität ökonomischer Variablen im Mittelpunkt stehen, wie sie z.B. bei ökonometrischen Untersuchungen über die Beziehung zwischen der Geldmenge und dem Sozialprodukt auftreten. Hier stellt die zeitreihenanalytisch orientierte Methode der Vektorautoregression ein Instrumentarium bereit, das ohne A-priori-Spezifikation der Kausalbeziehungen Einsichten in den Ablauf ökonomischer Prozesse vermitteln können.

In dem vorliegenden Lehrbuch werden diejenigen Methoden der Ökonometrie bevorzugt behandelt, die für die empirische Wirtschaftsforschung von großer Relevanz sind.

Eine fundierte Anwendung ökonometrischer Methoden setzt eine Diskussion ihrer Konzeption voraus, da sich nur dann die Grenzen ökonometrischer Erkenntnisgewinnung zuverlässig einschätzen lassen. Hierbei haben wir darauf geachtet, dass sich das verwendete mathematische Instrumentarium aus dem Repertoire konstituiert, das in der mathematischen Propädeutik für Ökonomen vermittelt wird. Außerdem werden Kenntnisse der Wahrscheinlichkeitsrechnung und induktiven Statistik aus der statistischen Grundausbildung für Ökonomen vorausgesetzt, die jedoch bei Bedarf unter Verwendung des von den Autoren im selben Verlag veröffentlichten Statistik-Lehrbuches repetiert werden können. Der Schwerpunkt liegt eindeutig in der Vermittlung ökonometrischen Basiswissens, das für eine empirisch orientierte wirtschaftswissenschaftliche

Tätigkeit unerlässlich ist. Hierzu dienen nicht zuletzt auch die zahlreichen Beispiele, mit denen immer wieder der Bezug zur ökonomischen Theorie gesucht wird. Mit Hilfe der Beispiele sollen außerdem die ökonometrischen Techniken anschaulich illustriert werden, was wir bei einer für Ökonomen adäquaten Präsentation didaktisch für unbedingt erforderlich halten.

Das Ökonometrie-Lehrbuch setzt sich aus drei Teilen zusammen. In Teil I wird die Rolle der Ökonometrie in der empirischen Wirtschaftsforschung diskutiert. Hier wird aufgezeigt, welche Tatbestände zu beachten sind, damit ökonomische Hypothesen und Theorie mit der wirtschaftlichen Wirklichkeit konfrontiert werden können. In diesem Kontext lassen sich generell die Aufgaben der Ökonometrie in der empirischen Wirtschaftsforschung erörtern. Außerdem werden hier einige Grundkonzepte betrachtet, die sich auf die ökonometrische Modellbildung beziehen.

In Teil II stehen ökonometrische Eingleichungsmodelle im Mittelpunkt der Betrachtung. Ausgehend vom klassischen multiplen Regressionsmodell werden in der ökonometrischen Praxis auftretende Verletzungen der Modellannahmen erörtert und alternative Schätzverfahren zur gewöhnlichen Methode der kleinsten Quadrate aufgezeigt. Hierzu gehören vor allem die verallgemeinerte Methode der kleinsten Quadrate und die robusten Schätzverfahren. Außerdem wird das Multikollinearitätsproblem diskutiert, das sich allerdings nicht auf die Modellspezifikation bezieht, sondern ein Datenproblem ist. Was die ökonometrische Prognose anbelangt, so beschränken wir uns hier auf eine Skizzierung des Prognoseproblems und einer Darstellung der wichtigsten Gütekriterien. Darüber hinaus werden hier spezielle Themenkomplexe wie z.B. ökonometrische Modelle mit qualitativen Variablen, Distributed-Lag-Modelle sowie die Kointegration behandelt.

Bei der Behandlung ökonometrischer Mehrgleichungsmodelle in Teil III steht zunächst einmal das Identifikationsproblem im Mittelpunkt der Diskussion. Anschließend wird gezeigt, dass die gewöhnliche Methode der kleinsten Quadrate vom Standpunkt der Inferenzstatistik nicht mehr begründbar ist. Aus diesem Grund werden alternative Schätzverfahren, wie z.B. die zweistufige Methode der kleinsten Quadrate, die dreistufige Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode bei voller Information vorgestellt. Die Vektorautoregression nimmt hier eine Sonderstellung ein, da bei dieser Methode eine alternative Modellierungsstrategie verfolgt wird. Bei dem abschließenden Vergleich ökonometrischer Schätzverfahren liegt der Schwerpunkt auf Simulationsstudien ökonometrischer Modelle, da der analytische Ansatz nur in Spezialfällen anwendbar ist.

Form und Inhalt des vorliegenden Lehrbuchs sind mit durch die Lehrveranstaltungen zur Ökonometrie geprägt, die die Autoren im Fachbereich Wirtschaftswissenschaften der Universität Gh Kassel durchgeführt haben. Hierdurch ist vor allem die didaktische Konzeption stark beeinflusst. Unser Dank hat somit zunächst einmal den Studenten zu gelten, die durch ihre Diskussionsbeiträge die Autoren veranlasst haben, einige Stellen didaktisch noch einmal zu überdenken. Frau Dipl.-Volkswirtin Petra Feldotto hat das vollständige Manuskript kritisch durchgearbeitet und einige wertvolle Verbesserungsvorschläge gemacht, die wir in der vorliegenden Fassung berücksichtigen konnten.

Hierfür sei ihr herzlich gedankt. Danken möchten wir auch Frau cand.rer.pol. Andrea Eisenberg, die die Graphiken mit großer Sorgfalt angefertigt hat. Frau Iris Röttger hat das Manuskript gewohnt zuverlässig in eine präsentierbare Form gebracht, wofür ihr ebenfalls unser Dank gilt.

Last but not least möchten wir uns bei der Lektorin des Gabler-Verlages, Frau Jutta Hauser-Fahr, für die umsichtige verlegerische Betreuung und angenehme Zusammenarbeit bedanken.

Hans-Friedrich Eckey
Reinhold Kosfeld
Christian Dreger

Inhaltsverzeichnis

Vorwort	V
Abbildungsverzeichnis	XVII
Tabellenverzeichnis	XIX
Abkürzungsverzeichnis	XXI
Symbolverzeichnis	XXIII
1. Ökonometrie und empirische Wirtschaftsforschung	1
1.1 Gegenstand und Arbeitsgebiete der Ökonometrie	1
1.2 Ökonomische Gesetze und Ätialprinzip	4
1.3 Beobachtungsmaterial und statistische Fehler	7
1.4 Variablen- und Modelltypen	11
Aufgaben	17
2. Ökonometrische Eingleichungsmodelle	19
2.1 Das multiple Regressionsmodell	19
2.1.1 Modellspezifikation	19
2.1.2 Methode der kleinsten Quadrate (OLS-Methode)	24
2.1.3 Schätzeigenschaften der OLS-Methode	40
2.1.3.1 Gütekriterien	40
2.1.3.2 Linearität	42
2.1.3.3 Erwartungstreue	42
2.1.3.4 Kovarianzmatrix des OLS-Schätzers $\hat{\beta}$	44
2.1.3.5 Effizienz	45
2.1.3.6 Konsistenz	47
2.1.4 Bestimmtheitsmaß und multipler Korrelationskoeffizient	49
Aufgaben	55
2.2 Maximum-Likelihood-Methode und Inferenzstatistik	57
2.2.1 Die Maximum-Likelihood-Methode	57
2.2.2 Erwartungstreue Schätzung der Störvarianz	61

2.2.3	Signifikanztest für die Regressionskoeffizienten	66
2.2.4	Konfidenzintervalle für die Regressionskoeffizienten	74
2.2.5	Varianzanalyse und Signifikanz des Gesamt- zusammenhangs	78
	Aufgaben	82
2.3	Multikollinearität	83
2.3.1	Begriff der Multikollinearität	83
2.3.2	Auswirkungen der Multikollinearität	86
2.3.3	Aufdeckung von Multikollinearität	89
2.3.4	Überwindung von Multikollinearität	94
	Aufgaben	97
2.4	Heteroskedastizität und Autokorrelation	98
2.4.1	Form und Auswirkungen der Modelldefekte	98
2.4.2	Tests auf Heteroskedastizität	102
2.4.2.1	Goldfeld-Quandt-Test	102
2.4.2.2	Breusch-Pagan-Test	107
2.4.2.3	White-Test	111
2.4.3	Tests auf Autokorrelation	113
2.4.3.1	Durbin-Watson-Test	113
2.4.3.2	Breusch-Godfrey-Test und Ljung-Box-Test	119
2.4.4	Verallgemeinerte Methode der kleinsten Quadrate (Generalized Least Squares)	122
2.4.4.1	GLS-Schätzung bei bekannter Kovarianz- matrix der Störterme	122
2.4.4.2	GLS-Schätzung bei unbekannter Kovarianz- matrix der Störterme	125
2.4.4.2.1	Kovarianzmatrix bei Hetero- skedastizität	126
2.4.4.2.2	Kovarianzmatrix bei Autokorrelation	131
2.4.4.2.3	Alternative Strategien bei auto- korrelierten Störtermen	137
	Aufgaben	141

2.5	Ökonometrische Modelle mit verteilten Verzögerungen	142
2.5.1	Begriff der verteilten Verzögerungen	142
2.5.2	Das allgemeine Modell verteilter Verzögerungen	143
2.5.3	Geometrische Lag-Modelle	148
2.5.3.1	Das Koyck-Modell	148
2.5.3.2	Anpassungs- und Erwartungshypothesen	153
2.5.3.3	Die OLS-Methode und ihre Schätzeigenschaften	156
2.5.3.4	Der Durbin-h-Test	162
2.5.3.5	Ein Beispiel	164
2.5.3.6	Die Methode der Instrumentvariablen (IV-Methode)	167
2.5.4	Das Almon-Verfahren	170
	Aufgaben	175
2.6	Modelle mit qualitativen Variablen	176
2.6.1	Vorbemerkungen	176
2.6.2	Qualitative Regressoren	177
2.6.3	Strukturbruchtest	183
2.6.4	Qualitative abhängige Variablen	187
2.6.4.1	Qualitative Wahlhandlungsprobleme	187
2.6.4.2	Lineares Wahrscheinlichkeitsmodell und Logit-Modell	188
2.6.4.3	Maximum-Likelihood-Schätzung des Logit-Modells	193
2.6.4.4	Likelihood-Verhältnis-Test und Pseudo- R^2	195
	Aufgaben	202
2.7	Ökonometrische Prognose	203
2.7.1	Punktprognose	203
2.7.2	Intervallprognose	204
2.7.3	Güte der Prognose	210
	Aufgaben	215

2.8	Tests auf Parameterinstabilität	216
2.8.1	Vorbemerkungen	216
2.8.2	CUSUM- und CUSUMSQ-Tests	217
2.8.3	RESET- und Harvey-Collier-Test	223
2.8.4	Jarque-Bera-Test	225
	Aufgaben	226
2.9	Nichtstationäre Variablen und Kointegration	227
2.9.1	Zeitreihenanalytische Grundlagen	227
2.9.2	Formen der Nichtstationarität	229
2.9.3	Tests auf Integration	234
2.9.4	Kointegration ökonomischer Variablen	241
2.9.5	Integration und Kointegration bei Strukturbruch	252
2.9.6	Nichtlineare Einheitswurzeltests und Kointegration	259
	Aufgaben	263
2.10	Bedingte Heteroskedastizität und ARCH-Modelle	264
2.10.1	Bedingte Erwartungswerte	264
2.10.2	ARCH-Modelle	265
2.10.3	GARCH-Modelle	268
2.10.4	ARCH-M-Modelle	269
	Aufgaben	270
2.11	Robuste Regression	271
2.11.1	Begriff der Robustheit	271
2.11.2	Verallgemeinerte Maximum-Likelihood-Schätzung (M-Schätzung)	273
2.11.3	Die Reweighted-Least-Squares-Methode (RLS-Methode)	277
	Aufgaben	284
2.12	Panelökonometrische Modelle	285
2.12.1	Querschnitts- und Zeitdimension	285
2.12.2	Gepoolte Regression und Panelmodelle	286
2.12.3	Panelmodell mit festen Effekten	288
2.12.4	Panelmodell mit zufälligen Effekten	293

2.12.5	Beispiel: Beschäftigungswirkungen einer Arbeitszeitverkürzung	299
2.12.6	Einheitswurzeltests bei Paneldaten	301
	Aufgaben	304
3.	Ökonometrische Mehrgleichungsmodelle	305
3.1	Modellspezifikation	305
3.1.1	Strukturelle Form	305
3.1.2	Stochastische Modellannahmen	311
3.1.3	Reduzierte Form	313
3.1.4	Finale Form	316
	Aufgaben	320
3.2	Identifizierbarkeit ökonometrischer Modelle	321
3.2.1	Das Identifikationsproblem	321
3.2.2	Identifikationskriterien	326
	Aufgaben	334
3.3	Schätzverfahren für interdependente Modelle	335
3.3.1	Inadäquanz der OLS-Methode	336
3.3.2	Seemingly Unrelated Regressions Equations (SURE)	339
3.3.3	Zweistufige Methode der kleinsten Quadrate	345
3.3.4	Methode der Instrumentvariablen	352
3.3.5	Dreistufige Methode der kleinsten Quadrate	357
3.3.6	Die Maximum-Likelihood-Methode bei voller Information	364
3.3.7	Vektorautoregressive Modelle	371
	Aufgaben	381
3.4	Vergleich ökonometrischer Schätzverfahren	382
3.4.1	Analytischer Vergleich	382
3.4.2	Simulationsstudien	384
	Aufgaben	390

Anhang	391
Literaturverzeichnis	403
Stichwortverzeichnis	415

Abbildungsverzeichnis

Abbildung 1.4.1:	Pfeilschema für das Keynesianische Grundmodell	13
Abbildung 1.4.2:	Pfeilschema für das Marktmodell	14
Abbildung 2.1.1:	Konsumfunktion	34
Abbildung 2.1.2:	Geldnachfragefunktion	38
Abbildung 2.2.1:	Annahme- und Ablehnungsbereiche bei dem zwei- seitigen Signifikanztest	68
Abbildung 2.2.2:	Annahme- und Ablehnungsbereich des F-Tests	79
Abbildung 2.3.1:	Multikollinearität	85
Abbildung 2.5.1:	Grafische Interpretation der Reaktionskoeffizienten (I)	144
Abbildung 2.5.2:	Grafische Interpretation der Reaktionskoeffizienten (II)	144
Abbildung 2.5.3:	Formale Analogie zwischen den Zufallsvariablen ξ und der Lag-Variablen θ	146
Abbildung 2.5.4:	Geometrische Lag-Verteilung für alternative Werte des Parameters λ	149
Abbildung 2.5.5:	Geldnachfragefunktion mit Koyck-Lag	166
Abbildung 2.6.1:	Verschiebung der Nachfragefunktion	179
Abbildung 2.6.2:	Veränderung der marginalen Konsumneigung	180
Abbildung 2.6.3:	Wahrscheinlichkeitsverteilung des Logit-Modells	191
Abbildung 2.6.4:	Abhängigkeit der Kaufwahrscheinlichkeit vom Einkommen	200
Abbildung 2.7.1:	Prognoseintervall	208
Abbildung 2.9.1:	OLS-Residuen der statistischen Regression	250
Abbildung 2.9.2:	Erwerbstätige (1000 Personen)	253
Abbildung 2.9.3:	Privater Verbrauch und verfügbare Einkommen	257
Abbildung 2.9.4:	Geldmarkt- und Kapitalmarktsatz	262
Abbildung 2.10.1:	Bedingte Varianz der DAX-Innovationen (ARCH(1)- Prozess)	267
Abbildung 3.2.1:	Marktmodell mit der Shift-Variablen Einkommen	328
Abbildung 3.3.1:	Gemeinsam abhängige Variablen des makro- ökonomischen Modells	356

Tabellenverzeichnis

Tabellen im Text

Tabelle 2.9.1:	Integrationsgrad von privatem Verbrauch und verfügbarem Einkommen	258
Tabelle 2.9.2:	Kointegration zwischen privatem Verbrauch und verfügbarem Einkommen	258

Tabellen im Anhang

Tabelle A1:	Standardnormalverteilung	393
Tabelle A2:	Quantile der Chi-Quadrat-Verteilung bei v Freiheitsgraden	394
Tabelle A3:	Quantile der t-Verteilung bei v Freiheitsgraden	395
Tabelle A4:	Quantile der F-Verteilung bei v_1 und v_2 Freiheitsgraden	396
Tabelle A5:	Durbin-Watson-Statistik	397
Tabelle A6:	Dickey-Fuller-Verteilung	399
Tabelle B1:	Daten für Energienachfragemodelle	400
Tabelle B2:	Daten zur Verkehrsmittelwahl	401

Abkürzungsverzeichnis

2SLS	Two-stage least-squares
3SLS	Three-stage least-squares
AR	Autoregressiver Prozess
ARCH-M	ARCH in mean
ARCH	Autoregressive conditional heteroskedasticity
ARMA	Autoregressiver Moving-Average-Prozess
ARIMA	Integrierte ARMA-Prozess
DF	Dickey-Fuller-Test
CUSUM	Cumulative sum of residuals
FIML	Full Information Maximum Likelihood
GARCH	Generalized autoregressive conditional heteroskedasticity
CUSUMSQ	CUSUM of squares
GLS	Generalized least-squares
IV	Instrumentvariable
MA	Moving-Average-Prozess
LIML	Limited Information Maximum Likelihood
MAE	Mittlerer absoluter Fehler
MAPE	Mittlerer absoluter prozentualer Fehler
ME	Median error
ML	Maximum Likelihood
M-Schätzung	Verallgemeinerte Maximum-Likelihood-Schätzung
MSE	Mean square error
OLS	Ordinary least-squares
RESET	Regression specification test
RLS	Reweighted least-squares
RMSE	Mittlerer quadratischer Fehler
RMSPE	Wurzel des mittleren quadratischen Fehlers
SE	Standard error
SEF	Standard error for forecast
SSE	Sum of squared residuals
SURE	Seemingly unrelated regression equations
VAR	Vektorautoregression
WLS	Weighted least-squares

Symbolverzeichnis

A	Koeffizientenmatrix der gemeinsam abhängigen und prädeteterminierten Variablen
A^{adj}	Adjungierte Matrix von A
a_j	1. Koeffizient eines autoregressiven Prozesses 2. Koeffizient einer Differenzgleichung der finalen Form
A(L)	Lag-Polynom eines autoregressiven Prozesses
b_k	Koeffizient eines Moving-Average-Prozesses
B(L)	Lag-Polynom eines Moving-Average-Prozesses
BP	Breusch-Pagan-Statistik
CI(d,b)	Kointegration der Ordnung d und b
Cov	Kovarianzoperator
Cov (x,y)	Kovarianz zwischen den Zufallsvariablen x und y
Cov (y)	Varianz-Kovarianz-Matrix des Zufallsvektors y
DW	Durbin-Watson-Statistik
d_n, d_o	Kritische Werte des Durbin-Watson-Tests
D	Dummy-Variablen-Matrix
exp	Exponentialfunktion
e	1. Eulersche Zahl (e=2,718 ...) 2. Index, der eine geplante oder erwartete Größe kennzeichnet.
E	Erwartungswertoperator
E(y)	Erwartungswert der Zufallsvariablen y
E(y)	Vektor der Erwartungswerte der Komponenten von y
f	Dichtefunktion
F	1. Prüfgröße eines F-Tests 2. Verteilungsfunktion
F_{1-α;v₁;v₂}	(1-α)-Quantil einer F-Verteilung mit v ₁ und v ₂ Freiheitsgraden
GQ	Goldfeld-Quandt-Statistik
h	Prüfgröße des Durbin-h-Tests
h_t	Bedingte Varianz
H₀	Nullhypothese
H₁	Alternativhypothese
I	Einheitsmatrix
I(d)	Integrierter Prozess der Ordnung d

JB	Jarque-Bera-Statistik
ℓ	Likelihood-Ratio-Statistik
L	1. Lag-Operator 2. Likelihoodfunktion
L^*	Logarithmierte Likelihoodfunktion
\ln	Natürlicher Logarithmus
$\tilde{\ell}_p$	Lag-Quantil ($0 < p < 1$)
med	Median
P	Wahrscheinlichkeitsoperator
N	Normalverteilung
p_i	Wahrscheinlichkeit für die i-te Ausprägung einer diskreten Zufallsvariablen
p_t	Wert einer Verteilungsfunktion
Q	Asymptotische Momentenmatrix
Q_k	Ljung-Box-Teststatistik
PR^2	Pseudo- R^2
R	Korrelationskoeffizient nach Bravais und Pearson
R	Multipler Korrelationskoeffizient
R^2	Bestimmtheitsmaß, Determinationskoeffizient
\bar{R}^2	Bereinigtes Bestimmtheitsmaß
$\text{rg}(\mathbf{X})$	Rang der Matrix \mathbf{X}
s^2	Stichprobenvarianz
$\text{sgn}(x)$	Vorzeichenfunktion
T	Transformationsmatrix
t	1. Zeitindex (bzw. Querschnittsindex) 2. Prüfgröße eines t-Tests
$t_{1-\alpha;v}$	$(1-\alpha)$ -Quantil einer t-Verteilung mit v Freiheitsgraden
$\text{tr}(\mathbf{A})$	Spur der Matrix \mathbf{A}
U	Theilscher Ungleichheitskoeffizient
u_t, v_t	Störgröße, stochastischer Störterm
$\mathbf{u}, \mathbf{v}, \mathbf{w}$	Vektor der Störgröße
\hat{u}_t	OLS-Residuen
$\hat{\mathbf{u}}$	Vektor der OLS-Residuen
U	Matrix der Störvariablen
Var	Varianzoperator

$\text{Var}(y)$	Varianz der Zufallsvariablen y
VR	Variance ratio
w_i	Lag-Gewichte der exogenen Variablen x_{t-i}
w_t	Gewichtsfunktion
$W(L)$	Lag-Polynom
WQ	Wendepunkt-Fehlerquote
x_{jt}	Exogene (unabhängige Variable) eines Eingleichungsmodells
x_{0t}	Scheinvariable
x_{kt}	k -te präterminierte Variable eines Mehrgleichungsmodells
x_t	1. Vektor der exogenen Variablen eines Eingleichungsmodells 2. Vektor der präterminierten Variablen eines Mehrgleichungsmodells
X	1. (Beobachtungs-) Matrix der exogenen Variablen eines Eingleichungsmodells 2. Matrix der präterminierten Variablen eines Mehrgleichungsmodells
x_{1t}	Vektor der präterminierten Variablen der ersten Gleichung eines Mehrgleichungsmodells zur Zeit t
X_1	Matrix der präterminierten Variablen der ersten Gleichung
y_t	Endogene (abhängige) Variable eines Eingleichungsmodells
y	Vektor der endogenen Variablen eines Eingleichungsmodells
\hat{y}_t	Regressionswerte
\hat{y}	Vektor der Regressionswerte
y_{ht}	h -te gemeinsam abhängige Variable eines Mehrgleichungsmodells
y_t	Vektor der gemeinsam abhängigen Variablen eines Mehrgleichungsmodells
Y	Matrix der gemeinsam abhängigen Variablen eines Mehrgleichungsmodells
$y_{h,t-i}$	Zeitlich verzögerte endogene Variable ($i > 0$)
y_{t-i}	Vektor zeitlich verzögerter endogener Variablen ($i > 0$)
Y_{1t}	Vektor der endogenen Variablen zur Zeit t (ohne y_1)
Y_1	Matrix der gemeinsam abhängigen Variablen (ohne y_1)
Z	Matrix von Instrumentvariablen
Z_1	Vektor der gemeinsam abhängigen und präterminierten Variablen (ohne y_1)
z_t	1. Vektor der exogenen Variablen der finalen Form 2. Vektor der gemeinsamen endogenen und präterminierten Variablen

Griechische Buchstaben

- α 1. Signifikanzniveau; Wahrscheinlichkeit für den Fehler 1. Art (α -Fehler)
2. Absolutes Glied des dynamischen Regressionsmodells beim Almon-Verfahren
- β Vektor der Regressionskoeffizienten eines Eingleichungsmodells
- $\hat{\beta}$ Schätzer (Schätzvektor) des Parametervektors β
- $\tilde{\beta}$ Linearer Schätzer (Schätzvektor) des Parametervektors β
- β_j Regressionskoeffizienten eines Eingleichungsmodells
- $\hat{\beta}_j$ Schätzer des Regressionskoeffizienten β_j
- $\tilde{\beta}_j$ Linearer Schätzer des Regressionskoeffizienten β_j
- β_{gk} Regressionskoeffizient der k-ten prädeterminierten Variablen in der g-ten Gleichung
- B** Koeffizientenmatrix der prädeterminierten Variablen eines Mehrgleichungsmodells
- χ^2 Chi-Quadrat-Verteilung
- Δ Differenzenoperator
- δ 1. Parameter heteroskedastischer Varianzen
2. Anpassungsparameter
- δ_1 Parametervektor der ersten Gleichung eines Mehrgleichungsmodells
- $\hat{\delta}_1$ Schätzer des Parametervektors δ_1
- ε Elastizität
- ϕ Autoregressiver Parameter
- $\hat{\phi}$ Schätzer des Parameters ϕ
- Φ Verteilungsfunktion einer Normalverteilung
- γ Anpassungsparameter
- γ_{gh} Koeffizient der h-ten gemeinsamen abhängigen Variablen in der g-ten Gleichung
- Γ** Koeffizientenmatrix der gemeinsam abhängigen Variablen
- γ_1 Parametervektor der endogenen Variablen (ohne y_1)
- $\hat{\gamma}_1$ Schätzer des Parametervektors γ_1
- \mathcal{G} Harvey-Collier-Statistik
- η_k KPSS-Teststatistik

- λ 1. Parameter der geometrischen Verteilung
 2. Anpassungsparameter
 3. Wurzel einer charakteristischen Gleichung
- λ_g Koeffizient einer Linearkombination
- μ Mittlerer Anpassungslag
- $\boldsymbol{\mu}$ Vektor der Erwartungswerte
- π Kreiskonstante ($\pi = 3,1415\dots$)
- Π Produktoperator
- π_{gh} Koeffizient der reduzierten Form
- $\mathbf{\Pi}$ Koeffizientenmatrix der reduzierten Form
- $\mathbf{\Pi}_1$ Koeffizientenvektor der zeitlich verzögerten endogenen Variablen der finalen Form
- $\mathbf{\Pi}_2$ Koeffizientenvektor der exogenen Variablen der finalen Form
- $\boldsymbol{\theta}_i$ Koeffizientenmatrix von \mathbf{y}_{t-1} einer Vektorautoregression
- $\boldsymbol{\theta}$ Parametervektor eines vektorautoregressiven Modells
- $\hat{\boldsymbol{\theta}}$ Schätzer des Parametervektors $\boldsymbol{\theta}$
- ρ_j Theoretischer Autokorrelationskoeffizient j-ter Ordnung
- $\hat{\rho}_j$ Empirischer Autokorrelationskoeffizient j-ter Ordnung
- $\rho(\mathbf{u})$ Modifizierte Likelihoodfunktion
- $\psi(\mathbf{u})$ Ableitung der ρ -Funktion
- σ^2 Varianz der Störvariablen
- $\hat{\sigma}^2$ Erwartungstreuer Schätzer der Störvarianz
- $\hat{\sigma}$ Erwartungstreuer Schätzer der Standardabweichung der Störvariablen
- σ_y^2 Varianz der Zufallsvariablen y
- σ_{gg} Varianz der Störgröße der g-ten Gleichung
- σ_{gh} Kontemporäre Kovarianzen zwischen den Störgrößen der g-ten und h-ten Gleichung
- $\boldsymbol{\Sigma}$ Kontemporäre Kovarianzmatrix der Störgrößen eines Mehrgleichungsmodells
- Σ Summenoperator
- $\boldsymbol{\Omega}$ 1. Kovarianzmatrix der Störgröße bei der GLS-Schätzung (bis auf multiplikative Konstante σ^2)
 2. Kontemporäre Kovarianzmatrix der Störvariablen der reduzierten Form
- $\hat{\boldsymbol{\Omega}}$ Konsistenter Schätzer für $\boldsymbol{\Omega}$

Sonstige Symbole

- \wedge 1. ohne Zusatz: Schätzer (meist OLS-Schätzer)
2. mit Zusatz: Symbol für einen anderen z.B. ML-Schätzer
- \sim 1. Symbol für einen linearen Schätzer
2. „ist verteilt“
3. Symbol für eine Abweichung vom Mittelwert
4. „ist integriert“ bzw. „kointegriert“
- $*$ 1. Symbol für transformierte Größen
2. Symbol für eine unbeobachtbare Indexvariable
- $'$ Symbol für die Transponation eines Vektors oder einer Matrix
- \otimes Kronecker-Produkt-Operator
- $|\mathbf{A}|$ Determinante einer Matrix \mathbf{A}
- $\lim_{n \rightarrow \infty}$ Limes, Grenzwert für n gegen unendlich
- $p \lim_{n \rightarrow \infty}$ Wahrscheinlichkeitslimes

1. Ökonometrie und empirische Wirtschaftsforschung

1.1 Gegenstand und Arbeitsgebiete der Ökonometrie

Ökonometrische Methoden gehören inzwischen zu einem festen Bestandteil der empirischen Wirtschaftsforschung. Die Anwendung ökonometrischer Methoden bei der empirischen Untersuchung des Nachfrageverhaltens nach Gütern und Dienstleistungen, des Geldangebotsprozesses einer Volkswirtschaft, der Investitionstätigkeit sowie einer Branchen- und Konjunkturanalyse lassen ihre Anwendungsbreite ersichtlich werden. Wenn die Ökonometrie in der empirischen Wirtschaftsforschung eine herausragende Bedeutung erlangt hat, so sind beide Disziplinen doch voneinander zu unterscheiden. Zu letzterer Disziplin gehören nämlich auch Methoden der deskriptiven Statistik wie z.B. Verhältniszahlen und Indikatoren, der multivariaten Statistik (z.B. Faktorenanalyse, Clusteranalyse) und der Input-Output-Analyse. In der empirischen Wirtschaftsforschung steht die Anwendung ökonometrischer Methoden bei ökonomischen Problemstellungen im Vordergrund. Die Ökonometrie stellt ein Methodenspektrum zur Überprüfung und Anwendung mathematisch formulierbarer ökonomischer Theorien bereit, das auf die ökonomische Modellbildung abgestellt ist. Ökonometrische Methoden werden in der empirischen Wirtschaftsforschung angewendet, um ökonomische Strukturzusammenhänge aufzudecken und ökonomische Hypothesen zu testen.

Als Geburtsstunde der **Ökonometrie** wird die Gründung der Econometric Society am 29.12.1930 in Cleveland, Ohio, USA, angesehen. Die ökonometrischen Forschungen, die zuvor vereinzelt betrieben wurden, sollten dadurch koordiniert werden. Die Ziele der Econometric Society umreißen bereits vortrefflich den Gegenstand der Ökonometrie: "The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics. ... Its main object shall be to promote studies that aim at a unification of the theoretical-quantitative approach to economic problems." (Frisch, 1933, S. 1) In den Versuchen namhafter Fachvertreter, die Ökonometrie zu definieren und abzugrenzen, kommt immer wieder die für die Disziplin konstituierende Verbindung von ökonomischer Theorie mit der Statistik und Mathematik zum Ausdruck. Die Ökonometrie soll dazu beitragen, eine Kluft zwischen der ökonomischen Theorie und der wirtschaftlichen Realität abzubauen. Ökonometrische Studien können durch die Konfrontation mit der Realität zu einer Entwicklung empirisch gehaltvoller ökonomischer Theorien beitragen. Von dem Grad der Überwindung der

Diskrepanz zwischen den Aussagen der ökonomischen Theorie und der Beobachtung der wirtschaftlichen Wirklichkeit hängt letztlich der Erfolg wirtschaftspolitischer Entscheidungen und Maßnahmen ab. Aufgrund dieser Denkweise ergibt sich eine Skepsis sowohl gegenüber einer rein "spekulativen" Wirtschaftstheorie auf der einen Seite und eines "measurement without theory" auf der anderen Seite.

Der Begriff Ökonometrie setzt sich aus den beiden griechischen Wörtern "oikonomia" (= Verwaltung, Wirtschaft) und "metron" (= Maß, Messung) zusammen. Gemessen werden sollen jedoch nicht etwa ökonomische Größen wie z.B. das Bruttosozialprodukt, der Konsum, die Investitionen oder die Exporte, da diese Art von Messung in die Sphäre der Wirtschaftsstatistik fällt. Das Interesse gilt vielmehr der Messung der Beziehungen zwischen ökonomischen Größen gegebenenfalls unter Einschluss nicht-ökonomischer Tatbestände. Die wirtschaftstheoretischen Hypothesen brauchen dabei noch nicht in mathematischer Form vorzuliegen; sie müssen jedoch mathematisch formulierbar sein. So wird z.B. in der Geldtheorie eine Geldnachfragefunktion entwickelt, in der als Argumente z.B. das Aktivitätsniveau, das Vermögen, Zinssätze und Renditen von Aktiva u.a. eingehen. Eine solche Geldnachfragefunktion lässt sich ökonometrisch schätzen, so dass der Einfluss und die Bedeutung der Determinanten auf die Geldnachfrage beurteilt werden können. Hieraus erhält man dann z.B. auch eine Einschätzung der Einkommens- und Zinselastizität der Geldnachfrage, womit wichtige Informationen für die Geldpolitik bereitgestellt werden. Von großer Bedeutung für die Geldpolitik ist außerdem die Frage nach der Stabilität der Geldnachfragefunktion, die ökonometrisch untersucht werden kann.

Wirtschaftswissenschaftliche Forschungsinstitute verwenden bevorzugt Konjunkturmodelle, die aus mehreren Gleichungen bestehen, zum Zwecke einer Konjunkturprognose und Politiksimulation. Sie enthalten neben Verhaltensgleichungen wie z.B. einer Konsum-, Investitions- und Exportfunktion auch die Identitäten der Volkswirtschaftlichen Gesamtrechnung oder Gleichgewichtsbedingungen. Ökonometrische Schätzungen können sich auch auf technische Relationen wie z.B. eine Produktionsfunktion oder institutionelle Beziehungen wie z.B. eine Steueraufkommensfunktion beziehen. Allgemein werden die Hypothesen als Eingleichungs- oder Mehrgleichungsmodelle formuliert. Es dürfte klar sein, dass Mehrgleichungsmodelle wie z.B. Markt- und Konjunkturmodelle größere Schätzprobleme aufwerfen als Eingleichungsmodelle.

Die Ökonometrie zeichnet sich dadurch aus, dass sie die ökonomische Theorie mit der wirtschaftlichen Wirklichkeit konfrontiert. Wie bedeutsam diese Konfrontation ist, zeigt

sich daran, dass es ohne Kenntnis der Größenordnung von ökonomischen Parametern nicht einmal immer möglich ist, auch nur die Richtung des Einflusses einer ökonomischen Größe auf eine andere Variable anzugeben. So hängt es z.B. im Multiplikator-Akzelerator-Modell von Samuelson von der Parameterkonstellation ab, ob das System in Schwingungen oder monoton verläuft. Nur im ersteren Fall kommt es empirisch als Baustein der Konjunkturanalyse in Betracht. Außerdem bietet erst die Einbeziehung der Fakten die Möglichkeit zu überprüfen, ob eine ökonomische Theorie als empirisch gehaltvoll eingestuft werden kann.

Aus den bisherigen Ausführungen lassen sich vier hauptsächliche Zwecke erkennen, auf die eine ökonometrische Untersuchung ausgerichtet sein kann:

- Testen einer ökonomischen Theorie (Hypothese),
- Strukturanalysen zu Planungs- und Entscheidungszwecken,
- Politiksimulation,
- Prognose.

Dabei ist zu beachten, dass sich die verschiedenen Zwecke bei Anwendungen gegenseitig bedingen und miteinander verflochten sein können.

Die Mathematik dient in der Ökonometrie als Mittel der Modellformulierung. Zum Zwecke einer kompakten Darstellung wird dabei häufig von der Matrizennotation Gebrauch gemacht. Die induktive Statistik spielt bei der Modellschätzung eine bedeutende Rolle. Hierzu ist das ökonomische Modell vorab in ein ökonometrisches Modell zu überführen, was unter Verwendung der Wahrscheinlichkeitsrechnung aufgrund der Erkenntnis erfolgt, dass sich wirtschaftliche Verhaltensweisen einer deterministischen Betrachtung entziehen. Die Schätzung und der Test von ökonomischen Parametern wie z.B. der marginalen Konsumneigung oder der Zinselastizität der Investitionen innerhalb eines Modells können im Prinzip mittels statistischer Methoden erfolgen. Allerdings hat die theoretische Ökonometrie aufgrund bestimmter Probleme bei ökonometrischen Modellen selbst spezielle Schätz- und Testmethoden entwickelt, die als ökonometrische Methoden bekannt geworden sind.

1.2 Ökonomische Gesetze und Ätialprinzip

Wenn die Ökonometrie zu einer Verringerung der Kluft zwischen ökonomischer Theorie und wirtschaftlicher Wirklichkeit beitragen soll, hat man sich Klarheit über die Art der Erkenntnisgewinnung zu verschaffen. In der ökonomischen Theorie werden die Hypothesen im Prinzip deterministisch betrachtet, was unter Rückgriff auf die *ceteris paribus*-Klausel eine Rechtfertigung findet. Auf diese Weise ist es möglich, modellmäßig die Wirkung von Einflussgrößen auf ökonomische Variablen isoliert zu betrachten. So wird im Rahmen der Keynesianischen Theorie z.B. eine makroökonomische Konsumfunktion

$$(1.2.1) \quad C_t = C_0 + c_1 \cdot Y_t$$

postuliert, in der der private Konsum C_t in der Periode t in Abhängigkeit vom Einkommen Y_t betrachtet wird. ¹ In (1.2.1) gibt C_0 den autonomen Konsum wieder, während c_1 die marginale Konsumquote bezeichnet. Aus der Konsumfunktion ergibt sich für ein gegebenes Einkommensniveau ein bestimmtes Konsumniveau. Ähnlich wie in den Naturwissenschaften gibt die Konsumfunktion eine kausale Beziehung zwischen dem Einkommen und dem Konsum an. Während in den Naturwissenschaften jedoch experimentell die isolierte Wirkung einer Einflussgröße auf eine interessierende Größe gemessen werden kann, ist dies in den Wirtschaftswissenschaften in der Regel nicht möglich. ² Das Beobachtungsmaterial entstammt der wirtschaftlichen Wirklichkeit, so dass sich hierin auch die Variationen ökonomischer Variablen niederschlagen, die in der Theorie als konstant betrachtet worden sind. Als Einflussgrößen für den Konsum sind neben dem Einkommen insbesondere auch die Einkommensverteilung und das Vermögen von Bedeutung. In dem Ausmaß, in dem sich diese Einflussgrößen verändern, variiert auch der Konsum bei gegebenem Einkommen.

Selbst wenn man die Einkommensverteilung EV und das Vermögen V als Argumente in eine Konsumfunktion aufnehmen würde,

$$(1.2.2) \quad C_t = C(Y_t, EV_t, V_t),$$

¹ Für die Diskussion ist es unerheblich, dass bei Einbeziehung des Staates Steuern und Transfers zu berücksichtigen sind, so dass als relevante Einkommensgröße hauptsächlich das verfügbare Einkommen in Betracht kommt.

² Eine experimentelle Wirtschaftsforschung ist allenfalls in Spezialgebieten der Mikroökonomik realisierbar. In der Makroökonomik kommen Experimente dagegen von vornherein nicht in Betracht.

wäre eine Kausalität in dem Sinne, dass gegebene Werte von Y_t , EV_t und V_t ein bestimmtes Konsumniveau erzeugen würden, nicht gegeben. Denn eine Vielzahl von weiteren Einflussgrößen wie z.B. die soziale Stellung, Erwartungen, finanzielle Mittel, Altersstruktur und Zinssätze wirken sich ebenfalls auf den Konsum aus, wenn sie auch isoliert genommen vielleicht nur geringe Effekte zeigen. Das bedeutet, dass der Konsum keineswegs eine deterministische Größe ist, wenn die relevanten Einflussgrößen in einer Konsumfunktion berücksichtigt worden sind. Vielmehr ist er eine **stochastische Größe** (Zufallsgröße), bei der ein bestimmter Rest immer unerklärt sein wird.

Doch selbst bei einem gegebenen Bedingungskomplex sind ökonomische Größen nicht vollständig determiniert, da die Wirtschaftssubjekte eine gewisse Freiheit bei ihren Entscheidungen haben. Wie die konkrete Entscheidung aussieht, wird nicht unwesentlich von den verfügbaren Informationen beeinflusst. In bestimmten Situationen kann ein Wirtschaftssubjekt durchaus außerökonomischen Faktoren den Vorrang geben, was vielleicht als nichtrationale Entscheidung aussehen mag. Auch hieran wird deutlich, dass das Prinzip der Kausalität bei einer Überwindung der Kluft zwischen ökonomischer Theorie und Beobachtung nicht hilfreich ist.

Ökonomische Gesetze zeichnen sich dadurch aus, dass bei gegebenen Bedingungen nicht eine eindeutige Folge resultiert, wie es das **Kausalprinzip** im strengen Sinne erfordert. Vielmehr sind unter gegebenen Randbedingungen alternative Folgen als Resultat ökonomischer Entscheidungen denkbar und möglich. So ist z.B. ein bestimmtes Einkommensniveau bei gegebenen wirtschaftlichen Rahmenbedingungen keinesfalls nur mit einem ganz bestimmten Konsumniveau kompatibel, sondern durchaus mit alternativen Konsumausgaben, die jedoch unterschiedliche Eintrittswahrscheinlichkeiten haben. An Stelle des Kausalprinzips tritt daher in der Ökonometrie das **Ätialprinzip**, das besagt, dass den allgemeinen Bedingungen mögliche Folgen ökonomischer Entscheidungen zugeordnet werden können, deren Realisation sich aus der zugrunde liegenden Wahrscheinlichkeitsverteilung ergibt³.

Angenommen, das Einkommen Y_t ist allein die entscheidende Einflussgröße für die Konsumausgaben C_t . Dann gibt die Beziehung (1.2.1) die systematische Komponente des gesamtwirtschaftlichen Konsums bei gegebenem Einkommen wieder. Alle weiteren Einflussgrößen werden durch eine Störgröße u_t erfasst, die die unsystematische Kom-

³ Das Ätialprinzip, das auf Hartwig (1956), S. 252ff., zurückgeht, ist von Menges (1959), S. 611ff., als Grundlage für die Erkenntnisaufgabe der Ökonometrie herausgearbeitet worden. Vgl. auch Menges (1961), S. 20ff.

ponente der Konsumausgaben darstellt. Die ökonometrische Formulierung der makroökonomischen Konsumfunktion (1.2.1) lautet demzufolge

$$(1.2.3) \quad C_t = C_0 + c_1 \cdot Y_t + u_t .$$

Die Störgröße u_t ist eine Zufallsvariable, deren Wahrscheinlichkeitsverteilung zu spezifizieren ist. Wenn viele Einflussfaktoren auf den gesamtwirtschaftlichen Konsum einwirken, von denen keiner dominierend ist, kann unter Umständen unter Berufung auf den Zentralen Grenzwertsatz der Statistik für u_t eine Normalverteilung angenommen werden. Ökonometrische Schätzverfahren setzen aber nicht notwendig einen bestimmten Verteilungstyp voraus, sondern beschränken sich z.B. auf die ersten beiden Momente der Wahrscheinlichkeitsverteilung. Auf jeden Fall muss gesichert sein, dass die Störgröße u_t keinen systematischen Einfluss mehr auf die zu erklärende ökonomische Variable ausübt, da ansonsten das Modell fehlspezifiziert wäre. Wenn z.B. die Einkommensverteilung EV_t einen systematischen Einfluss auf das Konsumniveau C_t besitzt, müsste sie explizit in die Konsumfunktion aufgenommen werden,

$$(1.2.4) \quad C_t = C_0 + c_1 \cdot Y_t + c_2 \cdot EV_t + u_t ,$$

sofern sie im Untersuchungszeitraum veränderlich ist. Bei Längsschnittanalysen kann auch der Fall eintreten, dass die Momente der Wahrscheinlichkeitsverteilung der Störgröße u_t im Zeitablauf nicht konstant bleiben. So ist es z.B. nicht unplausibel, dass ihre Streuung mit wachsendem Einkommen zunimmt. Um derartige Probleme adäquat zu behandeln, sind eigens in der Ökonometrie spezielle Methoden entwickelt worden. Während man lange Zeit bestimmte Annahmen als gegeben unterstellte und optimale Schätzverfahren unter idealen Voraussetzungen suchte, hat man sich in neuerer Zeit verstärkt einer Überprüfung von Modellannahmen zugewandt (Krämer, 1991). Außerdem werden bei der Schätzung ökonometrischer Modelle vermehrt auch "Verschmutzungen" des Datenmaterials berücksichtigt.⁴

⁴ Unter diesem Aspekt kommt robusten Schätzverfahren eine größere Bedeutung zu, die jedoch erst langsam Einzug in die Ökonometrie finden. Siehe z.B. Koenker (1982); Judge et al. (1985), S. 828ff.; Amemiya (1985), S. 70ff.

1.3 Beobachtungsmaterial und statistische Fehler

Das Beobachtungsmaterial für ökonomische Modelle wird überwiegend von der amtlichen Statistik bereitgestellt. Ergänzend werden vor allem Daten, die von wirtschaftswissenschaftlichen Forschungsinstituten bereitgestellt werden, herangezogen. So erhebt das Ifo-Institut für Wirtschaftsforschung, München, Daten über Urteile, Erwartungen und Antizipationen der Unternehmen im Rahmen von Konjunktur- und Investitionstests, die im Gegensatz zu den "harten" Daten der amtlichen Statistik als "weiche" Daten bezeichnet werden. In diese Kategorie fällt z.B. auch der Konsumklimaindex der Gesellschaft für Konsumforschung (GfK), Nürnberg. Für länderübergreifende ökonomische Analysen kann auf zum Teil harmonisiertes Datenmaterial von supranationalen Organisationen wie z.B. der Organisation für wirtschaftliche Entwicklung und Zusammenarbeit (OECD), Brüssel und dem Statistischen Amt der Europäischen Gemeinschaften (EUROSTAT), Luxemburg, zurückgegriffen werden.

Bei gesamtwirtschaftlichen und Branchenanalysen werden vor allem Längsschnittdaten herangezogen. In der Tat ist ein derartiges Aggregationsniveau in der Ökonometrie vorherrschend, so dass die ökonomischen Modelle primär als Zeitreihenmodelle konzipiert sind. Es handelt sich hierbei jedoch nicht um eine Zeitreihenanalyse im engeren Sinne, bei der die Zeitreihen aus sich selbst heraus erklärt werden⁵. Vielmehr geht es hier gerade darum, Beziehungen zwischen ökonomischen Variablen gegebenenfalls unter Einbeziehung sozialer und demographischer Einflussgrößen aufzudecken. Methoden der Zeitreihenanalyse im engeren Sinne können aber durchaus in ein ökonomisches Modell integriert werden.

Dennoch lassen sich ökonomische Methoden gleichermaßen im Mikrobereich anwenden. Hierzu gehören z.B. die Schätzung der Kostenfunktionen von Unternehmen und die Ermittlung von Preis-Absatz-Funktionen für Produkte. Bei mikroökonomischen Analysen werden verstärkt Querschnittsdaten, d.h. Daten von statistischen Einheiten aus einer gegebenen Periode oder zu einem bestimmten Zeitpunkt, verwendet. Für mikroökonomische Untersuchungen sind spezielle ökonomische Methoden entwickelt worden, die das Instrumentarium der Ökonometrie in vielfältiger Hinsicht erweitern (Ronning, 1991). Allerdings sind Querschnittsanalysen nicht allein auf Mikrodaten be-

⁵ Wald (1936) bezeichnet derartige zeitreihenanalytische Verfahren als "innere Methode". Die ökonomischen Verfahren können in dieser Betrachtungsweise als "äußere Methode" bezeichnet werden.

schränkt. Statistische Einheiten können z.B. auch Branchen und Regionen einer Volkswirtschaft oder bei internationalen Studien Länder sein.

Zwischen einer auf Querschnittsdaten basierenden mikroökonomischen Verhaltensgleichung und ihrem makroökonomischen Gegenstück tritt ein **Aggregationsproblem**, das bei einer Kombination von Quer- und Längsschnittsanalysen zu beachten ist. Um dieses Problem aufzuzeigen, gehen wir von einer mikroökonomischen Konsumfunktion aus, die für den i -ten Haushalt

$$(1.3.1) \quad c_i = c_{0i} + c_{1i} \cdot y_i, \quad i = 1, 2, \dots, N$$

lautet. c_i gibt dabei den Konsum und y_i das Einkommen des i -ten Haushalts an. Außerdem bezeichnet c_{0i} den autonomen Konsum und c_{1i} die marginale Konsumquote dieses Haushalts. Aggregation von (1.3.1) über alle N Haushalte einer Volkswirtschaft führt zu

$$\sum_{i=1}^N c_i = \sum_{i=1}^N c_{0i} + \sum_{i=1}^N c_{1i} \cdot y_i,$$

woraus man mit

$$C = \sum_{i=1}^N c_i \quad \text{und} \quad C_0 = \sum_{i=1}^N c_{0i}$$

die Beziehung

$$(1.3.2) \quad C = C_0 + \sum_{i=1}^N c_{1i} \cdot y_i$$

erhält. Nur in dem Spezialfall $c_{11} = c_{12} = \dots = c_{1N}$ kann (1.6) unmittelbar in der Form

$$(1.3.3) \quad C = C_0 + c_1 \cdot Y \quad \text{mit} \quad Y = \sum_{i=1}^N y_i$$

geschrieben werden, da dann kein Unterschied zwischen der marginalen Konsumneigung auf Mikro- und Makroebene besteht. Aus der Konsumtheorie ist jedoch bekannt, dass die marginale Konsumneigung mit zunehmendem Einkommen abnimmt, so dass allgemein $c_{11} \neq c_{12} \neq c_{1N}$ ist. Wenn a_i den Anteil des Einkommens des i -ten Haushalts am Volkseinkommen Y angibt,

$$y_i = a_i Y,$$

dann erhält man mit

$$\sum_{i=1}^N c_{1i} \cdot y_i = \left(\sum_{i=1}^N a_i \cdot c_{1i} \right) Y$$

aus (1.3.2)

$$(1.3.4) \quad C = C_0 + \left(\sum_{i=1}^N a_i \cdot c_{1i} \right) Y .$$

Aus einem Vergleich von (1.3.4) mit (1.3.3) geht hervor, dass die makroökonomische Konsumneigung c_1 ein gewogenes arithmetisches Mittel der mikroökonomischen Konsumneigungen c_{1i} ist, wobei die Einkommensanteile a_i die Gewichte sind. Es zeigt sich damit, dass die einfache Aggregation, die auf der Variablenebene möglich ist, bei ökonomischer Verhaltensgleichung nicht mehr anwendbar ist.

Da der Ökonometriker bei empirischen Untersuchungen fast ausnahmslos auf vorhandenes Datenmaterial zurückgreift, hat er stets einzuschätzen, inwieweit es im Hinblick auf sein Untersuchungsziel in geeigneter Form vorliegt. Sind z.B. die Investitionen in einer Form abgegrenzt, die es erlaubt, differenzierte Aussagen über Determinanten verschiedener Arten von Erweiterungsinvestitionen zu machen? In welcher Abgrenzung sollen das Vermögen oder die Geldmenge herangezogen werden, um adäquate Aussagen über die Nachfrage nach dauerhaften Konsumgütern oder die Inflationsentwicklung machen zu können? Wie muss in letzterem Fall ein Index beschaffen sein, um das Ausmaß der Inflation valide messen zu können? Man kann nicht erwarten, dass zwischen einem ökonomischen Modellbegriff wie z.B. der Inflation und einem statistischen Gattungsbegriff wie z.B. einem Preisindex eine vollkommene Übereinstimmung besteht. Aber die logische Diskrepanz zwischen dem theoretischen Konstrukt und dem statistischen Konzept darf nicht zu groß sein⁶, wenn gültige Aussagen über ökonomische Zusammenhänge gewonnen werden sollen. Der Ökonometriker hat daher bei der Interpretation Adäquationsfehler zu berücksichtigen, die sich aus der Operationalisierung ökonomischer Begriffe ergeben.

Statistische Fehler treten in verschiedener Form bei der Erhebung und Aufbereitung der Daten auf (Krug und Nourney, 1987). Problematisch ist, dass Erhebungs- und Aufbereitungsfehler im Allgemeinen nicht dazu tendieren, sich gegenseitig auszugleichen. Messfehler lassen sich selbst bei einer sorgfältigen Planung und Durchführung einer Erhebung nicht vollständig vermeiden, so dass auch qualitativ hochwertige Daten statistische

⁶ Die Erreichung einer "minimalen logischen Diskrepanz" zwischen einem theoretischen Konstrukt und einem statistischen Gattungsbegriff kennzeichnet das Adäquationsproblem der Wirtschaftsstatistik, das von Hartwig, H. (1956), S. 260ff., herausgearbeitet worden ist. Siehe hierzu auch Grohmann (1985).

Fehler aufweisen. Bei ökonometrischen Verfahren werden statistische Fehler jedoch weitgehend vernachlässigt ⁷.

Bei zu erklärenden ökonomischen Größen in Verhaltensgleichungen, technischen Relationen und institutionellen Beziehungen können allerdings Messfehler unter den explizit berücksichtigten Zufallsfehlern subsumiert werden. Die ökonometrische Formulierung der makroökonomischen Konsumfunktion (1.2.1) lautet demzufolge

$$(1.3.5) \quad C_t = C_0 + c_1 \cdot Y_t + u_t ,$$

wobei u_t als Zufallsvariable eine Störgröße bezeichnet, deren Wahrscheinlichkeitsverteilung noch zu spezifizieren ist. Die Zufallsvariable u_t trägt dem stochastischen Charakter ökonomischer Gesetze Rechnung. Unter diesem Aspekt treten Messfehler bei Konsumgrößen in den Hintergrund, denn sie können ohne weiteres durch die Störgröße u_t erfasst werden. Erklärende Variablen in einem Eingleichungsmodell wie z.B. hier das Einkommen werden jedoch in der Regel als gegeben angesehen, was bedeutet, dass auch Messfehler vernachlässigt werden. Gleichwohl würde eine stochastische Behandlung derartiger Größen keine substantiellen Unterschiede hervorbringen, sofern bestimmte statistische Voraussetzungen erfüllt sind ⁸. Erhebungs- und Messfehler bleiben insbesondere auch bei Identitäten wie z.B.

$$(1.3.6) \quad Y_t = C_t + I_t$$

mit I_t als Investition in der Periode t außer Betracht. Die Identität (1.3.6) ergibt sich definitorisch aus der Volkswirtschaftlichen Gesamtrechnung für eine geschlossene Volkswirtschaft ohne staatliche Aktivität. Wie später noch zu zeigen sein wird, ergeben sich für die marginale Konsumquote unterschiedliche Schätzwerte, je nachdem, ob man sie isoliert aus der makroökonomischen Konsumfunktion (1.3.5) oder unter Berücksichtigung der Identitätsgleichung (1.3.6) ermittelt.

⁷ Eine Ausnahme bilden die Fehler-in-den-Variablen-Modellen, die sich jedoch in der Regel auf Zufallsfehler beschränken. Siehe hierzu z.B. Schneeweiß (1992), S. 216ff. Systematischen Fehlern in Form von Ausreißern wird bei robusten Verfahren eine besondere Aufmerksamkeit gewidmet.

⁸ Hauptsächlich muss die erklärende Variable unkorreliert mit der Störgröße sein. Siehe hierzu Schönfeld (1971), S. 1ff.

1.4 Variablen- und Modelltypen

In der Ökonometrie werden ökonomische Hypothesen vor allem durch **lineare Modelle** dargestellt, was auf verschiedene Gründe zurückgeführt werden kann. Vielfach macht die Wirtschaftstheorie keine Aussage über die Form der "kausalen" Beziehungen zwischen den ökonomischen Variablen. Die Modellierung einer Hypothese in Form einer linearen Beziehung entspricht dann dem Einfachheitspostulat (Assenmacher und Braun, 1981). Grundsätzlich ist ein einfacheres Modell einem komplexeren Modell vorzuziehen, wenn eine sophisticatedere Modellierung keinen wesentlichen Vorteil an Erkenntnis und Information mit sich bringen würde. Innerhalb eines bestimmten Bereichs lässt sich praktisch jedoch fast jede Funktion durch eine lineare Beziehung approximieren. Außerdem können nichtlineare Modelle oft durch eine geeignete Transformation linearisiert werden. So kann z.B. die Produktionsfunktion vom Typ Cobb-Douglas

$$(1.4.1) \quad Y_t = \alpha \cdot L_t^\beta \cdot K_t^\gamma \cdot e^{u_t} ,$$

in der Y die gesamtwirtschaftliche Produktion, L den Arbeitseinsatz, K den Kapitaleinsatz und u_t eine Störgröße bezeichnen, durch Logarithmierung in die lineare Form

$$(1.4.2) \quad Y_t^* = \alpha^* + \beta \cdot L_t^* + \gamma \cdot K_t^* + u_t$$

mit

$$Y_t^* = \ln Y_t, \quad L_t^* = \ln L_t \quad \text{und} \quad K_t^* = \ln K_t$$

überführt werden. Die Cobb-Douglas-Produktionsfunktion (1.4.2) stellt ein lineares Regressionsmodell dar, das z.B. mit der Methode der kleinsten Quadrate geschätzt werden kann, sofern die Störgröße u_t bestimmte Voraussetzungen erfüllt.

Bei **ökonomischen Eingleichungsmodellen** sind praktisch nur zwei Typen von Variablen zu unterscheiden. Während hierbei stets eine Variable durch die Hypothese erklärt werden soll, kommen als erklärende Variablen eine oder mehrere Einflussgrößen in Betracht. Im Rahmen eines Regressionsmodells wird die zu erklärende Größe als **Regressand** bezeichnet; die erklärenden Größen heißen **Regressoren**. Bei der Cobb-Douglas-Produktionsfunktion ist die gesamtwirtschaftliche Produktion Y_t der Regressand und die Produktionsfaktoren Arbeit L_t und Kapital K_t sind die Regressoren. Mathematisch handelt es sich dabei um die abhängige Variable auf der einen Seite und die unabhängigen Variablen auf der anderen Seite. In der Ökonometrie hat sich für die

zu erklärende Variable der Begriff **endogene Variable** eingebürgert und die erklärenden Variablen werden **exogene Variablen** genannt. Die Störgröße u_t ist eine **latente Variable**, da sie nicht direkt beobachtbar ist.

Letztere Unterscheidung ist vor allem bei **ökonometrischen Mehrgleichungsmodellen** von großer Bedeutung, die vor allem zum Zwecke der Konjunktur- und Branchenanalyse sowie zur Politiksimulation herangezogen werden. Ein ökonometrisches Mehrgleichungsmodell haben wir schon kennen gelernt, indem wir die Keynesianische Konsumfunktion (1.2.3) durch die aufgrund der Kreislauftheorie zu fordernde Identität (1.3.6) erweitert haben ⁹:

$$C_t = \alpha + \beta \cdot Y_t + u_t$$

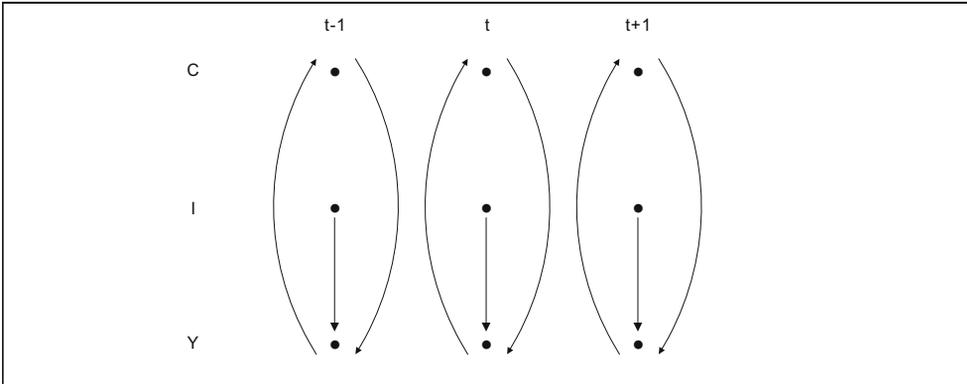
$$Y_t = C_t + I_t .$$

In diesem Keynesianischen Grundmodell sind der Konsum C_t und das Einkommen Y_t endogene Variablen, da diese Größen modellimmanent erklärt werden. Nicht erklärt werden dagegen die Investitionen I_t , die exogen vorgegeben werden müssen. Häufig werden in gesamtwirtschaftlichen Modellen demographische Größen wie z.B. die Bevölkerung und die Altersstruktur nicht durch das ökonometrische Modell erklärt, so dass sie exogene Variablen sind. Ob die Variablenwerte aus Bevölkerungsprognosen, Zeitreihenanalysen oder anderen Berechnungen stammen, ist bei der Klassifizierung unerheblich. Als exogen werden oft ebenso wirtschaftspolitische Steuerungsgrößen wie z.B. der Diskontsatz oder ein Steuersatz betrachtet. Gleiches trifft auf die Staatsausgaben zu, die ebenfalls in der Regel nicht innerhalb eines ökonometrischen Modells erklärt werden.

Wenn sich in einem ökonometrischen Modell endogene Variablen gegenseitig beeinflussen, spricht man von einem **interdependenten Modell**. Das hier betrachtete gesamtwirtschaftliche Modell ist interdependent, da der Konsum über die Konsumfunktion abhängig vom Einkommen ist und es andererseits eine Rückkoppelung in der Form gibt, dass das Konsumniveau das gesamtwirtschaftliche Einkommen determiniert. In einem Pfeilschema ¹⁰ lässt sich ein interdependentes Modell dadurch erkennen, dass in einer gegebenen Periode Pfeile, die die "kausale" Wirkung widerspiegeln, von einigen Variablen in beide Richtungen verlaufen (s. Abbildung 1.4. 1).

⁹ Dieses einfache ökonometrische Mehrgleichungsmodell dient hier natürlich nur zu Beispielszwecken. Ökonometrische Mehrgleichungsmodelle, mit denen wirtschaftspolitische Maßnahmen analysiert und Wirtschaftsabläufe prognostiziert werden, bestehen häufig aus über 100 Gleichungen.

¹⁰ Die Verwendung eines Pfeilschemas zur Darstellung der Abhängigkeitsstruktur in interdependenten ökonometrischen Modellen geht auf Tinbergen (1939, S. 75f.) zurück.

Abbildung 1.4. 1: Pfeilschema für das Keynesianische Grundmodell

Ein **rekursives Modell** liegt dagegen vor, wenn der Einfluss der Variablen nur in eine Richtung verläuft. Man kann sich vorstellen, dass die Schätzung eines rekursiven Modells nicht so verwickelt ist wie bei einem interdependenten Modell, da die endogenen Variablen sukzessive durch die Modellvariablen erklärt werden. Obwohl anfangs versucht wurde, die wirtschaftliche Wirklichkeit in Form von rekursiven Modellen abzubilden (Wold, 1954), hat sich diese Vorgehensweise als wenig Erfolg versprechend erwiesen. Ohne eine Berücksichtigung von Rückkoppelungseffekten und Wechselwirkungen zwischen ökonomischen Variablen hält ein ökonometrisches Mehrgleichungsmodell einer Konfrontation mit der Realität im Allgemeinen nicht stand. Ökonometrische Mehrgleichungsmodelle, die wirtschaftspolitische Relevanz beanspruchen, sind heutzutage praktisch durchweg interdependente Modelle. Aus diesem Grund verzichten wir auf eine ausführliche Diskussion der rekursiven Modelle; stattdessen sollen noch einige Charakteristiken interdependenter Modelle betrachtet werden.

Einem ökonometrischen Mehrgleichungsmodell kann man es nicht notwendigig "ansehen", welche Variablen endogen und exogen sind. Wenn z.B. das Marktmodell

$$(1.4.3) \quad q_t^N = \alpha_1 + \beta_1 \cdot p_t + \gamma \cdot y_t + u_{1t}$$

$$(1.4.4) \quad q_t^A = \alpha_2 + \beta_2 \cdot p_t + u_{2t}$$

$$(1.4.5) \quad q_t^N = q_t^A = q_t$$

die Mengen q_t und Preise p_t eines Gutes simultan erklären soll, sind diese beiden Größen die endogenen Variablen, obwohl der Güterpreis p_t auf der rechten Gleichungsseite

Der Güterpreis p_{t-1} ist eine zeitlich verzögerte Variable. In Abbildung 1.4. 2 ist die verzögerte Anpassung durch die gestrichelten Pfeile dargestellt, die die Übergänge von einer Periode zu der nächsten Periode aufzeigen. Die kleinen Pfeile würden bei einer Ersetzung der Angebotsfunktion (1.4.4) durch (1.4.4') entfallen. In diesem Fall liegt ein Marktmodell mit verzögerter Angebotsanpassung vor.

Verzögerungen (lags) können mithin nicht nur bei den exogenen Variablen auftreten, sondern gleichermaßen bei den endogenen Variablen. Letztere weisen in einem gewissen Sinne einen "doppelten" Charakter auf. Einerseits werden diese Größen durch das Modell erklärt. Andererseits sind sie in einer gegebenen Periode bekannt, so wie dies bei den exogenen Variablen vorausgesetzt wird. Aus diesem Grund gehören sie unter Berücksichtigung des ersteren Gesichtspunkts zwar zu den endogenen Variablen. Letzterer Aspekt lässt es jedoch auch sinnvoll erscheinen, sie mit den exogenen Variablen in Verbindung zu bringen. Bei ökonometrischen Modellen werden die verzögerten endogenen Variablen und die exogenen Variablen zusammen als **prädeterminierte (vorherbestimmte) Variablen** bezeichnet.

Anhand eines einfachen makroökonomischen Modells sollen die verschiedenen Variablentypen zusammenfassend herausgestellt und der Modelltyp identifiziert werden:

$$(1.4.6) \quad C_t = \beta_{11} + \beta_{12} \cdot Y_t + \beta_{13} \cdot C_{t-1} + u_{1t}$$

$$(1.4.7) \quad I_t = \beta_{21} + \beta_{22} \cdot Y_{t-1} + \beta_{23} \cdot Y_{t-2} + u_{2t}$$

$$(1.4.8) \quad Y_t = C_t + I_t + A_t + (EX_t - IM_t)$$

mit den makroökonomischen Variablen

Y	Bruttosozialprodukt
C	Privater Verbrauch
I	Investitionen
A	Staatsverbrauch
EX	Exporte
IM	Importe.

Für die Störvariablen ist die übliche Bezeichnung u verwendet worden. Der Konsumfunktion (1.4.6) liegt die Habit-Persistence-Hypothese (Brown, 1952) zugrunde, nach der der private Verbrauch C_t in einem Zeitraum t von dem Einkommen Y_t der aktuellen Periode t und dem Konsumniveau C_{t-1} der Vorperiode $t-1$ abhängig ist. Die Investitionsfunktion (1.4.7) gibt ein verallgemeinertes Akzelerationsprinzip wieder. Dass dies der Fall ist, zeigt sich, wenn man $\beta_{22} = -\beta_{23}$ setzt, wonach die Investitionen I_t in der Peri-

ode t durch die Veränderung der Nachfrage der Vorperiode, $Y_{t-1}-Y_{t-2}$, determiniert werden. Der Koeffizient β_{23} muss demzufolge ein negatives Vorzeichen haben. Gleichung (1.4.8) ist eine Definitionsgleichung, durch die das Bruttosozialprodukt nach seiner Verwendung definiert wird.

In dem gesamtwirtschaftlichen Modell der Gleichungen (1.4.6) bis (1.4.8) werden der Private Verbrauch, die Investitionen und das Bruttosozialprodukt erklärt. Sie werden als gemeinsam abhängige (= endogene) Variablen bezeichnet. Nicht erklärt werden dagegen der Staatsverbrauch, die Exporte und die Importe, die damit die exogenen Variablen des ökonometrischen Modells sind. Der Private Verbrauch und das Bruttosozialprodukt gehen außerdem mit einer Verzögerung von einer Periode bzw. einer und zwei Perioden als zu erklärende Variablen in das gesamtwirtschaftliche Modell ein. A_t , EX_t , IM_t , C_{t-1} , Y_{t-1} und Y_{t-2} bilden mithin die Gruppe der prädeterminierten Variablen.

In unserem makroökonomischen Modell ist Y in (1.4.6) eine erklärende Variable für C , und in (1.4.8) ist umgekehrt C eine erklärende Variable für Y . Beide Größen beeinflussen sich gegenseitig, so dass ein **interdependentes** Modell vorliegt. Außerdem ist das Modell **dynamisch**, da die endogenen Variablen C und Y in verzögerter Form vorkommen. Modelle, die sich auf eine Periode beziehen oder in denen nur exogene Variablen verzögert auftreten, heißen **statisch**. Ob ein ökonometrisches Modell dynamisch oder statisch ist, hängt allein davon ab, ob es verzögert endogene Variablen enthält. Die dynamischen Eigenschaften eines ökonometrischen Modells haben vor allem bei einer Konjunkturanalyse eine große Bedeutung.

Aufgaben

- (1.1) Erläutern Sie die konstitutionellen Elemente der Ökonometrie!
- (1.2) Grenzen Sie das Ätialprinzip gegenüber dem Kausalprinzip ab!
- (1.3) Zeigen Sie das Aggregationsproblem anhand einer Produktionsfunktion vom Typ Cobb-Douglas

$$Y = L^\beta K^\gamma$$

auf, in der Y den Output, L den Arbeitseinsatz und K den Kapitaleinsatz bezeichnen!

- (1.4) Welche Variablentypen sind in einem ökonometrischen (Mehrgleichungs-) Modell zu unterscheiden?
- (1.5) Diskutieren Sie anhand zweier ökonomischer Beispiele die Erforderlichkeit eines Einsatzes interdependenter Modelle!

2. Ökonometrische Eingleichungsmodelle

2.1 Das multiple Regressionsmodell

2.1.1 Modellspezifikation

In der ökonomischen Theorie werden Verhaltensgleichungen entwickelt, in denen eine ökonomische Variable wie z.B. der Konsum oder die Investitionen in Abhängigkeit von bestimmten Einflussgrößen betrachtet wird. So gibt z.B. die makroökonomische Konsumfunktion im keynesianischen System die Beziehung zwischen dem Privaten Verbrauch und dem verfügbaren Einkommen an. Die Habit-Persistence-Hypothese unterstellt außerdem einen Einfluss des in der Vergangenheit realisierten Konsumniveaus auf die Verbrauchsausgaben. Auch zur Erklärung des Investitionsvolumens bietet die ökonomische Theorie alternative Ansätze. Während die Akzeleratorhypothese auf die Veränderung der Nachfrage abstellt, steht bei einer Investitionsfunktion, in der die Gewinne als unabhängige Variable eingehen, die Finanzierungsseite im Mittelpunkt der Analyse des Investitionsverhaltens. In all den genannten Fällen handelt es sich um Eingleichungsmodelle, bei denen bestimmte ökonomische Hypothesen untersucht werden. Die Relevanz von Feedback-Effekten wird im Rahmen ökonometrischer Mehrgleichungsmodelle behandelt.

Bei Eingleichungsmodellen ist eine endogene Variable von einer oder mehreren exogenen Variablen abhängig. Die endogene Variable ist durch ein ökonometrisches Modell zu erklären, das sich aus einer systematischen und einer unsystematischen Komponente zusammensetzt. Bei linearen Modellen ist die systematische Komponente durch eine Linearkombination der k exogenen Variablen $x_{1t}, x_{2t}, \dots, x_{kt}$ und die unsystematische Komponente durch eine Störvariable u_t gegeben. Das ökonometrische Eingleichungsmodell lautet dann allgemein

$$(2.1.1) \quad y_t = \beta_1 + \beta_2 \cdot x_{2t} + \dots + \beta_k \cdot x_{kt} + u_t, \quad t = 1, 2, \dots, n,$$

wobei der Index t die Beobachtungsperiode oder den Beobachtungszeitpunkt angibt¹¹.

¹¹ Obwohl wir bei der Modelldarstellung primär auf zeitbezogene Variablen abstellen, kann das ökonometrische Eingleichungsmodell gleichermaßen bei Querschnittsanalysen eingesetzt werden. Der Index t würde dann z.B. für den t -ten Haushalt oder das t -te Unternehmen stehen. Allerdings stellen einige Modellkonzeptionen speziell auf zeitliche Beobachtungen ab.

Gelegentlich wird bei der methodischen Diskussion auf die gleichwertige kompaktere Modellform

$$y_t = \sum_{j=1}^k \beta_j \cdot x_{jt} + u_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$$

mit

$$\mathbf{x}'_t = \begin{pmatrix} 1 & x_{2t} & \dots & x_{kt} \end{pmatrix} \quad \text{und} \quad \boldsymbol{\beta}' = \begin{pmatrix} \beta_1 & \beta_2 & \dots & \beta_k \end{pmatrix}$$

zurückgegriffen. $\beta_1, \beta_2, \dots, \beta_k$ sind die Regressionskoeffizienten, die für den Stützbe-
reich von $t=1$ bis $t=n$ aus dem Datenmaterial zu schätzen sind. Die exogene Variable x_{1t}
hat hierin die Funktion einer Scheinvariablen, die für alle t gleich 1 gesetzt wird. Da-
durch ist der Regressionskoeffizient β_1 als absolutes Glied oder Achsenabschnitt inter-
pretierbar. Die Restgröße u_t stellt einen nicht beobachtbaren stochastischen Störterm
dar, für den die Annahmen

$$(2.1.2) \quad E(u_t) = 0 \quad \text{für alle } t=1, \dots, n$$

$$(2.1.3) \quad E(u_t^2) = \sigma^2 \quad \text{für alle } t=1, \dots, n$$

$$(2.1.4) \quad E(u_t u_{t-j}) = 0 \quad \text{für } j \neq 0$$

getroffen werden. Wegen (2.1.2) ist der Erwartungswert der Störvariablen in allen Peri-
oden gleich 0, was bedeutet, dass vom Störterm keine systematischen Einflüsse auf die
endogene Variable y_t ausgehen. Anders ausgedrückt, impliziert diese Annahme, dass
durch die k exogenen Variablen $x_{1t}, x_{2t}, \dots, x_{kt}$ sämtliche systematischen Einflüsse auf
 y_t modelliert sind. Die Annahmen (2.1.3) und (2.1.4) werden erst an späterer Stelle (Ka-
pitel 2.3) ausführlicher diskutiert, so dass hier lediglich ihre Bedeutung skizziert wird.
Wegen $E(u_t)=0$ ist die Varianz der Störvariablen durch

$$\text{Var}(u_t) = E[u_t - E(u_t)]^2 = E(u_t^2)$$

gegeben. Mit (2.1.3) wird also unterstellt, dass diese Varianz im Zeitablauf konstant
gleich σ^2 ist. Man bezeichnet diese Voraussetzung als Annahme der Homoskedastizität.

Da $E(u_t)=0$ ist, wird in (2.1.4) eine Annahme über die Kovarianz zwischen den Stör-
termen unterschiedlicher Perioden formuliert:

$$\text{Cov}(u_t, u_{t-j}) = E[(u_t - E(u_t))(u_{t-j} - E(u_{t-j}))] = E(u_t u_{t-j}) .$$

Speziell wird angenommen, dass zwischen den Störgrößen verschiedener Zeitpunkte keine linearen Abhängigkeiten bestehen, so dass ihre Kovarianz und damit auch die Korrelation gleich 0 ist. Man spricht bei Geltung von (2.1.4) auch von fehlender Autokorrelation, weil sich hier die Annahme über die Korrelation auf nur eine Zufallsvariable bezieht, die zu unterschiedlichen Zeitpunkten betrachtet wird.

Gleichung (2.1.1) gilt für alle Beobachtungszeitpunkte $t=1, \dots, n$, so dass man ausführlicher

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \dots + \beta_k x_{k1} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \dots + \beta_k x_{k2} + u_2 \\ &\vdots \\ y_n &= \beta_1 + \beta_2 x_{2n} + \beta_3 x_{3n} + \dots + \beta_k x_{kn} + u_n \end{aligned}$$

schreiben kann. Mit den Vektor- und Matrizendefinitionen

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{21} & \dots & x_{k1} \\ 1 & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & \dots & x_{kn} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

ist das obige Gleichungssystem kompakter als

$$(2.1.5) \quad \mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$$

(nx1) (nxk) (kx1) (nx1)

darstellbar. Unter den Vektoren und Matrizen wird in Klammern jeweils ihre Dimension angegeben. So ist z.B. \mathbf{X} eine $n \times k$ -Matrix, d.h. die Matrix \mathbf{X} besteht aus n Zeilen, die die Beobachtungen repräsentieren und aus k Spalten, die für die exogenen Variablen stehen. Das Element x_{jt} , $j=1, 2, \dots, k$; $t=1, \dots, n$ bezeichnet somit die Beobachtung der j -ten exogenen Variablen in der Periode t . Die erste Spalte in der \mathbf{X} -Matrix enthält die Werte der Scheinvariablen x_{1t} , die für alle t gleich 1 sind.

Mit der eingeführten Schreibweise lassen sich auch die Annahmen über die Störterme (2.1.2) bis (2.1.4) kompakter formulieren. Äquivalent wird nun

$$(2.1.6) \quad E(\mathbf{u}) = \mathbf{0}_{(nx1)}$$

$$(2.1.7) \quad \text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma^2 \mathbf{I}_{(nxn)}$$

unterstellt. Mit (2.1.6) kommt zum Ausdruck, dass der Vektor der erwarteten Störterme dem Nullvektor entspricht,

$$E \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

und somit der Erwartungswert der Störterme u_t zu jedem Zeitpunkt gleich 0 ist. Gleichung (2.1.7) enthält sowohl die Annahme homoskedastischer Störterme als auch die Annahme einer fehlenden Autokorrelation. Da \mathbf{u} ein $n \times 1$ -Vektor ist, ist der transponierte Vektor \mathbf{u}' von der Dimension $1 \times n$. Somit bezeichnet $\mathbf{u}\mathbf{u}'$ eine $n \times n$ -Matrix, die die Form

$$\mathbf{u}\mathbf{u}' = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} (u_1, u_2, \dots, u_n) = \begin{pmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_n \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_n \\ \vdots & \vdots & \cdots & \vdots \\ u_n u_1 & u_n u_2 & \cdots & u_n^2 \end{pmatrix}$$

hat. Durch Bildung des Erwartungswerts erhält man

$$\text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \begin{pmatrix} E(u_1^2) & E(u_1 u_2) & \cdots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \cdots & E(u_2 u_n) \\ \vdots & \vdots & \cdots & \vdots \\ E(u_n u_1) & E(u_n u_2) & \cdots & E(u_n^2) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & \\ \vdots & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix}$$

wobei die Annahmen (2.1.3) und (2.1.4) ausgenutzt werden. $E(\mathbf{u}\mathbf{u}')$ ist die Varianz-Kovarianz-Matrix der Störvariablen u_t , die unter den Annahmen der Homoskedastizität und fehlender Autokorrelation eine Diagonalmatrix ist. Der Skalar σ^2 kann noch vor die Matrix gezogen werden, so dass man schließlich die Darstellung

$$\text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix} = \sigma^2 \mathbf{I}$$

erhält, worin \mathbf{I} die $n \times n$ -Einheitsmatrix bezeichnet ¹².

¹² Statt Varianz-Kovarianz-Matrix wird oft einfacher die Bezeichnung Kovarianzmatrix verwendet, da die Varianz als Spezialfall der Kovarianz interpretierbar ist.

Wegen der Einbeziehung der Störterme ist das Modell (2.1.5) stochastisch. Die k erklärenden Variablen werden als feste, deterministische Größen interpretiert, so dass keinerlei Abhängigkeiten zwischen ihnen und den Störtermen bestehen. Da \mathbf{y} somit eine Linearkombination des Vektors \mathbf{u} ist, bezeichnet dieser Vektor eine n -dimensionale Zufallsvariable, deren Erwartungswert mit (2.1.6) durch den $n \times 1$ -Vektor

$$(2.1.8) \quad E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\mathbf{u}) = \mathbf{X}\boldsymbol{\beta}$$

gegeben ist. Die Störungen erfolgen daher nicht systematisch, so dass die endogene Variable im Mittel korrekt durch die exogenen Variablen erklärt wird. Für die $n \times n$ -Kovarianzmatrix von \mathbf{y} erhält man

$$E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))'] = E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})']$$

wegen (2.1.8) und

$$(2.1.9) \quad E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'] = E(\mathbf{u}\mathbf{u}') = \boldsymbol{\sigma}^2 \mathbf{I}$$

aufgrund von (2.1.7) und (2.1.5). Die Kovarianzmatrix der endogenen Variablen ist also gleich der entsprechenden Matrix der Störvariablen.

In Gleichung (2.1.5) sind die Elemente im \mathbf{y} -Vektor und in der \mathbf{X} -Matrix beobachtbar. Die Parameter im Vektor $\boldsymbol{\beta}$, die die quantitativen Einflüsse der exogenen Variablen auf die endogene Variable angeben, sind dagegen unbekannt und müssen somit geschätzt werden. Wir werden daher zunächst die grundlegende Schätztechnik zu erörtern haben.

2.1.2 Methode der kleinsten Quadrate (OLS-Methode)

Um Schätzwerte für die k Parameter im Vektor β zu erhalten, wird ein Kriterium benötigt, an dem sich der Schätzalgorithmus orientiert. Üblicherweise wird dabei von der gewöhnlichen Methode der kleinsten Quadrate (Ordinary Least Squares = OLS) ausgegangen. Die Begründung für diese Vorgehensweise liegt in den Güteeigenschaften, die diese Schätzmethode unter den gegebenen Annahmen impliziert und die später genauer diskutiert werden. Das Kriterium der kleinsten Quadrate verlangt, dass die Parameter im unbekanntem Vektor β so festzulegen sind, dass sich für die Summe der quadrierten Störterme ein Minimum ergibt. Die Summe der quadrierten Störterme ist zunächst durch

$$(2.1.10) \quad S = \sum_{t=1}^n u_t^2 = \mathbf{u}'\mathbf{u}$$

gegeben. Durch Einsetzen von (2.1.5) erhält man

$$S = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

oder

$$(2.1.11) \quad S = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta$$

aufgrund der Regeln über die Matrizenmultiplikation¹³. Die rechte Seite von (2.1.11) lässt sich noch weiter vereinfachen. Da alle Terme Skalare sind und ein Skalar nach seiner Transponierung erhalten bleibt, ist

$$\begin{pmatrix} \beta' & \mathbf{X}' & \mathbf{y} \\ 1 \times k & k \times n & n \times 1 \end{pmatrix} = \mathbf{y}'\mathbf{X}\beta$$

so dass sich schließlich als Summe der quadrierten Störterme

$$(2.1.12) \quad S = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta$$

ergibt. Das Kleinst-Quadrate-Kriterium verlangt nun, dass die Parameter im Vektor β so festzulegen sind, dass die Summe S minimiert wird. Dazu ist die Quadratsumme nach dem Vektor β zu differenzieren.

¹³ Insbesondere wird von der Regel $(\mathbf{X}\beta)' = \beta'\mathbf{X}'$ Gebrauch gemacht.

Bei der Differentiation wird die Ableitung des Skalars $\mathbf{y}'\mathbf{y}$ gleich 0, weil er nicht von $\boldsymbol{\beta}$ abhängig ist. Der Ausdruck $-2\mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ ergibt, abgeleitet nach $\boldsymbol{\beta}$, $-2\mathbf{X}'\mathbf{y}$, während die Ableitung des dritten Terms, der durch $\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ gegeben ist, $2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ ergibt¹⁴.

Zusammenfassend erhält man somit

$$(2.1.13) \quad \frac{dS}{d\boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} .$$

Notwendige Bedingung für ein Minimum von S ist, dass diese erste Ableitung nach $\boldsymbol{\beta}$ gleich dem Nullvektor $\mathbf{0}$ ist. Damit ergibt sich

$$-2\mathbf{X}'\mathbf{y} + 2(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{0}$$

und nach Division durch 2 sowie anschließender Umstellung

$$(2.1.14) \quad (\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} .$$

Mit (2.1.14) ist ein Gleichungssystem entstanden, das aus k Gleichungen besteht. Diese Gleichungen werden als Normalgleichungen bezeichnet. Die Schätzer $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ nach der Methode der kleinsten Quadrate (OLS-Schätzer) bilden nun den Vektor $\hat{\boldsymbol{\beta}}$, für den die Normalgleichungen erfüllt sind. Man erhält ihn nach Inversion der Matrix $\mathbf{X}'\mathbf{X}$,

$$(2.1.15) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} ,$$

sofern die Inverse von $\mathbf{X}'\mathbf{X}$ existiert. Wenn der in (2.1.15) unbekannte Vektor $\boldsymbol{\beta}$ durch $\hat{\boldsymbol{\beta}}$ geschätzt wird, ist die Quadratsumme der Störterme minimal. Die hinreichende Bedingung ist hier stets erfüllt, da das Minimum einer quadratischen Funktion gesucht ist.

Für die Herleitung des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ in (2.1.15) muss vorausgesetzt werden, dass die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ der symmetrischen $k \times k$ -Matrix $\mathbf{X}'\mathbf{X}$ existiert. Wie gezeigt werden kann, ist das genau dann der Fall, wenn die Datenmatrix \mathbf{X} vom Rang k ist,

$$(2.1.16) \quad \text{rg}(\mathbf{X}) = k ,$$

was voraussetzt, dass die erklärenden Variablen linear unabhängig voneinander sind. Da die Matrix $\mathbf{X}'\mathbf{X}$ stets den gleichen Rang wie die Datenmatrix \mathbf{X} hat,

¹⁴ Hierbei sind die folgenden Regeln für die Differentiation von Vektoren und Matrizen zu beachten:

- (a) Wenn $\mathbf{y}=\mathbf{a}'\mathbf{x}$ ist, dann ist $d\mathbf{y}/d\mathbf{x}=\mathbf{a}$, wobei y ein Skalar und \mathbf{a} und \mathbf{x} Vektoren sind.
- (b) Wenn $\mathbf{y}=\mathbf{x}'\mathbf{A}\mathbf{x}$ ist, wobei \mathbf{A} eine symmetrische Matrix ($\mathbf{A}=\mathbf{A}'$) bezeichnet, gilt $d\mathbf{y}/d\mathbf{x}=2\mathbf{A}\mathbf{x}$.
Speziell kommen hier diese Regeln mit $\mathbf{a}'=2\mathbf{y}'\mathbf{X}$ und $\mathbf{A}=\mathbf{X}'\mathbf{X}$ zur Anwendung.

$$\text{rg}(\mathbf{X}) = \text{rg}(\mathbf{X}'\mathbf{X}) = k,$$

ist dann eine Berechnung der Inversen $(\mathbf{X}'\mathbf{X})^{-1}$ möglich.

Wenn die Anzahl n der Beobachtungen die Anzahl k der exogenen Variablen (einschl. der Scheinvariablen) übersteigt, so ist die Voraussetzung (2.1.16) im Allgemeinen erfüllt. Denn bei empirischen Daten liegen dann in der Regel keine exakten linearen Abhängigkeiten zwischen den Beobachtungswerten der erklärenden Variablen vor¹⁵.

Wenn jedoch die Anzahl der erklärenden Variablen die Anzahl der Beobachtungen übersteigt, also $n < k$ ist, kann der Rang von \mathbf{X} maximal gleich n sein, so dass in diesem Fall

$$\text{rg}(\mathbf{X}) = \text{rg}(\mathbf{X}'\mathbf{X}) = n < k$$

gilt. Da $\mathbf{X}'\mathbf{X}$ jedoch eine $k \times k$ -Matrix ist, kann in dieser Situation ihre Inverse nicht berechnet werden, so dass keine Schätzwerte für die Regressionskoeffizienten $\beta_1, \beta_2, \dots, \beta_k$ bestimmbar sind. Eine Schätzung des Parametervektors $\boldsymbol{\beta}$ setzt somit stets voraus, dass mehr Beobachtungen vorliegen als exogene Variablen in das Regressionsmodell einbezogen werden.

Nach der OLS-Schätzung geht das ökonometrische Eingleichungsmodell (2.1.1) in die geschätzte Regressionsgleichung

$$(2.1.17) \quad y_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_{2t} + \dots + \hat{\beta}_k \cdot x_{kt} + \hat{u}_t, \quad t = 1, 2, \dots, n$$

über. Hierin geben $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ die geschätzten Regressionskoeffizienten an. \hat{u}_t ist das Residuum (= geschätzter Wert der Störgröße) der Periode t . Die Regressionswerte \hat{y}_t , die als geschätzte Werte der abhängigen Variablen y_t bei gegebenen x -Werten interpretiert werden können, lassen sich aus der Beziehung

$$(2.1.18) \quad \hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_{2t} + \dots + \hat{\beta}_k \cdot x_{kt}, \quad t = 1, 2, \dots, n$$

errechnen. Der Vektor $\hat{\mathbf{y}}$ der Regressionswerte \hat{y}_t , $\hat{\mathbf{y}} = (\hat{y}_1 \ \hat{y}_2 \ \dots \ \hat{y}_n)'$, kann aus dem Matrixprodukt $\mathbf{X} \cdot \hat{\boldsymbol{\beta}}$ bestimmt werden:

$$(2.1.19) \quad \hat{\mathbf{y}} = \mathbf{X} \cdot \hat{\boldsymbol{\beta}}.$$

¹⁵ Allerdings lassen sich oft hohe Korrelationen zwischen den exogenen Variablen feststellen, die zwar nicht die Berechnung des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ in Frage stellen, jedoch gewisse Schätzprobleme aufwerfen. Diese Problematik wird später unter dem Begriff Multikollinearität aufgegriffen und erörtert.

Mit (2.1.17) und (2.1.18) sind die Residuen \hat{u}_t durch

$$(2.1.20) \quad \hat{u}_t = y_t - \hat{y}_t, \quad t = 1, 2, \dots, n,$$

oder kompakt durch

$$(2.1.21) \quad \hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$$

mit $\hat{\mathbf{u}} = (\hat{u}_1 \ \hat{u}_2 \ \dots \ \hat{u}_n)'$ gegeben.

Das multiple Regressionsmodell (2.1.1) enthält das einfache Regressionsmodell

$$y_t = \beta_1 + \beta_2 \cdot x_t + u_t$$

im Spezialfall für $k=2$. Die Scheinvariable x_{1t} , die in allen Perioden t den Wert 1 annimmt, ist in 0 weggelassen worden. Während durch die OLS-Methode bei drei Regressoren (einschl. der Scheinvariablen) eine Regressionsebene im dreidimensionalen Raum festgelegt wird, ist bei zwei Regressoren (einschl. der Scheinvariablen) eine Regressionsgerade zu bestimmen:

$$(2.1.22) \quad \hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_t.$$

Der Ordinatenabschnitt der Regressionsgeraden, der als absolutes Glied bezeichnet wird, ist durch $\hat{\beta}_1$ gegeben. Ihre Steigung wird durch den Regressionskoeffizienten $\hat{\beta}_2$ gemessen.

Um die OLS-Schätzfunktionen $\hat{\beta}_1$ und $\hat{\beta}_2$ in expliziter Form aus (2.1.15) zu erhalten, ist zunächst einmal die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ für eine $(n \times 2)$ -Matrix \mathbf{X} zu bestimmen. Hierzu ist von der Produktmatrix $\mathbf{X}'\mathbf{X}$ auszugehen, die speziell

$$(2.1.23) \quad \mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_t \\ \sum x_t & \sum x_t^2 \end{bmatrix}$$

lautet. Summiert wird dabei stets über alle Perioden t von 1 bis n . Die Inverse von $\mathbf{X}'\mathbf{X}$ kann bestimmt werden, indem die adjungierte Matrix $(\mathbf{X}'\mathbf{X})^{\text{adj}}$ durch die Determinante $|\mathbf{X}'\mathbf{X}|$ dividiert wird:¹⁶

¹⁶ Diese Invertierungsmethode ist nicht auf die einfache Regression beschränkt, sondern sie kann allgemein in der multiplen Regression angewendet werden.

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{|\mathbf{X}'\mathbf{X}|} (\mathbf{X}\mathbf{X})^{\text{adj}} .$$

Da die Produktmatrix $\mathbf{X}'\mathbf{X}$ eine 2×2 -Matrix ist, ergibt sich ihre Determinante als Differenz aus dem Produkt der beiden Hauptdiagonalelemente und dem Produkt der beiden Nebendiagonalelemente:

$$|\mathbf{X}'\mathbf{X}| = n \cdot \sum x_t^2 - \left(\sum x_t \right)^2 .$$

Die adjungierte Matrix $(\mathbf{X}'\mathbf{X})^{\text{adj}}$ besteht aus den Adjunkten (Kofaktoren)

$$XX_{ij} = (-1)^{i+j} |\mathbf{X}\mathbf{X}_{ij}| ,$$

mit den Unterdeterminanten (Minoren) $|\mathbf{X}\mathbf{X}_{ij}|$, die nach Streichen der i -ten Zeile und j -ten Spalte der Produktmatrix $\mathbf{X}'\mathbf{X}$ zu berechnen sind. Für die 2×2 -Matrix $\mathbf{X}'\mathbf{X}$, deren Struktur aus (2.1.23) hervorgeht, erhält man

$$\begin{aligned} XX_{11} &= \sum x_t^2 \\ XX_{12} &= -\sum x_t \\ XX_{21} &= -\sum x_t \\ XX_{22} &= n \end{aligned} ,$$

so dass die adjungierte Matrix $(\mathbf{X}'\mathbf{X})^{\text{adj}}$ als transponierte Matrix der Adjunkten durch

$$(\mathbf{X}'\mathbf{X})^{\text{adj}} = \begin{bmatrix} XX_{11} & XX_{21} \\ XX_{12} & XX_{22} \end{bmatrix} = \begin{bmatrix} \sum x_t^2 & -\sum x_t \\ -\sum x_t & n \end{bmatrix}$$

gegeben ist. Somit lautet die Inverse von $\mathbf{X}'\mathbf{X}$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n \cdot \sum x_t^2 - \left(\sum x_t \right)^2} \begin{bmatrix} \sum x_t^2 & -\sum x_t \\ -\sum x_t & n \end{bmatrix} .$$

Außerdem ergibt das Produkt aus der Transponierten von \mathbf{X} und dem Vektor \mathbf{y} einen 2×1 -Vektor:

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_t \\ \sum x_t y_t \end{bmatrix} .$$

Damit nimmt der OLS-Schätzvektor $\hat{\boldsymbol{\beta}}$ bei der einfachen Regression die Form

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \frac{1}{n \sum x_t^2 - (\sum x_t)^2} \begin{bmatrix} \sum x_t^2 \sum y_t - \sum x_t \sum x_t y_t \\ n \sum x_t y_t - \sum x_t \sum y_t \end{bmatrix}$$

an. Der Ordinatenabschnitt ist demzufolge durch

$$(2.1.24) \quad \hat{\beta}_1 = \frac{\sum x_t^2 \sum y_t - \sum x_t \sum x_t y_t}{n \sum x_t^2 - (\sum x_t)^2}$$

und das Steigungsmaß durch

$$(2.1.25) \quad \hat{\beta}_2 = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{n \sum x_t^2 - (\sum x_t)^2}$$

gegeben.

Durch elementare Umformung lässt sich zeigen, dass das Steigungsmaß $\hat{\beta}_2$ dem Quotienten aus der Kovarianz s_{xy} zwischen den Werten der abhängigen Variablen y_t und der unabhängigen Variablen x_t und der Varianz s_x^2 letzter Größe entspricht:

$$(2.1.26) \quad \hat{\beta}_2 = \frac{s_{xy}}{s_x^2}$$

mit

$$s_{xy} = \frac{1}{n} \sum (x_t - \bar{x})(y_t - \bar{y})$$

und

$$s_x^2 = \frac{1}{n} \sum (x_t - \bar{x})^2.$$

Die Kennzahlen \bar{x} und \bar{y} bezeichnen die Mittelwerte beider Größen in der Stichprobe. Dass $\hat{\beta}_2$ in der Form (2.1.26) darstellbar ist, ergibt sich unmittelbar aus den Verschiebungssätzen für die Kovarianz und Varianz:

$$s_{xy} = \frac{\sum x_t y_t}{n} - \frac{\sum x_t}{n} \cdot \frac{\sum y_t}{n}$$

und

$$s_x^2 = \frac{1}{n} \sum x_t^2 - \left(\frac{\sum x_t}{n} \right)^2.$$

Gleichermaßen lässt sich zeigen, dass der Achsenabschnitt unter Verwendung des Steigungsmaßes $\hat{\beta}_2$ von der Form

$$(2.1.27) \quad \hat{\beta}_1 = \frac{\sum y_t}{n} - \hat{\beta}_2 \frac{\sum x_t}{n} = \bar{y} - \hat{\beta}_2 \cdot \bar{x}$$

ist. Die Identität von (2.1.24) und (2.1.27) ergibt sich unmittelbar durch Einsetzen von (2.1.25) in (2.1.27) nach einigen elementaren Umformungen. Löst man Gleichung (2.1.27) nach \bar{y} auf, so erhält man die Beziehung

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \cdot \bar{x} ,$$

aus der ersichtlich ist, dass die Regressionsgerade durch den Schwerpunkt (\bar{x}, \bar{y}) verlaufen muss.

Beispiel 2.1. 1: Das verfügbare Einkommen Y_t^v wird als entscheidende Einflussgröße des privaten Verbrauchs C_t angesehen. Man erhält es, indem man das Bruttoeinkommen von Steuern und Sozialabgaben bereinigt, die bei den privaten Haushalten nicht ausgabenwirksam werden können, und außerdem die Transferzahlungen berücksichtigt. Wenn potentielle Einflussgrößen, wie z.B. die Einkommensverteilung, das Vermögen, der Zinssatz für Konsumentenkredite oder die soziale Stellung vernachlässigbar sind, kann die makroökonomische Konsumfunktion unter Einbeziehung einer Störvariablen u_t mit einem Erwartungswert von Null ökonometrisch geschätzt werden:

$$(2.1.28) \quad C_t = C_0 + c_1 \cdot Y_t^v + u_t .$$

Der Parameter c_1 ist aus der Makroökonomik als marginale Konsumneigung bekannt:

$$\frac{dC_t}{dY_t^v} = c_1 .$$

Er gibt die Veränderung der Konsumausgaben an, die durch eine Erhöhung oder Verringerung des verfügbaren Einkommens um 1 Einheit induziert wird. Der Ordinatenabschnitt C_0 steht dagegen für den autonomen Konsum, der vom verfügbaren Einkommen unabhängig ist.

Die Konsumfunktion soll für die Bundesrepublik Deutschland für den Zeitraum von 1974 bis 1992 unter Verwendung von Jahresdaten mit der OLS-Methode geschätzt werden. Während der Private Verbrauch in Preisen des Jahres 1985 vorliegt, ist das verfügbare

bare Einkommen in jeweiligen Preisen angegeben. Eine Preisbereinigung kann unter Verwendung des Preisindex des Privaten Verbrauchs vorgenommen werden.

Jahr	Verfügbares Einkommen Mrd. DM (in jeweiligen Preisen)	Preisindex Privater Verbrauch (1985=100)	Privater Verbrauch Mrd. DM (in Preisen von 1985)
1974	613,25	63,7	837,60
1975	667,95	67,5	863,82
1976	716,45	70,4	897,32
1977	760,72	72,7	937,80
1978	816,76	74,7	971,48
1979	881,03	77,9	1003,06
1980	921,94	82,4	1015,37
1981	961,89	87,7	1007,92
1982	995,09	92,3	992,55
1983	1036,10	95,4	1005,92
1984	1078,16	98,0	1021,68
1985	1119,93	100,0	1036,53
1986	1199,41	99,5	1072,01
1987	1240,56	100,1	1106,88
1988	1319,22	101,5	1137,00
1989	1384,49	104,6	1167,37
1990	1525,65	107,4	1230,68
1991	1580,15	111,5	1274,63
1992	1648,80	116,0	1287,11

Quelle: Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung,
Jahresgutachten 1993/94

Es bezeichne y_t den realen Wert des Privaten Verbrauchs (C_t) und x_t den realen Wert des verfügbaren Einkommens (Y_t^v) im Jahr t . Die Beobachtungsmatrix X und der Vektor y der abhängigen Variablen sind dann durch

$$\mathbf{X} = \begin{bmatrix} 1 & 962,72 \\ 1 & 989,56 \\ 1 & 1017,68 \\ 1 & 1046,38 \\ 1 & 1093,39 \\ 1 & 1130,98 \\ 1 & 1118,86 \\ 1 & 1096,80 \\ 1 & 1078,10 \\ 1 & 1086,06 \\ 1 & 1100,16 \\ 1 & 1119,93 \\ 1 & 1205,44 \\ 1 & 1239,32 \\ 1 & 1299,72 \\ 1 & 1323,60 \\ 1 & 1420,53 \\ 1 & 1417,17 \\ 1 & 1421,38 \end{bmatrix} \quad \text{und} \quad \mathbf{y} = \begin{bmatrix} 837,60 \\ 863,82 \\ 897,32 \\ 937,80 \\ 971,48 \\ 1003,06 \\ 1015,57 \\ 1007,92 \\ 992,55 \\ 1005,92 \\ 1021,68 \\ 1036,53 \\ 1072,01 \\ 1106,88 \\ 1137,00 \\ 1167,37 \\ 1230,68 \\ 1274,63 \\ 1287,11 \end{bmatrix}$$

gegeben. Der OLS-Schätzer

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

enthält dann die Schätzwerte für die Parameter C_0 und c_1 der Konsumfunktion. Da es sich bei der Konsumfunktion jedoch um ein einfaches Regressionsmodell handelt, lassen sich die OLS-Schätzer \hat{C}_0 und \hat{c}_1 für den autonomen Konsum und die marginale Konsumneigung unmittelbar aus (2.1.24) und (2.1.25) ermitteln. Hierzu bietet es sich an, die dafür benötigten Summen in einer Arbeitstabelle zu berechnen:

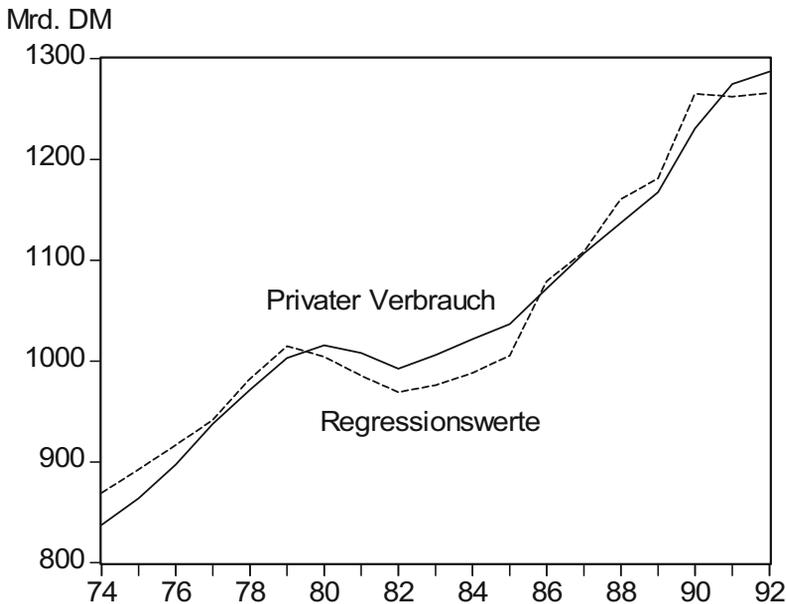
t	x_t	y_t	x_t^2	$x_t \cdot y_t$
1	962,72	837,60	926829,80	806374,27
2	989,56	863,82	979228,99	854801,72
3	1017,68	897,32	1035672,58	913184,62
4	1046,38	937,80	1094911,10	981295,16
5	1093,39	971,48	1195501,69	1062206,52
6	1130,98	1003,06	1279115,76	1134440,80
7	1118,86	1015,57	1251847,70	1136280,65
8	1096,80	1007,92	1202970,24	1105486,66
9	1078,10	992,55	1162299,61	1070068,16
10	1086,06	1005,92	1179526,32	1092489,48
11	1100,16	1021,68	1210352,03	1124011,47
12	1119,93	1036,53	1254243,21	1160841,04
13	1205,44	1072,01	1453085,59	1202243,73
14	1239,32	1106,88	1535914,06	1371778,52
15	1299,72	1137,00	1689272,08	1477781,64
16	1323,60	1167,37	1751916,96	1545130,93
17	1420,53	1230,68	2017905,48	1748217,86
18	1417,17	1274,63	2008370,81	1806367,40
19	1421,38	1287,11	2020321,10	1829472,41
Σ	22167,78	19866,93	26249285,11	23512473,04

Bei 19 Jahreswerten ($n=19$) erhält man damit

$$\begin{aligned}\hat{C}_0 &= \frac{\sum x_t^2 \sum y_t - \sum x_t \sum x_t y_t}{n \sum x_t^2 - (\sum x_t)^2} \\ &= \frac{26249285,11 \cdot 19866,93 - 22167,78 \cdot 23512473,04}{19 \cdot 26249285,11 - (22167,78)^2} \\ &= \frac{273380223,76}{7325946,96} = 37,324603\end{aligned}$$

und

$$\begin{aligned}\hat{C}_1 &= \frac{n \cdot \sum x_t y_t - \sum x_t \sum y_t}{n \sum x_t^2 - (\sum x_t)^2} \\ &= \frac{19 \cdot 23512473,04 - 22167,78 \cdot 19866,93}{19 \cdot 26249285,11 - (22167,78)^2} \\ &= \frac{6331254,26}{7325946,96} = 0,8642164\end{aligned}$$

Abbildung 2.1. 1: Konsumfunktion

Der Achsenabschnitt C_0 beträgt also 37,325 und das Steigungsmaß c_1 nimmt den Wert 0,864 an.

Eine ökonomische Interpretation des Achsenabschnitts wäre eigentlich nur dann sinnvoll, wenn der Stützbereich den Koordinatenursprung mit einschließen würde, was bei einer makroökonomischen Konsumfunktion jedoch nicht gegeben ist. Insofern ist bei der Interpretation des Wertes 37,325 als autonomer Konsum der Volkswirtschaft Vorsicht geboten. Jedenfalls ist er nicht als Mindestkonsum im Sinne eines Subsistenzniveaus zu interpretieren.

Gut interpretierbar ist dagegen das Steigungsmaß. Der Schätzwert von 0,864, der die marginale Konsumneigung im Stützbereich angibt, besagt, dass im betrachteten Zeitraum ein Einkommenszuwachs von 1 Mrd. DM im Mittel zu einer Erhöhung der Konsumausgaben um 864 Mill. DM geführt hat. Wenn die Konsumfunktion stabil ist, kann davon ausgegangen werden, dass etwa 86 % des zusätzlichen verfügbaren Einkommens für Konsumausgaben verwendet werden. Aufgrund der Komplementarität ergibt sich daraus eine marginale Sparquote von 14 %.

◆

Beispiel 2.1. 2: Die Nachfrage nach Geld ist Gegenstand vielfältiger geldtheoretischer und geldpolitischer Diskussionen. So kommt z.B. dem Problem der Stabilität der Geldnachfragefunktion bei einer Beurteilung der Effizienz geldpolitischer Maßnahmen ein hoher Stellenwert zu. Eine Spezifikation der Geldnachfragefunktion setzt eine Definition der relevanten Geldmenge voraus. Außerdem ist zu klären, ob sich die Modellierung auf die nominale oder reale Geldmenge beziehen soll.

Als entscheidende Einflussgrößen für die Nachfrage nach Geld werden das Transaktionsvolumen und die Opportunitätskosten der Geldhaltung angesehen: Während das Transaktionsvolumen im allgemeinen durch das Brutto sozialprodukt y gemessen wird, werden die Opportunitätskosten durch die Renditen alternativer Aktiva erfasst, die meist durch einen repräsentativen Zinssatz r charakterisiert werden. Bei fehlender Geldillusion lässt sich die Nachfrage nach Geld m dann durch die Funktion

$$m = f(y, r)$$

wiedergeben, in der m und y reale Größen sind, die man dadurch erhält, dass man die nominale Geldmenge M und das nominale Sozialprodukt Y auf das Preisniveau P bezieht:

$$m = \frac{M}{P} \quad \text{und} \quad y = \frac{Y}{P} .$$

Sofern die Einheitsperiode hinreichend lang ist, können Anpassungsverzögerungen unberücksichtigt bleiben. Bei einem logarithmisch-linearen Ansatz, den man bei empirischen Untersuchungen zur Geldnachfrage häufig zugrunde legt, lautet die ökonometrische Geldnachfragefunktion dann

$$(2.1.29) \quad \ln m_t = \beta_1 + \beta_2 \ln y_t + \beta_3 \ln r_t + u_t$$

mit der Störvariablen u_t , die die klassischen Modellannahmen erfüllen soll. In diesem Modell lassen sich die Regressionskoeffizienten β_2 und β_3 unmittelbar als Einkommens- und Zinselastizität interpretieren. Da der Geldbedarf mit zunehmendem Transaktionsvolumen steigt, wird in der ökonomischen Theorie eine positive Einkommenselastizität vorausgesetzt. Andererseits ist eine negative Zinselastizität zu erwarten, da die Opportunitätskosten der Geldhaltung mit steigendem Zinssatz zunehmen.

Bei der ökonometrischen Schätzung der Geldnachfragefunktion wird die Geldmenge M_1 , die sich aus dem Bargeld und den Sichtguthaben des Publikums zusammensetzt, als

zu erklärende Variable verwendet. Der repräsentative Zinssatz wird durch den FIBOR (Frankfurt interbank offered rate) als kurzfristigem Zinssatz erfasst. Als Deflator zur Bestimmung der realen Geldmenge wird der Preisindex für das Brutto Sozialprodukt zur Basis 1985 herangezogen. Im Beobachtungszeitraum von 1970 bis 1989 ergibt sich damit folgende Datenbasis:

Jahr	Geldmenge M_1 Mrd. DM	Brutto Sozialprodukt Mrd. DM (in Preisen des Jahres 1985)	Kurzfristiger Zinssatz %	Preisindex für das Brutto Sozialprodukt (1985=100)
1970	108,219	1322,8	9,41	51,1
1971	121,522	1363,1	7,15	55,1
1972	139,298	1422,3	5,61	58,0
1973	142,862	1491,1	12,14	61,6
1974	158,432	1491,9	9,90	65,9
1975	179,898	1473,0	4,96	69,8
1976	186,852	1554,7	4,25	72,3
1977	208,076	1594,4	4,37	75,0
1978	237,909	1649,4	3,70	78,2
1979	247,869	1715,9	6,69	81,2
1980	257,335	1733,8	9,54	85,2
1981	255,277	1735,7	12,11	88,7
1982	273,047	1716,5	8,88	92,6
1983	295,795	1748,4	5,78	95,8
1984	314,235	1802,0	5,99	97,9
1985	329,737	1834,5	5,44	100,0
1986	358,747	1874,4	4,64	103,3
1987	385,170	1902,3	4,03	105,3
1988	426,997	1971,8	4,33	106,9
1989	450,746	2046,8	7,12	109,7

Quellen: Monatsberichte der Deutschen Bundesbank (verschiedene Hefte); Jahresgutachten 1991/92 des Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung

Da die Geldnachfragefunktion eine logarithmisch-lineare Form besitzt, sind zunächst die Logarithmen der Beobachtungswerte zu ermitteln ¹⁷:

¹⁷ Wir arbeiten hier mit den natürlichen Logarithmen der Variablen.

t	$\ln m_t$	$\ln y_t$	$\ln r_t$
1	5,355543	7,187506	2,241773
2	5,396116	7,217517	1,967112
3	5,481343	7,260031	1,724551
4	5,446387	7,307270	2,496506
5	5,482357	7,307806	2,292535
6	5,551926	7,295056	1,601406
7	5,554663	7,349038	1,446919
8	5,625586	7,374253	1,474763
9	5,717789	7,408167	1,308333
10	5,721156	7,447693	1,900614
11	5,710547	7,458071	2,255493
12	5,662260	7,459166	2,494031
13	5,686525	7,448042	2,183802
14	5,732574	7,466456	1,754404
15	5,771365	7,496653	1,790091
16	5,798295	7,514527	1,693779
17	5,850150	7,536044	1,534714
18	5,902041	7,550819	1,393766
19	5,990053	7,586702	1,465567
20	6,018325	7,624033	1,962908

Um die Parameter β_1 , β_2 und β_3 mit der OLS-Methode zu bestimmen, benötigen wir zunächst einmal die Produktmatrizen $\mathbf{X}'\mathbf{X}$ und $\mathbf{X}'\mathbf{y}$:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum \ln y_t & \sum \ln r_t \\ \sum \ln y_t & \sum (\ln y_t)^2 & \sum \ln y_t \cdot \ln r_t \\ \sum \ln r_t & \sum \ln r_t \cdot \ln y_t & \sum (\ln r_t)^2 \end{bmatrix} = \begin{bmatrix} 20,000 & 148,295 & 36,983 \\ 148,295 & 1099,858 & 273,976 \\ 36,983 & 273,976 & 71,041 \end{bmatrix}$$

und

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum \ln m_t \\ \sum \ln y_t \cdot \ln m_t \\ \sum \ln r_t \cdot \ln m_t \end{bmatrix} = \begin{bmatrix} 113,455 \\ 841,671 \\ 209,244 \end{bmatrix}.$$

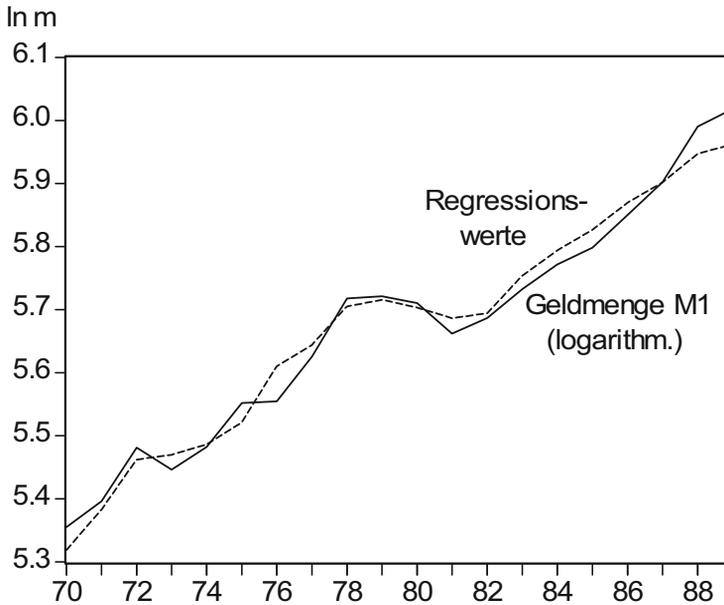
Die Inverse der Produktmatrix $\mathbf{X}'\mathbf{X}$ lautet

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 218,194 & -28,592 & -3,322 \\ -28,592 & 3,770 & 0,346 \\ -3,322 & 0,346 & 0,409 \end{bmatrix}.$$

Der OLS-Schätzvektor $\hat{\beta}$ lässt sich nun unmittelbar als Produkt der Matrizen $(\mathbf{X}'\mathbf{X})^{-1}$ und $(\mathbf{X}'\mathbf{y})$ bestimmen:

$$\hat{\beta}_{3 \times 1} = \begin{bmatrix} \hat{\beta}_1 = -4,806 \\ \hat{\beta}_2 = 1,432 \\ \hat{\beta}_3 = -0,076 \end{bmatrix}.$$

Abbildung 2.1. 2: Geldnachfragefunktion



Während das konstante Glied $\hat{\beta}_1$ bei der ökonomischen Interpretation nicht von Belang ist, richtet sich das Interesse auf die Schätzer $\hat{\beta}_2$ und $\hat{\beta}_3$. Zunächst einmal weisen beide Regressionskoeffizienten im Stützbereich das erwartete Vorzeichen auf: Die Geldnachfrage steigt mit zunehmendem Einkommen und sinkt mit zunehmendem Zinssatz. Darüber hinaus lassen sich die Regressionskoeffizienten in unserem Modell als Elastizitäten interpretieren. Aus dem Wert der Einkommenselastizität der Geldnachfrage ist zu entnehmen, dass die Geldnachfrage überproportional auf Veränderungen des Einkommens reagiert hat. Eine 1%ige Erhöhung des Einkommens ging im betrachteten Zeitraum im Mittel mit einer etwa 1,4%igen Erhöhung der realen Geldnachfrage einher. Hierin zeigen sich "diseconomies of scale", da die Geldhaltung bei einer Steigerung des Einkommens relativ stärker zugenommen hat als die Haltung ertragbringender Aktiva. Damit bestätigt sich die Luxusguthypothese, nach der bei steigendem Vermögen und Einkommen ein überproportionaler Anteil als Kasse gehalten wird, da die Tauschbereit-

schaft als Luxusbedürfnis angegeben wird (vgl. Ehrlicher, 1972, S. 384). Der Wert der Zinselastizität der Geldnachfrage von -0,076 gibt dagegen an, dass sich die Geldnachfrage bei einer 10%igen Zinserhöhung im Mittel um etwa 0,8 % verringert hat. Das Publikum hat eine Erhöhung der Opportunitätskosten der Geldhaltung somit zum Anlass genommen, sein Portfolio zugunsten Ertrag bringender Finanzanlagen umzuschichten. ♦

Wenn man den Parametervektor $\hat{\beta}$ nach dem Kleinst-Quadrate-Kriterium ermittelt hat, lässt sich der Vektor \hat{u} der Residuen, der die Schätzfehler \hat{u}_t enthält, wegen (2.1.21) und (2.1.19) durch

$$(2.1.30) \quad \hat{u} = y - X\hat{\beta}$$

berechnen. Die n Residuen $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n$ im $n \times 1$ -Vektor \hat{u} lassen sich als realisierte Störterme interpretieren. Da $\hat{\beta}$ nach der OLS-Methode entwickelt wurde, spricht man in diesem Zusammenhang auch von OLS-Residuen. Wird Gleichung (2.1.23) von links mit der Transponierten von X multipliziert, ergibt sich

$$(2.1.31) \quad X'\hat{u} = X'y - X'X\hat{\beta} = \mathbf{0} .$$

Das Produkt aus der transponierten Beobachtungsmatrix X' und dem Residuenvektor \hat{u} ist danach gleich einem $k \times 1$ -Nullvektor $\mathbf{0}$. Diese Bedingung ist aufgrund der Normalgleichungen (2.1.14) erfüllt, aus denen der OLS-Parametervektor $\hat{\beta}$ berechenbar ist. Für die n OLS-Residuen gelten also k lineare Restriktionen der Form

$$X'\hat{u} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^n \hat{u}_t \\ \sum_{t=1}^n X_{2t} \hat{u}_t \\ \vdots \\ \sum_{t=1}^n X_{kt} \hat{u}_t \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} ,$$

die zum Ausdruck bringen, dass die Residuen \hat{u} orthogonal auf den exogenen Variablen stehen. Aus der ersten Restriktion

$$\sum_{t=1}^n \hat{u}_t = 0$$

ist ersichtlich, dass die Summe der OLS-Residuen und daher auch ihr arithmetisches Mittel

$$(2.1.32) \quad \frac{1}{n} \sum_{t=1}^n \hat{u}_t = \bar{\hat{u}}$$

gleich 0 sind. Diese Implikation ist mit der Annahme über die Störterme $E(u_t)=0$ gut verträglich. Sie ergibt sich tatsächlich jedoch nur dann, wenn im Regressionsansatz ein Absolutglied berücksichtigt wird. Die $k-1$ restlichen Restriktionen, die

$$(2.1.33) \quad \sum_{t=1}^n x_{jt} \hat{u}_t = 0 \quad \text{für } j=2, \dots, k$$

lauten, lassen sich ebenfalls gut interpretieren. Sie verkörpern nämlich implizit die Aussage, dass die Kovarianz $s_{\hat{u},x}$ zwischen den OLS-Residuen und jeder exogenen Variablen gleich 0 ist. Denn unter Berücksichtigung von (2.1.32) gilt

$$s_{\hat{u},x} = \frac{1}{n} \sum_{t=1}^n (\hat{u}_t - \bar{\hat{u}})(x_{jt} - \bar{x}_j) = \frac{1}{n} \sum_{t=1}^n \hat{u}_t (x_{jt} - \bar{x}_j) = \frac{1}{n} \sum_{t=1}^n \hat{u}_t x_{jt} - \bar{x}_j \frac{1}{n} \sum_{t=1}^n \hat{u}_t$$

mit
$$\bar{x}_j = \frac{1}{n} \sum_{t=1}^n x_{jt} ,$$

woraus

$$s_{\hat{u},x} = \frac{1}{n} \sum_{t=1}^n \hat{u}_t \cdot x_{jt}$$

folgt. Die OLS-Residuen sind also mit den exogenen Variablen nicht korreliert. Wie sich zeigen wird, kommt diesem Befund bei der Diskussion der Güteeigenschaften des OLS-Schätzers $\hat{\beta}$ eine erhebliche Bedeutung zu.

2.1.3 Schätzeigenschaften der OLS-Methode

2.1.3.1 Gütekriterien

Die Eignung der OLS-Methode zur Schätzung ökonometrischer Modelle hängt von den Eigenschaften ihrer Schätzfunktionen ab. Ihre Anwendung ist daher durch die Qualität der Schätzung zu begründen. Bei der Beurteilung der Güte der Schätzung unterscheidet man Schätzeigenschaften, die vom Stichprobenumfang unabhängig sind, und solche, die

ausschließlich bei großem Stichprobenumfang gelten. Insbesondere sind die Eigenschaften im ersteren Fall auch bei kleinen Stichproben gültig, die häufig in der ökonomischen Praxis vorzufinden sind. Güteeigenschaften, die allein bei großen Stichproben Gültigkeit besitzen, bezeichnet man als asymptotische Eigenschaften.

Wir diskutieren hier als Güteeigenschaften, die unabhängig vom Stichprobenumfang sind, die **Erwartungstreue und Effizienz** des OLS-Schätzers $\hat{\beta}$. Wenn $\hat{\beta}$ erwartungstreu ist, dann wird er zwar im Einzelfall nicht völlig mit dem Vektor β der "wahren" Regressionskoeffizienten identisch sein. Wenn man die OLS-Schätzer $\hat{\beta}$ aber über alle möglichen Stichproben gleichen Umfangs mitteln würde, so würden die Unterschiede jedoch bei einem erwartungstreuen Schätzer verschwinden. Der Vektor β der unbekanntenen Regressionskoeffizienten könnte dann durch den OLS-Schätzer $\hat{\beta}$ unverzerrt geschätzt werden.

Damit ist jedoch noch nichts über die Genauigkeit der Schätzung ausgesagt. Je kleiner die Varianzen der geschätzten Regressionskoeffizienten sind, umso präziser kann β geschätzt werden. Ein Schätzer mit der kleinstmöglichen Varianz heißt effizient. Wir werden uns hier bei der Effizienzüberprüfung auf die Klasse der linearen Schätzfunktionen beschränken. Wie sich nämlich zeigen lässt, haben die Komponenten des OLS-Schätzers $\hat{\beta}$ unter allen linearen unverzerrten Schätzern die kleinstmöglichen Varianzen. Sie sind daher beste lineare unverzerrte Schätzfunktionen, was als BLUE-Eigenschaft (best linear unbiased estimator) bezeichnet wird. Dies ist die Aussage des **Gauß-Markow-Theorems**, das hier aus didaktischen Gründen in die Bereiche Linearität, Erwartungstreue und Effizienz aufgespalten wird.

Welches Verhalten ist von einer guten Schätzfunktion zu erwarten, wenn der Stichprobenumfang nach und nach vergrößert wird? Offenbar würde ein zunehmender Stichprobenumfang Vorteile mit sich bringen, wenn die zusätzliche Information zum Zwecke einer präziseren Schätzung genutzt werden könnte. Wenn der Stichprobenumfang immer größer wird, sollte die Schätzfunktion dann schließlich mit dem unbekanntenen Parametervektor β zusammenfallen. Ein Schätzer, der diese Eigenschaft besitzt, heißt konsistent. Die **Konsistenz** kennzeichnet einen Grenzzustand, der theoretisch erst dann erreicht ist, wenn der Stichprobenumfang über alle Grenzen wächst.

2.1.3.2 Linearität

Das **Gauß-Markow-Theorem** macht eine Aussage über bestimmte Schätzeigenschaften innerhalb der Klasse der linearen Schätzer für den Parametervektor β . Allgemein lässt sich ein beliebiger linearer Schätzer $\tilde{\beta}$ für β in der Form

$$(2.1.34) \quad \tilde{\beta} = Cy + a$$

schreiben, wobei C ein $k \times n$ -Matrix und a ein $k \times 1$ -Vektor ist. Während die Komponenten y_t des Vektors y in Abhängigkeit von den Störvariablen u_t Zufallsvariablen sind, setzen sich die Matrix C und der Vektor a aus konstanten Elementen zusammen. Da β aus dem Datenmaterial geschätzt wird, sind die Elemente der Matrix C genauer Funktionen der exogenen Variablen x_{jt} . Wie sich später zeigen wird, kommen für die BLUE-Eigenschaft allein lineare Schätzfunktionen in Betracht, bei denen der Vektor a dem Nullvektor $\mathbf{0}$ entspricht.

Der OLS-Schätzer

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'y$$

hat die Form (2.1.34) mit

$$C = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \quad \text{und} \quad a = \mathbf{0}.$$

Die Matrix C setzt sich hierbei aus dem Produkt der Inversen $(\mathbf{X}'\mathbf{X})^{-1}$ und der transponierten Beobachtungsmatrix \mathbf{X} zusammen, so dass die Elemente von C Funktionen der exogenen Variablen x_{jt} sind. Die Linearität der Schätzfunktion $\hat{\beta}$ ergibt sich dann unter der Voraussetzung, dass die Werte der exogenen Variablen als feste Größen angesehen werden können.

2.1.3.3 Erwartungstreue

Ein Schätzer ist erwartungstreu, wenn sein Erwartungswert mit dem Parameter der Grundgesamtheit übereinstimmt. Wenn sich Schätzfehler schon nicht vermeiden lassen, so ist ein erwartungstreuer Schätzer zumindest frei von systematischen Verzerrungen. Es ist dann sichergestellt, dass der Schätzfehler allein aus stichprobenbedingten Zufallsschwankungen um den unbekanntem Parameter der Grundgesamtheit besteht.

Um die Erwartungstreue des OLS-Schätzers $\hat{\beta}$ zu zeigen, setzen wir zunächst das Regressionsmodell (2.1.5) in (2.1.15) ein:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) .$$

Nach Ausmultiplizieren erhält man

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} ,$$

so dass

$$(2.1.35) \quad \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}$$

folgt.

Nun kann der Erwartungswert von $\hat{\boldsymbol{\beta}}$ leicht bestimmt werden. Aufgrund der Linearität des Erwartungswertoperators erhält man

$$E(\hat{\boldsymbol{\beta}}) = E[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}] = E(\boldsymbol{\beta}) + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}] .$$

Da der Erwartungswert einer Konstanten gleich der Konstanten selbst ist, folgt

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}] .$$

Für den letzten Schritt ist noch zu berücksichtigen, dass die exogenen Variablen feste, nicht-stochastische Größen sind. Man kann daher

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{u})$$

schreiben, so dass sich mit (2.1.6) schließlich

$$(2.1.36) \quad E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

ergibt. Damit ist die Erwartungstreue des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ gezeigt. Aus der ausführlichen Schreibweise von (2.1.36),

$$\begin{pmatrix} E(\hat{\beta}_1) \\ E(\hat{\beta}_2) \\ \vdots \\ E(\hat{\beta}_k) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} ,$$

ist ersichtlich, dass die Erwartungswerte der OLS-Schätzer $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ mit den unbekanntenen Regressionskoeffizienten des ökonomischen Modells (2.1.1) übereinstimmen. Das bedeutet, dass der unbekannte Parametervektor $\boldsymbol{\beta}$ des Regressionsmodells (2.1.5) mit dem OLS-Schätzer $\hat{\boldsymbol{\beta}}$ unverzerrt geschätzt werden kann.

2.1.3.4 Kovarianzmatrix des OLS-Schätzers $\hat{\beta}$

Die Kovarianzmatrix $\text{Cov}(\hat{\beta})$ des OLS-Schätzers $\hat{\beta}$, die eine symmetrische $k \times k$ -Matrix ist, wird bei der Diskussion der Effizienz und Konsistenz benötigt. Aus diesem Grund bietet es sich an, ihre Struktur vorab zu betrachten, womit eine Grundlage zur Diskussion der beiden genannten Güteeigenschaften gegeben wird. Für die Effizienz sind insbesondere die Varianzen $\text{Var}(\hat{\beta}_j)$ der geschätzten Regressionskoeffizienten $\hat{\beta}_j$ von Bedeutung, die in der Hauptdiagonale der Kovarianzmatrix von $\hat{\beta}$ stehen. Sie werden außerdem bei der Konstruktion von Konfidenzintervallen und Signifikanztests benötigt. Außerhalb der Hauptdiagonale stehen dagegen die Kovarianzen $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j)$ zwischen den Schätzern $\hat{\beta}_i$ und $\hat{\beta}_j$, $i \neq j$.

Bei der Definition der Kovarianzmatrix $\text{Cov}(\hat{\beta})$ geht man von den Abweichungen des OLS-Schätzers $\hat{\beta}$ von seinem Erwartungswert $E(\hat{\beta})$ aus. Genauer erhält man die Varianzen und Kovarianzen in der entsprechenden Anordnung aus der Definition

$$(2.1.37) \quad \text{Cov}(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\}$$

Wegen (2.1.36) kann (2.1.37) gleichwertig in der Form

$$(2.1.38) \quad \text{Cov}(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

geschrieben werden. Die ausführliche Darstellung von (2.1.38) lautet:

$$\text{Cov}(\hat{\beta}) = \begin{bmatrix} E(\hat{\beta}_1 - \beta_1)^2 & E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2)] & \cdots & E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k)] \\ E[(\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1)] & E(\hat{\beta}_2 - \beta_2)^2 & \cdots & E[(\hat{\beta}_2 - \beta_2)(\hat{\beta}_k - \beta_k)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(\hat{\beta}_k - \beta_k)(\hat{\beta}_1 - \beta_1)] & E[(\hat{\beta}_k - \beta_k)(\hat{\beta}_2 - \beta_2)] & \cdots & E(\hat{\beta}_k - \beta_k)^2 \end{bmatrix}$$

mit

$$\text{Var}(\hat{\beta}_j) = E(\hat{\beta}_j - \beta_j)^2$$

und

$$\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = E[(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)], \quad i \neq j.$$

Unter Verwendung von Gleichung (2.1.35) lässt sich die Struktur der Kovarianz-Matrix wie folgt bestimmen:

$$\begin{aligned}
\text{Cov}(\hat{\boldsymbol{\beta}}) &= E\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right] \\
&= E\left\{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right]'\right\} \\
&= E\left\{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]'\right\}.
\end{aligned}$$

Da bei symmetrischen Matrizen die Operationen des Transponierens und der Invertierung in ihrer Reihenfolge vertauschbar sind, gilt $[(\mathbf{X}'\mathbf{X})^{-1}]' = [(\mathbf{X}'\mathbf{X})']^{-1}$. Weiter ist $(\mathbf{X}'\mathbf{X})' = \mathbf{X}'\mathbf{X}$, so dass insgesamt $[(\mathbf{X}'\mathbf{X})^{-1}]' = (\mathbf{X}'\mathbf{X})^{-1}$ ist. Damit hat man

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

und mit (2.1.7)

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

woraus schließlich

$$(2.1.39) \quad \text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

folgt. Mithin ergibt sich die Varianz-Kovarianz-Matrix von $\hat{\boldsymbol{\beta}}$ als Produkt aus der konstanten Varianz σ^2 der Störvariablen und der Inversen der Matrix $\mathbf{X}'\mathbf{X}$.

Wenn man die Diagonalelemente der inversen Produktmatrix $(\mathbf{X}'\mathbf{X})^{-1}$ mit x^{ij} bezeichnet, dann sind die Varianzen der geschätzten Regressionskoeffizienten $\hat{\beta}_j$ in der Form

$$(2.1.40) \quad \text{Var}(\hat{\beta}_j) = \sigma^2 \cdot x^{jj}$$

darstellbar.

2.1.3.5 Effizienz

Ein erwartungstreuer Schätzer, der im Vergleich zu allen alternativen unverzerrten Schätzern die kleinste Varianz hat, heißt effizient. Man bezeichnet eine solche Schätzfunktion als besten unverzerrten Schätzer. Beim Gauß-Markow-Theorem wird jedoch eine beste unverzerrte Schätzung innerhalb der Klasse der linearen Schätzfunktionen untersucht. Innerhalb dieser Klasse erweist sich der OLS-Schätzer als effizient.

Um dies zu zeigen, betrachten wir eine beliebige lineare Schätzfunktion $\tilde{\boldsymbol{\beta}}$, wie sie in (2.1.34) definiert ist:

$$\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y} + \mathbf{a}.$$

Durch Einsetzen des Regressionsmodells (2.1.5) für y erhält man

$$(2.1.41) \quad \tilde{\boldsymbol{\beta}} = \mathbf{C}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) + \mathbf{a} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\mathbf{u} + \mathbf{a} ,$$

so dass sich für den Erwartungswert von $\tilde{\boldsymbol{\beta}}$ mit der Annahme (2.1.6)

$$E(\tilde{\boldsymbol{\beta}}) = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}E(\mathbf{u}) + \mathbf{a} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{a}$$

ergibt. Damit $\hat{\boldsymbol{\beta}}$ erwartungstreu ist, d.h.

$$E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

gilt, müssen offenbar die Restriktionen

$$(2.1.42) \quad \mathbf{C}\mathbf{X} = \mathbf{I} \quad \text{und} \quad \mathbf{a} = \mathbf{0}$$

mit der $k \times k$ -Einheitsmatrix \mathbf{I} und dem $k \times 1$ -Vektor $\mathbf{0}$ gelten.

Die $k \times k$ -Kovarianzmatrix für $\tilde{\boldsymbol{\beta}}$ lässt sich dann aus der Beziehung (2.1.32)

$$\text{Cov}(\tilde{\boldsymbol{\beta}}) = E\left[\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \cdot \left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right]$$

bestimmen. Unter Berücksichtigung von (2.1.42) geht (2.1.41) in

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{C}\mathbf{u}$$

über, so dass die Abweichung zwischen $\tilde{\boldsymbol{\beta}}$ und $\boldsymbol{\beta}$ durch

$$(2.1.43) \quad \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{C}\mathbf{u}$$

gegeben ist. Unter Verwendung von (2.1.43) erhält man für die Kovarianzmatrix von $\tilde{\boldsymbol{\beta}}$ die Form

$$\text{Cov}(\tilde{\boldsymbol{\beta}}) = E[(\mathbf{C}\mathbf{u})(\mathbf{C}\mathbf{u})'] = E(\mathbf{C}\mathbf{u}\mathbf{u}'\mathbf{C}) = \mathbf{C} E(\mathbf{u}\mathbf{u}')\mathbf{C} ,$$

woraus sich mit der Annahme (2.1.7)

$$(2.1.44) \quad \text{Cov}(\tilde{\boldsymbol{\beta}}) = \sigma^2 \mathbf{C}\mathbf{C}'$$

ergibt. Der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ hat die BLUE-Eigenschaft, d.h. er ist der beste lineare unverzerrte Schätzer für den unbekannt Parametervektor $\boldsymbol{\beta}$, wenn gezeigt werden kann, dass die Hauptdiagonalelemente von (2.1.39) stets kleiner oder gleich den Hauptdiagonalelementen von (2.1.44) sind. Denn dann können die Varianzen der Regressionskoeffizienten $\tilde{\beta}_j$ eines alternativen linearen unverzerrten Schätzers $\tilde{\boldsymbol{\beta}}$ die entsprechenden Varianzen nach der OLS-Methode nicht unterschreiten:

$$\text{Var}(\hat{\beta}_j) \leq \text{Var}(\tilde{\beta}_j) , \quad j=1, 2, \dots, k.$$

Aus formalen Gründen definieren wir hierzu eine $k \times n$ -Matrix \mathbf{D} :

$$\mathbf{D} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'.$$

Die $k \times n$ -Matrix \mathbf{C} ist dann in der Form

$$\mathbf{C} = \mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$$

darstellbar, womit (2.1.44) in die Darstellung

$$\begin{aligned} \text{Cov}(\tilde{\boldsymbol{\beta}}) &= \sigma^2 [\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'] [\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'] \\ &= \sigma^2 [\mathbf{D}\mathbf{D}' + \mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{D}' + (\mathbf{X}'\mathbf{X})^{-1}] \end{aligned}$$

übergeht. Wegen $\mathbf{C}\mathbf{X} = \mathbf{X}'\mathbf{C}' = \mathbf{I}$ sind die beiden mittleren Terme in der eckigen Klammer Nullmatrizen,

$$\mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{C}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$$

und

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{D}' = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{C}' - (\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0},$$

so dass man für die Kovarianzmatrix von $\tilde{\boldsymbol{\beta}}$ schließlich

$$\text{Cov}(\tilde{\boldsymbol{\beta}}) = \sigma^2 [\mathbf{D}\mathbf{D}' + (\mathbf{X}'\mathbf{X})^{-1}]$$

erhält. Da alle Hauptdiagonalelemente d_{jj} der Produktmatrix $\mathbf{D}\mathbf{D}'$ Quadratsummen sind, muss offenbar $d_{jj} \geq 0$ gelten. Damit ist aber auch gesichert, dass

$$\text{Var}(\hat{\beta}_j) = \sigma^2 \cdot x_{jj} \leq \text{Var}(\tilde{\beta}_j) = \sigma^2 \cdot (d_{jj} + x_{jj})$$

gilt, womit die BLUE-Eigenschaft des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ gezeigt ist. Die geschätzten Regressionskoeffizienten $\hat{\beta}_j$ haben von allen linearen erwartungstreuen Schätzfunktionen für β_j die kleinste Varianz. In diesem Sinne ist der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ ein bester Schätzer für den Vektor $\boldsymbol{\beta}$ der unbekanntenen Regressionskoeffizienten β_j des ökonomischen Modells (2.1.5).

2.1.3.6 Konsistenz

Wenn ein Schätzer mit zunehmendem Stichprobenumfang einen Parametervektor immer genauer treffen soll, dann ist insbesondere sein Grenzverhalten für $n \rightarrow \infty$ von Interesse, also eine Eigenschaft bei großen Stichproben. Bei der Konsistenz geht es um die Frage, ob ein Schätzer bei einem über alle Grenzen wachsenden Stichprobenumfang schließlich mit dem Parametervektor $\boldsymbol{\beta}$ zusammenfällt. Diese Eigenschaft ist dann ge-

sichert, wenn gezeigt werden kann, dass ein Schätzer asymptotisch unverzerrt ist und seine Streuung asymptotisch verschwindet.

Was den OLS-Schätzer $\hat{\beta}$ angeht, so haben wir gezeigt, dass er erwartungstreu ist. Da diese Eigenschaft für einen beliebigen Stichprobenumfang gilt, ist sie natürlich erst recht für $n \rightarrow \infty$ gegeben:

$$(2.1.45) \quad \lim_{n \rightarrow \infty} E(\hat{\beta}) = \beta .$$

Somit ist die erste Konsistenzbedingung erfüllt.

Die zweite Konsistenzeigenschaft bezieht sich auf die Streuung des Schätzers $\hat{\beta}$, die durch die Kovarianzmatrix $\text{Cov}(\hat{\beta})$ erfasst wird. $\hat{\beta}$ fällt für $n \rightarrow \infty$ genau dann mit dem unbekanntem Parametervektor β zusammen, wenn seine Kovarianzmatrix gegen eine $k \times k$ -Nullmatrix $\mathbf{0}$ strebt. Denn in diesem Fall verschwindet seine Variabilität, so dass β präzise geschätzt wird.

Für den Konsistenzbeweis ist eine zusätzliche Annahme erforderlich. Sie bezieht sich genauer auf die Produktmatrix $\mathbf{X}'\mathbf{X}$ nach Multiplikation mit dem Skalar $1/n$, deren Verhalten für $n \rightarrow \infty$ untersucht wird. Mit

$$(2.1.46) \quad \mathbf{Q} = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}'\mathbf{X} \quad \text{und} \quad \mathbf{Q} \text{ regulär}$$

wird unterstellt, dass die Matrix der empirischen Momente der exogenen Variablen

$$\frac{1}{n} \mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & \frac{1}{n} \sum X_{2t} & \cdots & \frac{1}{n} \sum X_{kt} \\ \frac{1}{n} \sum X_{2t} & \frac{1}{n} \sum X_{2t}^2 & \cdots & \frac{1}{n} \sum X_{2t} \cdot X_{kt} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} \sum X_{kt} & \frac{1}{n} \sum X_{kt} \cdot X_{2t} & \cdots & \frac{1}{n} \sum X_{kt}^2 \end{bmatrix}$$

an ihrer Grenze gegen eine reguläre und damit invertierbare $k \times k$ -Matrix \mathbf{Q} konvergiert. Dann lässt sich die asymptotische Kovarianzmatrix

$$\lim_{n \rightarrow \infty} \text{Cov}(\hat{\beta}) = \lim_{n \rightarrow \infty} \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

als Produkt zweier Grenzwerte schreiben:

$$\lim_{n \rightarrow \infty} \text{Cov}(\hat{\boldsymbol{\beta}}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} \cdot \lim_{n \rightarrow \infty} \mathbf{Q}^{-1},$$

weil die Inverse

$$\mathbf{Q}^{-1} = \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1}$$

für $n \rightarrow \infty$ mit (2.1.46) existiert. Da die Varianz σ^2 der Störvariablen eine Konstante ist, geht der Grenzwert des Ausdrucks σ^2/n gegen null, so dass die Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$ asymptotisch in eine Nullmatrix übergeht:

$$\lim_{n \rightarrow \infty} \text{Cov}(\hat{\boldsymbol{\beta}}) = 0 \cdot \lim_{n \rightarrow \infty} \mathbf{Q}^{-1} = \mathbf{0}.$$

Der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ wird damit mit wachsendem Stichprobenumfang ein immer besserer Schätzer für den unbekannt Parametervektor $\boldsymbol{\beta}$, bis er schließlich für einen über alle Grenzen wachsenden Stichprobenumfang mit ihm zusammenfällt. $\hat{\boldsymbol{\beta}}$ ist somit ein konsistenter Schätzer für $\boldsymbol{\beta}$.

2.1.4 Bestimmtheitsmaß und multipler Korrelationskoeffizient

Einen visuellen Eindruck über die Güte der Anpassung des Regressionsmodells können verschiedene graphische Darstellungen vermitteln. Aufgrund von Residuendiagrammen, bei denen die Residuen \hat{u}_t gegen die Zeit t oder gegen die Regressionswerte \hat{y}_t geplottet werden, lassen sich konkrete Aufschlüsse über Defekte in der Anpassung der Regressionshyperebene bei bestimmten Beobachtungen erkennen. Gleichwohl muss eine Beurteilung der Güte der Anpassung anhand graphischer Darstellungen subjektiv orientiert bleiben, so dass die Vergleichbarkeit erschwert wird. Aus Residuendiagrammen lassen sich hauptsächlich Probleme der Anpassung bei einzelnen Beobachtungen ausfindig machen. Die Beurteilung der globalen Güte der Anpassung des Regressionsmodells kann durch die Verwendung einer Maßzahl objektiviert werden. Eine derartige Maßzahl ist das **Bestimmtheitsmaß (Determinationskoeffizient)**.

Das Bestimmtheitsmaß bietet eine Entscheidungshilfe darüber, ob die exogenen Variablen insgesamt die endogene Variable erklären können. Wenn dies gegeben ist, kann man mit dem ökonomischen Modell erst einmal arbeiten, da es einen sinnvollen Beitrag zur Erklärung der endogenen Variablen liefert. Die Bedeutung des Bestimmtheitsmaßes

liegt genau darin, dass es den "Erklärungsgehalt" der exogenen Variablen in einem ökonometrischen Modell quantifiziert. Seine Interpretation ist anschaulich und wird dadurch erleichtert, dass es sich bei dieser Kennzahl um eine normierte Größe handelt.

Aufgrund des Verschiebungssatzes lässt sich die Varianz s_y^2 der zu erklärenden Variablen y in der Form

$$(2.1.47) \quad s_y^2 = \frac{1}{n} \sum y_t^2 - \bar{y}^2 = \frac{1}{n} \mathbf{y}'\mathbf{y} - \bar{y}^2$$

schreiben, so dass man mit $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$ die äquivalente Form

$$s_y^2 = \frac{1}{n} (\hat{\mathbf{y}} + \hat{\mathbf{u}})'(\hat{\mathbf{y}} + \hat{\mathbf{u}}) - \bar{y}^2 = \frac{1}{n} \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\frac{1}{n} \hat{\mathbf{y}}'\hat{\mathbf{u}} + \frac{1}{n} \hat{\mathbf{u}}'\hat{\mathbf{u}} - \bar{y}^2$$

erhält. Aus den Eigenschaften der OLS-Residuen in (2.1.31) folgt aber

$$(2.1.48) \quad \hat{\mathbf{y}}'\hat{\mathbf{u}} = (\mathbf{X}\hat{\boldsymbol{\beta}})'\hat{\mathbf{u}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\hat{\mathbf{u}} = 0 ,$$

was bedeutet, dass mit den Regressoren x_1, x_2, \dots, x_m die Regressionswerte \hat{y}_t mit den Residuen \hat{u}_t unkorreliert sind. Unter Berücksichtigung der Beziehungen $\bar{y} = \bar{\hat{y}}$ und $\hat{\mathbf{u}} = 0$ ergibt sich daher die Streuungszersetzung

$$(2.1.49) \quad s_y^2 = s_{\hat{y}}^2 + s_{\hat{u}}^2$$

mit
$$s_{\hat{y}}^2 = \frac{1}{n} \hat{\mathbf{y}}'\hat{\mathbf{y}} - \bar{y}^2$$

und
$$s_{\hat{u}}^2 = \frac{1}{n} \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

Nach der Streuungszersetzung (2.1.49) kann die Varianz der endogenen Variablen y in zwei Komponenten aufgespalten werden: die Varianz $s_{\hat{y}}^2$ der Regressionswerte und die Residualvarianz $s_{\hat{u}}^2$. Während die Varianz der Regressionswerte die durch das ökonometrische Modell "erklärte" Streuung der endogenen Variablen y wiedergibt, erfasst die Residualvarianz die "nicht-erklärte" Streuung. Je größer die Streuung der Regressionswerte ist, umso mehr wird prinzipiell der Gesamtzusammenhang durch das ökonometrische Modell erklärt.

Das **Bestimmtheitsmaß (Determinationskoeffizient)** basiert genau auf dieser Interpretation der Streuungszersetzung. Es misst die Güte der Anpassung durch das Verhält-

nis aus der durch die Regression "erklärten" Varianz zur Gesamtvarianz der endogenen Variablen:

$$(2.1.50) \quad R^2 = \frac{s_{\hat{y}}^2}{s_y^2}.$$

Wegen $0 \leq s_{\hat{y}}^2 \leq s_y^2$ gilt

$$0 \leq R^2 \leq 1,$$

so dass sich der Determinationskoeffizient als Anteilswert interpretieren lässt. So besagt z.B. ein Bestimmtheitsmaß von 0,9, dass 90 % der Streuung der endogenen Variablen auf die Streuung der exogenen Variablen zurückgeführt werden kann, d.h. durch das ökonometrische Modell "erklärt" wird. 10 % der Varianz von y werden in diesem Fall nicht durch die Regression erklärt, was dem Anteil der Residualvarianz an der Gesamtstreuung der endogenen Variablen entspricht. Anhand der Streuungszerlegung (2.1.49) lässt sich verifizieren, dass der Determinationskoeffizient unter Verwendung der Residualvarianz $s_{\hat{u}}^2$ gleichwertig aus

$$(2.1.51) \quad R^2 = 1 - \frac{s_{\hat{u}}^2}{s_y^2}$$

bestimmt werden kann. Eine Berechnungsformel für das Bestimmtheitsmaß ist unter Berücksichtigung der Definitionen der Varianzen von y und \hat{y} nach Multiplikation mit n durch

$$(2.1.52) \quad R^2 = \frac{\hat{y}'\hat{y} - n \cdot \bar{y}^2}{y'y - n \cdot \bar{y}^2} = \frac{n\hat{y}'\hat{y} - (\sum y_t)^2}{ny'y - (\sum y_t)^2}$$

gegeben. Ohne explizite Verwendung des Vektors \hat{y} der Regressionswerte lässt sich das Bestimmtheitsmaß mit dem OLS-Schätzer $\hat{\beta}$ alternativ aus

$$(2.1.53) \quad R^2 = \frac{\hat{\beta}'X'y - n \cdot \bar{y}^2}{y'y - n \cdot \bar{y}^2} = \frac{n\hat{\beta}'X'y - (\sum y_t)^2}{ny'y - (\sum y_t)^2}$$

bestimmen¹⁸.

¹⁸ Diese Berechnungsformel ergibt sich, indem man in dem Skalarprodukt $\hat{y}'\hat{y}$ die Beziehungen $\hat{y}' = (X\hat{\beta})'$ und $\hat{y} = y - \hat{u}$ verwendet:

$$\hat{y}'\hat{y} = (X\hat{\beta})'(y - \hat{u}) = \hat{\beta}'X'y - \hat{\beta}'X'\hat{u} = \hat{\beta}'X'y \text{ wegen (2.1.48).}$$

Beispiel 2.1.3: Gesucht ist der Anteil der Varianz des Privaten Verbrauchs, der auf die Streuung des verfügbaren Einkommens zurückgeführt werden kann. Um hierüber eine Aussage machen zu können, wird das Bestimmtheitsmaß für die makroökonomische Konsumfunktion für den Stützbereich von 1974 bis 1992 unter Verwendung der Berechnungsformel (2.1.53) ermittelt. Es ist

$$\hat{\beta}' = [37,324603 \quad 0,8642164],$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum y_t \\ \sum x_t y_t \end{bmatrix} = \begin{bmatrix} 19866,93 \\ 23512473,04 \end{bmatrix}$$

und

$$\mathbf{y}'\mathbf{y} = \sum y_t^2 = 21070671,2 \quad ,$$

woraus man

$$\hat{\mathbf{y}}'\hat{\mathbf{y}} = \hat{\beta}'\mathbf{X}'\mathbf{y} = 21061390,1$$

erhält. Mit $n=19$ und $\bar{y} = 1045,6$ ergibt sich damit ein Bestimmtheitsmaß in Höhe von

$$R^2 = \frac{19 \cdot 21061390,1 - 19866,93^2}{19 \cdot 21070671,2 - 19866,93^2} = \frac{5471504,28}{5647845,18} = 0,969 \quad ,$$

was bedeutet, dass 96,9 % der Streuung des Privaten Konsums durch die Regression "erklärt" wird. ♦

Bei der einfachen Regression entspricht das Bestimmtheitsmaß dem Quadrat des Korrelationskoeffizienten r_{yx} zwischen der endogenen Variablen y und der exogenen Variablen x :

$$(2.1.54) \quad R^2 = r_{yx}^2 \quad .$$

Diese Beziehung lässt sich zeigen, indem man die Varianz $s_{\hat{y}}^2$ der Regressionswerte in Abhängigkeit von der Varianz s_x^2 des Regressors x darstellt:¹⁹

$$s_{\hat{y}}^2 = \hat{\beta}_2^2 \cdot s_x^2 \quad .$$

¹⁹ Dahinter steht die Eigenschaft von Varianzen bei Lineartransformationen. Zum Beweis siehe Eckey, Kosfeld und Dreger (2002), S. 178.

Unter Berücksichtigung der Darstellung von $\hat{\beta}_2$ in der Form (2.1.26) als Verhältnis der Kovarianz s_{yx} zwischen y und x und der Varianz s_x^2 von x erhält man für den Determinationskoeffizienten den Ausdruck

$$R^2 = \frac{\hat{\beta}_2^2 \cdot s_x^2}{s_y^2} = \frac{\left(\frac{s_{xy}}{s_x^2}\right)^2 \cdot s_x^2}{s_y^2} = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2},$$

was genau dem Quadrat des Korrelationskoeffizienten

$$r_{yx} = \frac{s_{xy}}{s_x s_y}$$

entspricht. Anders ausgedrückt gibt die Wurzel des Bestimmtheitsmaßes bei der einfachen Regression den Korrelationskoeffizienten zwischen dem Regressanden y und dem Regressor x bis auf das Vorzeichen wieder:

$$(2.1.55) \quad |r_{yx}| = \sqrt{R^2}.$$

Damit ist eine Verallgemeinerung bei der multiplen Regression angezeigt. Die Wurzel des Bestimmtheitsmaßes lässt sich auch hier als Korrelationskoeffizient interpretieren. Sie gibt genauer die Korrelation zwischen dem Regressanden y und den Regressionswerten \hat{y} an. Da \hat{y} jedoch aus einer Linearkombination der Regressoren x_1, x_2, \dots, x_k hervorgeht, spricht man auch von einem **multiplen Korrelationskoeffizienten**. Das Bestimmtheitsmaß ist dann das Quadrat des multiplen Korrelationskoeffizienten $r_{yx_1, 2, \dots, k}$ zwischen der endogenen Variablen y und den exogenen Variablen x_1, x_2, \dots, x_k :

$$(2.1.56) \quad R^2 = r_{y\hat{y}}^2 = r_{yx_1, 2, \dots, k}^2.$$

Das Vorzeichen des multiplen Korrelationskoeffizienten bleibt hier erst recht unbestimmt, da positive und negative Einflüsse der Regressoren gleichzeitig auftreten können.

Obwohl der Determinationskoeffizient anschaulich interpretierbar ist, enthält er bestimmte Nachteile, die seinen Wert als Maß für die Güte der Anpassung einschränken. Ein Nachteil resultiert daraus, dass die Freiheitsgrade der Quadratsummen unberücksichtigt bleiben, was dazu führt, dass das Bestimmtheitsmaß bei Aufnahme eines zusätzlichen Regressors niemals sinken kann. Im Prinzip könnte man den Determinationskoeffizienten somit durch Aufnahme weiterer Regressoren, die nicht notwendig von

ökonomischer Relevanz sein müssen, vergrößern, ohne tatsächlich eine bessere empirische Fundierung ökonomischer Hypothesen erreicht zu haben. Aus diesem Grund ist streng genommen ein Vergleich der Güte der Anpassung alternativer ökonometrischer Modelle nur bei gleicher Anzahl von Regressoren zulässig. Auf jeden Fall ist bei unterschiedlicher Anzahl von Regressoren dieser Sachverhalt in eine Bewertung einzubeziehen.

Man kann nun den Determinationskoeffizienten durch Berücksichtigung der Freiheitsgrade bereinigen, wozu wir von der Definition (2.1.51) ausgehen. Da die Residualvarianz $s_{\hat{u}}^2$ nun $n-k$ und die Varianz s_y^2 der endogenen Variablen $n-1$ Freiheitsgrade besitzen, ist das korrigierte Bestimmtheitsmaß durch

$$(2.1.57) \quad \bar{R}^2 = 1 - \frac{\frac{1}{n-k} \hat{u}'\hat{u}}{\frac{1}{n-1} (y'y - n\bar{y}^2)}$$

gegeben. In Abhängigkeit von R^2 erhält man das **korrigierte Bestimmtheitsmaß** in der Form

$$(2.1.58) \quad \bar{R}^2 = 1 - \frac{n-1}{n-k} (1 - R^2).$$

Bei großem Stichprobenumfang im Vergleich zur Anzahl der Regressoren geht der Faktor $(n-1)/(n-k)$ gegen eins, so dass \bar{R}^2 gut mit R^2 übereinstimmt. Allgemein ist \bar{R}^2 jedoch stets kleiner als R^2 . Das bereinigte Bestimmtheitsmaß braucht bei einer Aufnahme zusätzlicher Regressoren nicht notwendig zu steigen. Vor allem wenn auf einer frühen Stufe ein Regressor nur einen geringfügigen Beitrag zur Erklärung der endogenen Variablen leistet, kann \bar{R}^2 im Extremfall sogar negativ werden. Abgesehen davon bleibt die Interpretation dieses Maßes als Anteil der durch die Regression erklärte Varianz der abhängigen Variablen im Wesentlichen erhalten. Vorteilhaft ist, dass alternative ökonometrische Modelle mit unterschiedlicher Anzahl von Regressoren besser miteinander verglichen werden können.

Bei Anwendungen nimmt das Bestimmtheitsmaß oft eine dominierende Rolle bei der Modellbewertung ein, die allerdings äußerst kritisch zu beurteilen ist. Sicher ist ein hohes Bestimmtheitsmaß bei einem ökonometrischen Modell vorteilhaft und wünschenswert, da es die Stringenz eines ökonomischen Zusammenhangs unterstreicht. Es kommt jedoch dabei entscheidend darauf an, ob die einbezogenen Regressoren auch separat

einen bedeutsamen Einfluss auf die endogene Variable besitzen, was mit Signifikanztests der Regressionskoeffizienten beurteilt werden kann. Außerdem kann eine ökonomische Hypothese durchaus bei einem nicht so hohen Bestimmtheitsmaß eine empirische Bestätigung finden, wenn sich z.B. alle einbezogenen Regressoren als signifikant erweisen. Das Bestimmtheitsmaß ist im Allgemeinen schon bei recht niedrigen Werten in seiner Interpretation als Zufallsvariable gegen Null gesichert, was bedeutet, dass eine Erklärung durch die exogenen Variablen gegeben ist. Wie der Korrelationskoeffizient darf das Bestimmtheitsmaß keinesfalls ohne weiteres kausal interpretiert werden.

Ein besonderes Problem ergibt sich bei ökonomischen Zeitreihen, die zum überwiegenden Teil trendbehaftet sind. Ein hohes Bestimmtheitsmaß ergibt sich dann oft schon allein aufgrund eines Trends, der einen Gleichlauf der endogenen Variablen mit einer oder mehrerer exogenen Variablen erzeugt. Gelegentlich wird daher versucht, durch Bildung von Differenzen oder Wachstumsraten den Trend zu eliminieren. Jedoch bleibt fraglich, ob man damit tatsächlich die relevanten Variablen für ein ökonometrisches Modell erhält, weil der eigentlich interessierende Niveauezusammenhang zwischen den ökonomischen Größen bei einer solchen Strategie verloren geht.²⁰

Aufgaben

- (2.1.1) Welche Annahmen liegen dem klassischen multiplen Regressionsmodell zugrunde?
- (2.1.2) Wir haben hier die erklärenden Variablen als exogen betrachtet, so dass stets vorgegebene, fixe Werte dieser Größen vorliegen. Zeigen Sie auf, welche Modifikation der Modellannahmen im Fall stochastischer x -Variablen vorgenommen werden müsste!
- (2.1.3) Leiten Sie die Kleinst-Quadrate-Schätzer $\hat{\beta}_1$ und $\hat{\beta}_2$ für die Regressionskoeffizienten β_1 und β_2 ohne Verwendung des Matrixkalküls her!
- (2.1.4) In einer Energienachfragefunktion soll die Nachfrage nach Erdgas (GASV) in Abhängigkeit vom verfügbaren Einkommen (VEINKR) erklärt werden

²⁰ Ökonometrische Verfahren, die speziell auf diese Problematik abstellen, heißen Kointegrationstechniken. Sie werden im Abschnitt 2.8 erörtert.

($\hat{=}$ Energienachfragemodell I). Ermitteln Sie die OLS-Schätzer für β_1 und β_2 einer linearen Regression von GASV auf VEINKR (Daten s. Anhang B, Tab. B1)!

- (2.1.5)** Die Nachfrage nach Erdgas soll durch die auf den Preisindex des Privaten Verbrauchs bezogenen Preise für Erdgas (GASPR) und Fernwärme (FERNWPR) durch eine lineare Nachfragefunktion erklärt werden ($\hat{=}$ Energienachfragemodell II). Bestimmen Sie die OLS-Schätzer der Nachfragefunktion nach Erdgas (Daten s. Anhang B, Tab. B1)!

Hinweis: Die Inverse der Produktmatrix lautet:

$$(\mathbf{X}\mathbf{X})^{-1} = \begin{bmatrix} 18,48237 & 5,069422 & -22,42939 \\ 5,069422 & 7,154167 & -11,31260 \\ -22,42939 & -11,31260 & 31,89869 \end{bmatrix}$$

- (2.1.6)** Was versteht man unter einem BLUE-Schätzer?
- (2.1.7)** Zeigen Sie, dass der OLS-Schätzer $\hat{\beta}_2$ einer einfachen Regression ein erwartungstreuer Schätzer für β_2 ist!
- (2.1.8)** Bestimmen Sie die Varianz des OLS-Schätzers $\hat{\beta}_2$ einer einfachen Regression und weisen Sie seine Effizienzeigenschaft innerhalb der Klasse der linearen Schätzfunktionen nach!
- (2.1.9)** Erläutern Sie verbal die Konsistenzeigenschaft! Welchen Stellenwert würden Sie dieser Schätzeigenschaft in der angewandten Ökonometrie in Abhängigkeit von der Datenart zubilligen?
- (2.1.10)** Zeigen Sie die Gültigkeit der Streuungszerlegung (2.1.49) für das einfache Regressionsmodell ohne Verwendung des Matrixkalküls auf!
- (2.1.11)** Berechnen Sie den Determinationskoeffizienten für das ökonometrisch geschätzte Energiemodell I [s. Aufgabe (2.1.4)] und interpretieren Sie ihn!
- (2.1.12)** Erläutern Sie die beiden unterschiedlichen Interpretationsmöglichkeiten des Determinationskoeffizienten!
- (2.1.13)** Welche Rolle könnte der korrigierte Determinationskoeffizient bei der Modellauswahl einnehmen?

2.2 Maximum-Likelihood-Methode und Inferenzstatistik

2.2.1 Die Maximum-Likelihood-Methode

Die Schätzung des Parametervektors $\boldsymbol{\beta}$ im multiplen Regressionsmodell erfolgte mit der gewöhnlichen Methode der kleinsten Quadrate (OLS-Methode), ohne dass eine Annahme über den Verteilungstyp der Störvariablen u_t erforderlich gewesen ist. Inferenzstatistische Konzepte, die über eine Punktschätzung von $\boldsymbol{\beta}$ hinausgehen, gehen jedoch in der Regel von einer Normalverteilung der Störvariablen aus, die zumindest approximativ gegeben sein sollte. So werden insbesondere Konfidenzintervalle und Signifikanztests für die Regressionskoeffizienten auf der Grundlage einer Normalverteilung der Störvariablen konstruiert. Zum Zwecke einer über die Punktschätzung hinausgehenden Inferenzstatistik im multiplen Regressionsmodell setzen wir aus diesem Grund neben den Annahmen (2.1.6) und (2.1.7),

$$E(\mathbf{u}) = \mathbf{0} \text{ und } E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I},$$

über die Störvariable zusätzlich eine Normalverteilung voraus:

$$(2.2.1) \quad \mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}).$$

Die Annahme (2.2.1) besagt, dass der Vektor \mathbf{u} der Störvariablen normalverteilt ist mit dem Erwartungswertvektor $\mathbf{0}$ und der Kovarianzmatrix $\sigma^2\mathbf{I}$.

Die Dichtefunktion f der Störvariablen u_t hat somit in jeder Periode t die Form

$$(2.2.2) \quad f_t(u_t; 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2} u_t^2\right). \quad ^{21}$$

Da die Kovarianzmatrix $\text{Cov}(\mathbf{u})$ eine Diagonalmatrix ist, sind die Störgrößen u_t unkorreliert, was bei einer Normalverteilung zugleich eine stochastische Unabhängigkeit impliziert. Ihre gemeinsame Dichtefunktion f lässt sich daher als Produkt der marginalen Dichtefunktionen f_t darstellen:

²¹ Der Ausdruck $\exp(z)$ gibt die e-Funktion e^z wieder, wobei e die Eulersche Zahl ist ($e=2,718 \dots$).

$$(2.2.3) \quad f(u_1, u_2, \dots, u_n; \mathbf{0}, \sigma^2 \mathbf{I}) = \prod_{t=1}^n f_t(u_t; 0, \sigma^2).$$

Unter Verwendung von (2.2.2) erhält man nach geeigneter Zusammenfassung der Terme für die gemeinsame Dichtefunktion

$$\begin{aligned} f(\mathbf{u}; \mathbf{0}, \sigma^2 \mathbf{I}) &= \frac{1}{(2\pi)^{1/2} \sigma} \exp\left(-\frac{u_1^2}{2\sigma^2}\right) \cdot \frac{1}{(2\pi)^{1/2} \sigma} \exp\left(-\frac{u_2^2}{2\sigma^2}\right) \cdot \dots \cdot \frac{1}{(2\pi)^{1/2} \sigma} \exp\left(-\frac{u_n^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} (u_1^2 + u_2^2 + \dots + u_n^2)\right) = \frac{1}{(2\sigma)^{n/2} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum u_t^2\right) \end{aligned}$$

oder wegen $\sum u_t^2 = \mathbf{u}'\mathbf{u}$

$$(2.2.4) \quad f(\mathbf{u}; \mathbf{0}, \sigma^2 \mathbf{I}) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \mathbf{u}'\mathbf{u}\right).$$

Die Maximum-Likelihood-Methode (ML-Methode) geht bei der Schätzung der unbekannt Parameter im multiplen Regressionsmodell von der gemeinsamen Dichtefunktion der Störvariablen u_t aus. Im Gegensatz zur OLS-Methode setzt die ML-Methode somit die Kenntnis eines Verteilungstyps voraus. Hier handelt es sich um die Normalverteilung, die sich für die Störvariable u_t durch den Zentralen Grenzwertsatz begründen lässt. Danach ist die Summe von unabhängigen und identisch verteilten Zufallsvariablen bei großen Stichproben unter sehr allgemeinen Bedingungen approximativ normalverteilt. Wenn das multiple Regressionsmodell aber korrekt spezifiziert ist, umfassen die exogenen Variablen x_j alle wesentlichen Einflussgrößen der endogenen Variablen y . Die Störvariable u enthält dann nur noch eine Vielzahl von unsystematischen Einflüssen, deren Effekte sich aufgrund der Annahme (2.1.6) gegenseitig ausgleichen. Die Vorstellung unabhängiger Störgrößen u_t steht bereits hinter der Annahme (2.1.7). Allerdings sollte die Normalverteilungsannahme im Zweifel überprüft werden, wozu verschiedene Testverfahren verwendet werden können. Wenn jedoch eine Normalverteilung der Störvariablen u_t zu inferenzstatistischen Zwecken ohnehin vorausgesetzt werden muss, dann bietet es sich auch an, zur Schätzung der Parameter des multiplen Regressionsmodells eine Methode in Betracht zu ziehen, die von diesem Verteilungstyp explizit Gebrauch macht. Der Vorteil liegt zum einen darin, dass für die Maximum-Likelihood-Methode Güteeigenschaften bekannt sind, die unabhängig von der Problem-

stellung Gültigkeit besitzen. Zum anderen lässt sich mit der ML-Methode zusätzlich zum Parametervektor $\boldsymbol{\beta}$ die Varianz σ^2 der Störvariablen u schätzen, was mit der OLS-Methode nicht möglich ist.

Hinzu kommt, dass das Schätzprinzip der Maximum-Likelihood-Methode trotz seiner wahrscheinlichkeitstheoretischen Fundierung überaus einleuchtend und anschaulich ist. Nach der Maximum-Likelihood-Methode ergeben sich nämlich unter allen in Betracht kommenden Werten der Parameter diejenigen als Schätzer, unter denen die gegebene Stichprobe die größtmögliche Wahrscheinlichkeit (bzw. Wahrscheinlichkeitsdichte) besitzt. Dieses Schätzprinzip wird realisiert, indem man von einer Likelihood-Funktion L ausgeht, bei der die Parameter als variabel und die Variablenwerte (Merkmalswerte) als gegeben angesehen werden, was der Situation nach Ziehen einer Stichprobe entspricht. Für vorgegebene Parameterwerte entspricht dagegen die Likelihood-Funktion der Wahrscheinlichkeitsfunktion (Dichtefunktion), bei der die Parameter als konstant und die Merkmalswerte als veränderlich betrachtet werden.

Im Folgenden geht es darum, bei gegebenen Werten der exogenen und endogenen Variablen die Schätzer für $\boldsymbol{\beta}$ und σ^2 zu finden, die die größte Wahrscheinlichkeitsdichte haben. Mit

$$\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

lässt sich die Likelihood-Funktion L in der Form

$$(2.2.5) \quad L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$

schreiben, die nach dem ML-Prinzip zu maximieren ist, da die Maximum-Likelihood-Schätzer $\hat{\boldsymbol{\beta}}_{ML}$ und $\hat{\sigma}_{ML}^2$ die Bedingung

$$(2.2.6) \quad L(\hat{\boldsymbol{\beta}}_{ML}, \hat{\sigma}_{ML}^2) = \max_{\boldsymbol{\beta}, \sigma^2} L(\boldsymbol{\beta}, \sigma^2)$$

erfüllen müssen. Da die Extremwerte einer Funktion von einer monotonen Transformation nicht beeinflusst werden, können wir die ML-Schätzer für $\boldsymbol{\beta}$ und σ^2 gleichermaßen aus der logarithmierten Likelihood-Funktion

$$(2.2.7) \quad \ln L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

bestimmen, was rechentechnisch einfacher ist. Da β allein im letzten Term auftritt, der ein negatives Vorzeichen besitzt, ist eine Maximierung der logarithmierten Likelihood-Funktion in Bezug auf β offenbar gleichwertig mit einer Minimierung des Ausdrucks

$$Q = \mathbf{u}'\mathbf{u} = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) .$$

Es handelt sich dabei um eine Minimierung der Summe der quadrierten Modellabweichungen u_i^2 , was genau bei der OLS-Methode erfolgt ist. Der ML-Schätzer $\hat{\beta}_{ML}$ stimmt daher mit dem OLS-Schätzer $\hat{\beta}$ für den Parametervektor β überein:

$$(2.2.8) \quad \hat{\beta}_{ML} = \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Das Gauß-Markow-Theorem ist damit natürlich auch für den ML-Schätzer $\hat{\beta}_{ML}$ gültig. Da für die Störvariable u jedoch eine Normalverteilung vorausgesetzt worden ist, entfällt die Restriktion auf die Klasse der linearen Schätzer für β . $\hat{\beta}_{ML}$ hat unter allen erwartungstreuen Schätzern die kleinste Varianz, so dass er der beste erwartungstreue Schätzer für β , d.h. effizient, ist ²². Die Konsistenzeigenschaft trifft natürlich auch auf $\hat{\beta}_{ML}$ zu.

Um einen ML-Schätzer für die Varianz σ^2 der Störvariablen zu erhalten, wird die logarithmierte Likelihood-Funktion (2.2.7) partiell nach σ^2 differenziert und gleich null gesetzt:

$$\frac{\partial \ln L(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \cdot (\mathbf{y} - \mathbf{X}\beta)' \cdot (\mathbf{y} - \mathbf{X}\beta) = 0 .$$

Aus dieser notwendigen Bedingung für ein Maximum von $\ln L$ bezüglich σ^2 erhält man durch Umformung

$$n = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) ,$$

so dass sich unter Verwendung des ML-Schätzers $\hat{\beta}_{ML}$ für den unbekannt Parametervektor β

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta}_{ML})' (\mathbf{y} - \mathbf{X}\hat{\beta}_{ML})$$

²² Siehe hierzu Fomby, Hill und Johnson (1984), S. 32ff.

ergibt. Wegen $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\text{ML}}$ sind die Komponenten von $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{ML}}$ mit den OLS-Residuen \hat{u}_t identisch,

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{ML}},$$

so dass der ML-Schätzer $\hat{\sigma}_{\text{ML}}^2$ für die Störvarianz

$$(2.2.9) \quad \hat{\sigma}_{\text{ML}}^2 = \frac{1}{n} \hat{\mathbf{u}}' \hat{\mathbf{u}} = \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2$$

lautet. Danach ergibt sich der Maximum-Likelihood-Schätzer für σ^2 als ungewogenes arithmetisches Mittel der quadrierten OLS-Residuen (= ML-Residuen) \hat{u}_t .

Wir werden allerdings den ML-Schätzer $\hat{\sigma}_{\text{ML}}^2$ nicht unmodifiziert für inferenzstatistische Zwecke im multiplen Regressionsmodell verwenden. Denn $\hat{\sigma}_{\text{ML}}^2$ besitzt zwar asymptotisch wünschenswerte Eigenschaften wie asymptotische Erwartungstreue, Konsistenz und asymptotische Effizienz, die bei großen Stichproben von Relevanz sind²³, jedoch ist er bei kleinen Beobachtungsumfängen nicht erwartungstreu. Die Modifikation wird also darauf hinauslaufen, einen erwartungstreuen Schätzer für die Störvarianz σ^2 zu erhalten.

2.2.2 Erwartungstreue Schätzung der Störvarianz

Eine erwartungstreue Schätzung der Störvarianz ist nicht allein bei der Beurteilung der Modellanpassung von Relevanz. Sie spielt ebenfalls bei der Konstruktion von Konfidenzintervallen und Hypothesentests für die Regressionskoeffizienten eine bedeutende Rolle. Ohne sie ist keine unverzerrte Schätzung der Varianz der Regressionskoeffizienten $\hat{\beta}_j$ gegeben.

Wie sich zeigen wird, muss bei einer erwartungstreuen Schätzung von σ^2 die Anzahl der Freiheitsgrade des Schätzers in Betracht gezogen werden. Ein Freiheitsgrad geht durch die Restriktion verloren, dass die Residuen \hat{u}_t im Mittel gleich null sein müssen. Außerdem müssen die Residuen \hat{u}_t mit jedem der $k-1$ Regressoren unkorreliert sein.

²³ Maximum-Likelihood-Schätzer besitzen unter allgemeinen Regularitätsbedingungen stets diese asymptotisch wünschenswerten Eigenschaften. Siehe hierzu z.B. Fisz (1976), S. 563ff.

Dies bedeutet, dass von den n Residuen \hat{u}_t nur noch $n-k$ "frei wählbar" sind, da sie insgesamt k Restriktionen genügen müssen, die sich aus dem System (2.1.24) ergeben. Bei der Schätzung von σ^2 bestehen somit nur $n-k$ Freiheitsgrade. Man gelangt dann zu einem erwartungstreuen Schätzer $\hat{\sigma}^2$ für die Störvarianz, indem man den Faktor $\frac{1}{n}$ bei dem ML-Schätzer durch den Faktor $\frac{1}{n-k}$ ersetzt:

$$(2.2.10) \quad \hat{\sigma}^2 = \frac{1}{n-k} \hat{\mathbf{u}}' \hat{\mathbf{u}} = \frac{1}{n-k} \sum_{t=1}^n \hat{u}_t^2 .$$

Bei dem erwartungstreuen Schätzer $\hat{\sigma}^2$ wird die Summe der Residualquadrate \hat{u}_t nicht durch die Anzahl n der Residuen, sondern durch die Anzahl $n-k$ der Freiheitsgrade der Schätzung geteilt.

Der Beweis der Erwartungstreue von $\hat{\sigma}^2$,

$$(2.2.11) \quad E(\hat{\sigma}^2) = \sigma^2 ,$$

kann unter Verwendung einiger Regeln der linearen Algebra geführt werden. Setzt man den OLS-Schätzer $\hat{\boldsymbol{\beta}}$ in die Beziehung $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ ein, so erhält man für den Residualvektor $\hat{\mathbf{u}}$ den Ausdruck

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} ,$$

der nach Ausklammern von \mathbf{y} in

$$\hat{\mathbf{u}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{y}$$

übergeht. Unter Verwendung der $n \times n$ -Matrix \mathbf{M} ,

$$(2.2.12) \quad \mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' ,$$

erhält man $\hat{\mathbf{u}}$ in der Form,

$$(2.2.13) \quad \hat{\mathbf{u}} = \mathbf{M}\mathbf{y} ,$$

woraus ersichtlich ist, dass die Residuen \hat{u}_t eine Linearkombination der endogenen Variablen y_1, y_2, \dots, y_n sind. Gleichmaßen lassen sich die Residuen \hat{u}_t jedoch auch als Linearkombination der Störvariablen u_1, u_2, \dots, u_n darstellen. Denn wenn man das multiple Regressionsmodell (2.1.5) für \mathbf{y} in (2.2.13) einsetzt, dann folgt wegen

$$\mathbf{M}\mathbf{X} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']\mathbf{X} = \mathbf{0}$$

$$(2.2.14) \quad \hat{\mathbf{u}} = \mathbf{M}\mathbf{u} .$$

Die Matrix \mathbf{M} hat die beiden Eigenschaften

$$(2.2.15) \quad \mathbf{M}' = \mathbf{M} \quad \text{und} \quad \mathbf{M}^2 = \mathbf{M} \cdot \mathbf{M} = \mathbf{M} .$$

Erstere Eigenschaft gibt die Symmetrie der Matrix \mathbf{M} wieder, die durch die Identität mit ihrer Transponierten gekennzeichnet ist:

$$\mathbf{M}' = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' = \mathbf{I}' - \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M} .$$

Letztere Eigenschaft besagt, dass das Produkt der Matrix \mathbf{M} mit sich selbst die Ausgangsmatrix \mathbf{M} unverändert lässt:

$$\begin{aligned} \mathbf{M}^2 &= \mathbf{M} \cdot \mathbf{M} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ (2.2.16) \quad &= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I} - 2 \cdot \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{M} \end{aligned}$$

Eine Matrix, die diese Eigenschaft besitzt, heißt idempotent. Damit können wir die Summe der Residuenquadrate in der Form

$$\sum_{t=1}^n \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} = (\mathbf{M}\mathbf{u})' \mathbf{M}\mathbf{u} = \mathbf{u}' \mathbf{M}' \mathbf{M}\mathbf{u} = \mathbf{u}' \mathbf{M}\mathbf{u}$$

schreiben. Nun machen wir von dem Konzept der Spur (trace) tr einer Matrix Gebrauch²⁴, das für die weiteren Betrachtungen benötigt wird. Zunächst einmal gilt

$$\hat{\mathbf{u}}' \hat{\mathbf{u}} = \mathbf{u}' \mathbf{M}\mathbf{u} = \text{tr}(\mathbf{u}' \mathbf{M}\mathbf{u}) ,$$

da die Spur eines Skalars gleich dem Skalar selbst ist. Die Spur eines Matrixprodukts bleibt unverändert, wenn die Reihenfolge der Produktbildung vertauscht wird, sofern nur die Multiplikation der Matrizen von ihrer Ordnung her möglich ist²⁵. Mithin ist die Spur der quadratischen Form $\mathbf{u}' \mathbf{M}\mathbf{u}$ mit der Spur der Matrix $\mathbf{M}\mathbf{u}\mathbf{u}'$ identisch:

$$\hat{\mathbf{u}}' \hat{\mathbf{u}} = \text{tr}(\mathbf{u}' \mathbf{M}\mathbf{u}) = \text{tr}(\mathbf{M}\mathbf{u}\mathbf{u}')$$

Wir bilden nun den Erwartungswert von $\hat{\mathbf{u}}' \hat{\mathbf{u}}$,

$$E(\hat{\mathbf{u}}' \hat{\mathbf{u}}) = E[\text{tr}(\mathbf{M}\mathbf{u}\mathbf{u}')] ,$$

der sich unter Berücksichtigung der Vertauschbarkeit der Erwartungswert- und Spurbildung in der Form

$$E(\hat{\mathbf{u}}' \hat{\mathbf{u}}) = \text{tr}[\mathbf{M} \cdot E(\mathbf{u}\mathbf{u}')] = \text{tr}[\mathbf{M} \cdot E(\mathbf{M}\mathbf{u}\mathbf{u}')] ,$$

darstellen lässt²⁶, so dass man mit der Annahme (2.1.7)

$$E(\hat{\mathbf{u}}' \hat{\mathbf{u}}) = \sigma^2 \text{tr}(\mathbf{M})$$

erhält. Da die Spur der Differenz zweier Matrizen separat berechnet werden kann, gilt weiter

$$\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}) - \text{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$$

mit der $n \times n$ -Einheitsmatrix \mathbf{I} . Die Spur von \mathbf{I} ist gleich n ,

$$\text{tr}(\mathbf{I}) = n ,$$

und die Spur von $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ist gleich k ,

$$\text{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = k ,$$

²⁴ Die Spur einer quadratischen Matrix ist allgemein als Summe ihrer Hauptdiagonalelemente definiert.

²⁵ Bei der Matrix \mathbf{A} , \mathbf{B} und \mathbf{C} gilt $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{ACB}) = \dots = \text{tr}(\mathbf{CBA})$, falls die entsprechenden Matrixprodukte definiert sind.

²⁶ Aus diesem Grund bezeichnet man den Erwartungswert E und die Spur tr auch als lineare Operatoren.

wenn man aufgrund der Rangannahme (2.1.16) über die Beobachtungsmatrix \mathbf{X} für das Matrixprodukt $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ denselben Rang voraussetzt. Damit erhält man

$$E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = \sigma^2(n - k) ,$$

womit gezeigt ist, dass der Schätzer (2.2.10) für σ^2 erwartungstreu ist:

$$E\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n - k}\right) = \sigma^2 .$$

Der Schätzer $\hat{\sigma}^2$ für die Varianz σ^2 ermöglicht es uns nun, Konfidenzintervalle für die geschätzten Regressionskoeffizienten zu berechnen und Signifikanztests durchzuführen.

Beispiel 2.2. 1: Die Schätzung der Varianz σ^2 der Störvariablen soll hier für die makroökonomische Konsumfunktion

$$\hat{C}_t = 37,317 + 0,864 \cdot Y_t$$

erfolgen. Hierzu wird eine Arbeitstabelle angelegt, in der die zur Berechnung des Schätzers

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{n - k} = \frac{\sum (C_t - \hat{C}_t)^2}{n - k}$$

benötigten Größen berechnet werden. \hat{C}_t ist der Regressionswert des privaten Verbrauchs in der Periode t, der z.B. für 1974

$$\hat{C}_1 = 37,317 + 0,864 \cdot 962,72 = 869,11$$

lautet. Auf diese Weise lassen sich die weiteren Regressionswerte berechnen, die in die Arbeitstabelle übernommen werden:

t	Y_t^v	C_t	\hat{C}_t	$C_t - \hat{C}_t$	$(C_t - \hat{C}_t)^2$
1	962,72	837,60	869,11	-31,51	992,88
2	989,56	863,82	892,30	-28,48	811,11
3	1017,68	897,32	916,59	-19,27	371,33
4	1046,38	937,80	941,39	-3,59	12,89
5	1093,39	971,48	982,01	-10,53	110,88
6	1130,98	1003,06	1014,48	-11,42	130,42
7	1118,86	1015,57	1004,01	11,56	133,63
8	1096,80	1007,92	984,95	22,97	527,62
9	1078,10	992,55	968,80	23,75	564,06
10	1086,06	1005,92	975,67	30,25	915,06
11	1100,16	1021,68	987,86	33,82	1143,79
12	1119,93	1036,53	1004,94	31,59	997,93
13	1205,44	1072,01	1078,82	-6,81	46,38
14	1239,32	1106,88	1108,09	-1,21	1,46
15	1299,72	1137,00	1160,28	-23,28	541,96
16	1323,60	1167,37	1180,91	-13,54	183,33
17	1420,53	1230,68	1264,65	-33,97	1153,96
18	1417,17	1274,63	1261,75	12,88	165,89
19	1421,38	1287,11	1265,39	21,72	471,76
Σ					9276,34

Die Summe der Abweichungsquadrate beträgt also 9276,34. Mit $n=19$ und $k=2$ erhält man hiermit den erwartungstreuen Schätzer

$$\hat{\sigma}^2 = \frac{9276,34}{19-2} = 545,667$$

für die Störvarianz σ^2 . Der Standardfehler der Schätzung lautet demzufolge

$$\hat{\sigma} = \sqrt{545,667} = 23,360 .$$

♦

2.2.3 Signifikanztest für die Regressionskoeffizienten

Wenn das multiple Regressionsmodell (2.1.5) mit der ökonomischen Theorie im Einklang steht, ist zu erwarten, dass die exogenen Variablen x_j einen Einfluss in einer bestimmten Richtung auf die endogene Variable y zeigen. So würden z.B. Zweifel an der Spezifikation einer Investitionsfunktion auftreten, in der der Zinssatz ein positives Vorzeichen aufweisen würde. Bei einer Konsumfunktion kommt dagegen aufgrund der ökonomischen Theorie allein ein positives Vorzeichen der marginalen Konsumneigung in Betracht. In einem ökonometrischen Modell müssen die Schätzer der Regressionskoeffizienten auf jeden Fall das theoretisch erwartete Vorzeichen aufweisen, damit eine ökonomische Hypothese eine empirische Bestätigung finden kann.

Allerdings handelt es sich dabei erst um eine unabdingbare Voraussetzung für eine empirisch gehaltvolle ökonomische Theorie. Es ist zusätzlich zu prüfen, ob die Einflussgrößen überhaupt eine Relevanz für die Erklärung der endogenen Variablen besitzen. Der Schätzer eines Regressionskoeffizienten einer erklärenden Variablen kann nämlich durchaus das theoretisch erwartete Vorzeichen besitzen, ohne dass ihr Einfluss auf die endogene Variable notwendig substantiell sein muss. Denn der geschätzte Regressionskoeffizient zeigt möglicherweise zufallsbedingt in der gegebenen Stichprobe ein negatives Vorzeichen, während er bei einer anderen Stichprobe ein positives Vorzeichen annehmen würde. Einen signifikanten Einfluss einer exogenen Variablen wird man unterstellen können, wenn mit großer Wahrscheinlichkeit sichergestellt werden kann, dass die Richtung des Einflusses, die durch ihren Regressionskoeffizienten angezeigt wird, nicht als Zufallsergebnis angesehen werden kann. Aus diesem Grund ist die Nullhypothese

$$(2.2.16) \quad H_0 : \beta_j = 0 ,$$

die von einem fehlenden Einfluss einer exogenen Variablen x_j auf die endogene Variable y ausgeht, statistisch zu testen. Erst wenn die Nullhypothese widerlegt werden kann, ist empirisch die Relevanz der exogenen Variablen zur Erklärung der endogenen Variablen erhärtet. Eine Prüfung der Nullhypothese in dieser Art und Weise heißt **Signifikanztest**.

Unter der Nullhypothese (2.2.16) ist der Schätzvektor $\hat{\beta}$ aufgrund der Normalverteilung der Störvariablen u normalverteilt mit dem Erwartungswertvektor $\mathbf{0}$ und der Kovarianzmatrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$(2.2.17) \quad \hat{\beta} \sim N(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Die Varianzen der Regressionskoeffizienten $\hat{\beta}_j$ sind daher durch

$$\text{Var}(\hat{\beta}_j) = \sigma^2 \cdot x^{jj}$$

gegeben, wobei die Größen x^{jj} die Hauptdiagonalelemente der Inversen $(\mathbf{X}'\mathbf{X})^{-1}$ bezeichnen. Man erhält Schätzer dieser Größen, indem man die unbekannte Störvarianz σ^2 durch den Schätzer $\hat{\sigma}^2$ aus (2.2.10) ersetzt:

$$(2.2.18) \quad \hat{\text{Var}}(\hat{\beta}_j) = \hat{\sigma}^2 \cdot x^{jj} = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n - k} \cdot x^{jj}.$$

Da $\hat{\sigma}^2$ erwartungstreu ist, werden die Varianzen $\text{Var}(\hat{\beta}_j)$ mit (2.2.18) selbst ebenfalls unverzerrt geschätzt. Im Spezialfall einer einfachen Regression erhält man als Schätzer für die Varianzen von $\hat{\beta}_1$ und $\hat{\beta}_2$

$$\hat{\text{Var}}(\hat{\beta}_1) = \hat{\sigma}^2 \frac{\sum x_t^2}{n \sum x_t^2 - (\sum x_t)^2}$$

und

$$\hat{\text{Var}}(\hat{\beta}_2) = \hat{\sigma}^2 \frac{n}{n \sum x_t^2 - (\sum x_t)^2}.$$

Die Prüfgröße des Signifikanztests

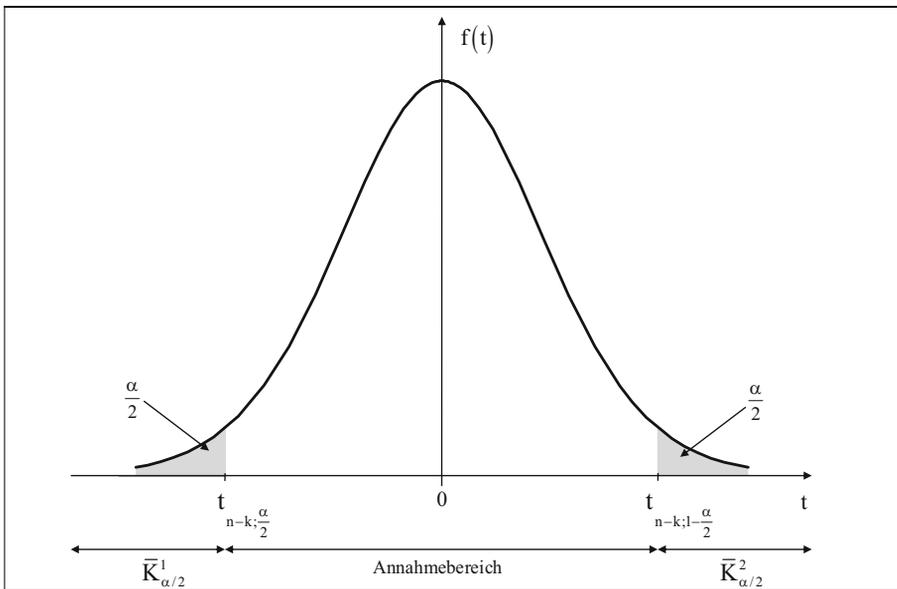
$$(2.2.19) \quad t = \frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{x^{jj}}}$$

ist t-verteilt mit $n-k$ Freiheitsgraden. Bei einem zweiseitigen Test wird die Nullhypothese (2.2.16) bei einem Signifikanzniveau von α verworfen, wenn der absolute Wert der Prüfgröße t das $(1-\alpha/2)$ -Quantil $t_{n-k;1-\alpha/2}$ der t-Verteilung mit $n-k$ Freiheitsgraden übersteigt:

$$|t| > t_{n-k;1-\alpha/2} \rightarrow H_0 \text{ ablehnen.}$$

Bei einem Signifikanzniveau von 5 % hätte ein Wert der Prüfgröße, der in den Ablehnungsbereich $\bar{K}_\alpha = \bar{K}_{\alpha/2}^1 \cup \bar{K}_{\alpha/2}^2$ fällt (Abbildung 2.2. 1), unter der Nullhypothese eine Wahrscheinlichkeit von 5 % oder weniger. Eine derart geringe Irrtumswahrscheinlichkeit wird bei der Ablehnung der Nullhypothese in Kauf genommen. Während geringe Abweichungen der Nullhypothese als zufallsbedingt interpretiert werden, kommt der Zufall bei größer werdenden Abweichungen als Ursache immer weniger in Betracht. Ab einer bestimmten Größe der Abweichung wird der Unterschied daher substantiell interpretiert.

Abbildung 2.2. 1: Annahme- und Ablehnungsbereiche bei dem zweiseitigen Signifikanztest



Sofern die ökonomische Theorie eine bestimmte Richtung des Einflusses vorgibt, lässt sich ein einseitiger Signifikanztest durchführen. Bei einem rechtsseitigen Test wird die Nullhypothese (2.2.16) abgelehnt, wenn die Ungleichung $t > t_{n-k; 1-\alpha}$ erfüllt ist,

$$t > t_{n-k; 1-\alpha} \rightarrow H_0 \text{ ablehnen,}$$

und bei einem linksseitigen Test ist eine Signifikanz im Falle von $t < t_{n-k; 1-\alpha}$ gegeben:

$$t < t_{1-\alpha; n-k} \rightarrow H_0 \text{ ablehnen.}$$

Ein zweiseitiger Signifikanztest gibt unter dieser Voraussetzung ein strengeres Kriterium für die Bewertung einer ökonomischen Beziehung vor als ein einseitiger Signifikanztest.

Beispiel 2.2. 1: Für die makroökonomische Konsumfunktion (2.1.28) haben wir für den Zeitraum von 1974 bis 1992 die OLS-Schätzer $\hat{\beta}_1=37,317$ und $\hat{\beta}_2=0,864$ ermittelt. Unter der Voraussetzung einer normalverteilten Störvariablen u sind dies zugleich die ML-Schätzer. Für einen Signifikanztest der Regressionskoeffizienten benötigen wir zusätzlich ihre Varianzen. Mit dem Schätzer $\hat{\sigma}^2=545,667$ für die Störvarianz sowie den Größen $n=19$, $\sum x_t = 22167,78$ und $\sum x_t^2 = 26249285,11$ erhält man

$$\begin{aligned}\widehat{\text{Var}}(\hat{\beta}_1) &= \hat{\sigma}^2 \cdot \frac{\sum x_t^2}{n \sum x_t^2 - (\sum x_t)^2} \\ &= 545,667 \cdot \frac{26249285,11}{19 \cdot 26249285,11 - 22167,78^2} \\ &= 545,667 \cdot 3,5583 = 1955,125\end{aligned}$$

und

$$\begin{aligned}\widehat{\text{Var}}(\hat{\beta}_2) &= \hat{\sigma}^2 \cdot \frac{n}{n \sum x_t^2 - (\sum x_t)^2} \\ &= 545,667 \cdot \frac{19}{19 \cdot 26249285,11 - 22167,78^2} \\ &= 545,667 \cdot 2,5935 \cdot 10^{-6} \\ &= 0,001415\end{aligned}$$

Damit ergeben sich für die geschätzten Regressionskoeffizienten die t-Werte

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}} = \frac{37,317}{\sqrt{1955,125}} = 0,844$$

und

$$t(\hat{\beta}_2) = \frac{\hat{\beta}_2}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_2)}} = \frac{0,864}{\sqrt{0,001415}} = 22,969 .$$

Bei einem zweiseitigen Test auf dem 5 %-Signifikanzniveau beträgt der kritische Wert $t_{17;0,975}$ der t-Verteilung mit 17 Freiheitsgraden 2,110. Wegen

$$|t(\hat{\beta}_1)| = 0,844 < t_{17;0,975} = 2,110$$

ist das absolute Glied $\hat{\beta}_1$ nicht signifikant von Null verschieden, wohingegen das Steigungsmaß $\hat{\beta}_2$ wegen

$$|t(\hat{\beta}_2)| = 22,969 > t_{17;0,975} = 2,110$$

signifikant ist. Der Regressionskoeffizient $\hat{\beta}_2$ ist sogar als hochsignifikant zu bezeichnen, da sein t-Wert auch bereits die kritische Grenze eines Signifikanztests auf dem 1 %-Niveau ($t_{17;0,995}=2,898$) übersteigt. ♦

Bei bestimmten Problemstellungen ist nicht nur von Interesse, ob eine Einflussgröße überhaupt eine substantielle Wirkung auf eine zu erklärende Variable hat. Vielmehr möchte man wissen, ob die empirischen Daten mit einer Hypothese

$$(2.2.20) \quad H_0: \beta_j = c, \quad c \neq 0,$$

vereinbar sind. Wenn man z.B. für eine bestimmte Region oder Bevölkerungsgruppe eine marginale Konsumquote von 0,8 ermittelt hat, so mag interessieren, ob diese Konsumneigung auch in einer anderen Region oder Bevölkerungsgruppe Gültigkeit besitzt. Zur Einschätzung des Effekts von Investitionsmaßnahmen kann es entscheidungsrelevant sein zu wissen, ob von einer bestimmten Höhe des Investitionsmultiplikators ausgegangen werden kann. Bei der Schätzung einer Geldnachfragefunktion stellt sich die Frage, ob die Hypothese einer Einkommenselastizität von eins Gültigkeit besitzt. Zur Prüfung der Nullhypothese kann der erörterte t-Test unter Verwendung der modifizierten Prüfgröße

$$(2.2.21) \quad t = \frac{\hat{\beta}_j - c}{\hat{\sigma} \sqrt{x_{jj}}}$$

in ansonsten unveränderter Form herangezogen werden.

Beispiel 2.2. 2: Bei der Geldnachfragefunktion (2.1.29) interessiert nicht allein, ob die erklärenden Variablen Einkommen und Zinssatz einen statistisch gesicherten Einfluss auf die Geldnachfrage ausüben. Zusätzlich ist für die Geldpolitik von Bedeutung,

inwieweit sich Größenvorteile oder -nachteile bei der Geldhaltung einstellen. Wenn von Größenvorteilen im Sinne einer rationelleren Geldhaltung ausgegangen werden kann, müsste die Einkommenselastizität der Geldnachfrage signifikant kleiner als Eins sein. Umgekehrt treten diseconomies of scale auf, wenn die Einkommenselastizität signifikant größer als Eins ist. Ein Signifikanztest der Einkommenselastizität auf Eins gibt somit Aufschluss darüber, ob die durch den Schätzer der entsprechenden Regressionskoeffizienten angezeigten Größenvorteile oder -nachteile tatsächlich substantiell interpretiert werden können.

Zunächst soll jedoch die Signifikanz der Einflussgrößen überhaupt geprüft werden. Hier benötigen wir die Varianzen der Regressionskoeffizienten $\hat{\beta}_1$, $\hat{\beta}_2$ und $\hat{\beta}_3$. Bei einer Residualquadratsumme von

$$\sum \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} = 0,014855$$

erhält man mit $n=20$ und $k=3$ den erwartungstreuen Schätzer $\hat{\sigma}^2$ der Störvarianz σ^2 aus

$$\hat{\sigma}^2 = \frac{1}{n-k} \hat{\mathbf{u}}' \hat{\mathbf{u}} = \frac{0,014855}{20-3} = 0,000874 .$$

Unter Verwendung der Diagonalelemente

$$x^{11}=218,194 , \quad x^{22}=3,770 \quad \text{und} \quad x^{33}=0,409$$

lassen sich die erwartungstreuen Schätzer für die Varianzen der geschätzten Regressionskoeffizienten berechnen:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \hat{\sigma}^2 \cdot x^{11} = 0,000874 \cdot 218,194 = 0,190702 ,$$

$$\widehat{\text{Var}}(\hat{\beta}_2) = \hat{\sigma}^2 \cdot x^{22} = 0,000874 \cdot 3,770 = 0,003295$$

und

$$\widehat{\text{Var}}(\hat{\beta}_3) = \hat{\sigma}^2 \cdot x^{33} = 0,000874 \cdot 0,409 = 0,000357 .$$

Damit erhält man die Prüfgrößen für die Signifikanztests der geschätzten Regressionskoeffizienten auf null:

$$\text{für } \hat{\beta}_1: \quad t = \frac{\hat{\beta}_1}{\sqrt{\hat{\text{Var}}(\hat{\beta}_1)}} = \frac{-4,806}{\sqrt{0,190702}} = -11,005 ,$$

$$\text{für } \hat{\beta}_2: \quad t = \frac{\hat{b}_2}{\sqrt{\hat{\text{Var}}(\hat{b}_2)}} = \frac{1,432}{\sqrt{0,003295}} = 24,947 ,$$

$$\text{für } \hat{\beta}_3: \quad t = \frac{\hat{\beta}_3}{\sqrt{\hat{\text{Var}}(\hat{\beta}_3)}} = \frac{-0,076}{\sqrt{0,000357}} = -4,022 .$$

Alle Regressionskoeffizienten sind auf einem Signifikanzniveau von 5% statistisch gesichert, da ihre t-Werte absolut größer sind als der kritische Wert $t_{17;0,975}=2,110$ eines zweiseitigen Tests.

Bei einem Test der Einkommenselastizität auf eins,

$$H_0 : \beta_2 = 1 ,$$

lautet die Prüfgröße

$$t = \frac{\hat{\beta}_2 - 1}{\sqrt{\hat{\text{Var}}(\hat{\beta}_2)}} = \frac{1,432 - 1}{\sqrt{0,003295}} = 7,526 .$$

Die Nullhypothese ist bei einem zweiseitigen Signifikanztest ($t_{0,975;17}=2,110$) und erst recht bei einem einseitigen (rechtsseitigen) Signifikanztest ($t_{0,95;17}=1,740$) zu verwerfen. Was die Geldmenge M_1 betrifft, so ist innerhalb des Beobachtungszeitraums von 1970 bis 1989 somit tatsächlich von "diseconomies of scale" der Geldhaltung auszugehen. Die Aufdeckung der Ursachen hierfür muss einer monetären Analyse vorbehalten bleiben. ♦

In den beiden vorgestellten Tests sind die Regressionskoeffizienten separat betrachtet worden. Ökonometrische Studien haben jedoch gleichermaßen gemeinsame Effekte von Einflussgrößen aufzudecken. So ist z.B. bei einer Cobb-Douglas-Produktionsfunktion

$$y_t = \beta_1 \cdot L_t^{\beta_2} \cdot K_t^{\beta_3} \cdot e^{u_t} ,$$

in der der Output y von den Produktionsfaktoren Arbeit L und Kapital K abhängig ist, bei bestimmten Anwendungen von Interesse, ob sie homogen vom Grade 1 ist. In diesem Fall müssen die aus der linearisierten Produktionsfunktion

$$\ln y_t = \ln \beta_1 + \beta_2 \cdot \ln L_t + \beta_3 \cdot \ln K_t + u_t$$

Schätzer $\hat{\beta}_2$ und $\hat{\beta}_3$ der Produktionselastizitäten mit der Hypothese

$$H_0: \beta_2 + \beta_3 = 1$$

im Einklang stehen. Es handelt sich hierbei somit um einen Test über eine Linearkombination der Regressionskoeffizienten:

$$(2.2.22) \quad c_1 \beta_1 + c_2 \beta_2 + \dots + c_k \beta_k = \mathbf{c}' \boldsymbol{\beta}$$

mit

$$\mathbf{c}' = (c_1 \ c_2 \ \dots \ c_k) .$$

Die Nullhypothese dieses Tests lautet allgemein

$$(2.2.23) \quad H_0: \mathbf{c}' \boldsymbol{\beta} = c ,$$

wobei c ein fester Wert ist. Bei dem Test der Cobb-Douglas-Produktionsfunktion auf lineare Homogenität lautet z.B. der Koeffizientenvektor \mathbf{c} der Linearkombination $\mathbf{c}' = (0 \ 1 \ 1)$, und die Konstante c nimmt den Wert 1 an. Die Prüfgröße

$$(2.2.24) \quad t = \frac{\mathbf{c}' \hat{\boldsymbol{\beta}} - c}{\hat{\sigma} \sqrt{\mathbf{c}' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{c}}}$$

des Tests ist t-verteilt mit $n-k$ Freiheitsgraden.

2.2.4 Konfidenzintervalle für die Regressionskoeffizienten

Die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode liefern zunächst einmal Punktschätzer für die Regressionskoeffizienten. Ob sie tatsächlich bedeutungsvoll sind, lässt sich mit einem Signifikanztest herausfinden. Wenn jedoch z.B. die Wirkung wirtschaftspolitischer Maßnahmen aufgrund eines ökonometrischen Modells eingeschätzt werden soll, ist es zudem wichtig, Kenntnisse über den Bereich zu erhalten, in dem der "wahre" Regressionskoeffizient mit einer großen Wahrscheinlichkeit liegen wird. Dieser Bereich, der durch eine Intervallschätzung ermittelt werden kann, heißt **Konfidenzintervall**.

Bei der Konstruktion eines Konfidenzintervalls für den unbekanntes Regressionskoeffizienten β_j gehen wir von der standardisierten Größe

$$(2.2.25) \quad t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_j)}} = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{x_{jj}}}$$

aus, die t-verteilt ist mit $n-k$ Freiheitsgraden. Der t-Wert kann dabei als standardisierter Regressionskoeffizient $\hat{\beta}_j$ interpretiert werden. Die Wahrscheinlichkeit, dass der standardisierte Regressionskoeffizient in einem symmetrischen Intervall um seinen Erwartungswert 0 liegt, ist durch

$$P(-t_{n-k;1-\alpha/2} \leq t \leq t_{n-k;1-\alpha/2}) = 1 - \alpha$$

gegeben. Setzt man hierin (2.2.25) ein, so erhält man

$$P\left(-t_{n-k;1-\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{x_{jj}}} \leq t_{n-k;1-\alpha/2}\right) = 1 - \alpha$$

und nach weiterer Umformung

$$(2.2.26) \quad P\left(\hat{\beta}_j - t_{n-k;1-\alpha/2} \cdot \hat{\sigma}\sqrt{x_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{n-k;1-\alpha/2} \cdot \hat{\sigma}\sqrt{x_{jj}}\right) = 1 - \alpha .$$

Gleichung (2.2.26) gibt das Konfidenzintervall für β_j zum Konfidenzniveau $1-\alpha$ an. Bei Normalverteilung der Störvariablen u liegt der unbekanntes Regressionskoeffizient β_j mit einer Wahrscheinlichkeit von $1-\alpha$ in dem Intervall

$$\hat{\beta}_j \pm t_{n-k;1-\alpha/2} \cdot \hat{\sigma} \sqrt{x^{jj}}.$$

Das Konfidenzintervall gibt aber aufgrund seiner Robustheit auch eine brauchbare Auskunft über die Lokalisation des Regressionskoeffizienten, wenn die Verteilung der Störvariablen u Abweichungen von der Normalverteilungsannahme aufweist.

Unter Verwendung der Schätzer für die Varianzen der Regressionskoeffizienten $\hat{\beta}_1$ und $\hat{\beta}_2$ erhält man für die einfache Regression die Konfidenzintervalle

$$\hat{\beta}_1 \pm t_{n-2;1-\alpha/2} \cdot \hat{\sigma} \cdot \sqrt{\frac{\sum x_t^2}{n \sum x_t^2 - (\sum x_t)^2}}$$

und

$$\hat{\beta}_2 \pm t_{n-2;1-\alpha/2} \cdot \hat{\sigma} \cdot \sqrt{\frac{n}{n \sum x_t^2 - (\sum x_t)^2}}$$

zum Konfidenzniveau $1-\alpha$. Die $(1-\alpha)$ -Quantile $t_{n-2;1-\alpha/2}$ beziehen sich hier auf eine t -Verteilung mit $n-2$ Freiheitsgraden.

Beispiel 2.2. 4: Als Punktschätzer für die marginale Konsumquote c_1 der makroökonomischen Konsumfunktion ist der Wert 0,864 ermittelt worden. Dieser Wert ist als Realisation einer Zufallsvariablen zu interpretieren, die sich aus den Stichprobendaten des Einkommens und Konsums im Zeitraum von 1974 bis 1992 ergibt. Für die Beurteilung wirtschaftspolitischer Maßnahmen ist jedoch gleichwohl von Interesse, in welchem Bereich die marginale Konsumquote mit einem hohen Vertrauensniveau liegen wird. Auf diese Weise kann z.B. versucht werden, wirtschaftliche Entwicklungstendenzen unter günstigen und ungünstigen Bedingungen einzuschätzen.

Die Punktschätzung für die marginale Konsumquote ist dann durch ein Konfidenzintervall zu ergänzen. Bei einem Konfidenzniveau von 95 % ist es durch

$$\hat{c}_1 \pm t_{n-2;0,975} \cdot \hat{\sigma} \cdot \sqrt{\frac{n}{n \sum x_t^2 - (\sum x_t)^2}}$$

gegeben. Bekannt sind die Größen, die zur Berechnung der Standardabweichung von c_1 benötigt werden:

$$\hat{\sigma} = 23,360, \quad n = 19, \quad \sum x_t^2 = 26249285,11, \quad \sum x_t = 22167,78.$$

Unter Verwendung des 0,975-Quantils $t_{17;0,975} = 2,110$ der t-Verteilung mit 17 Freiheitsgraden erhält man damit das Konfidenzintervall

$$\left[0,864 - 2,110 \cdot 23,360 \cdot \sqrt{\frac{19}{19 \cdot 26249285,11 - 22167,78^2}}; 0,864 + 2,110 \cdot 23,360 \cdot \sqrt{\frac{19}{19 \cdot 26249285,11 - 22167,78^2}} \right]$$

$$\left[0,864 - 2,110 \cdot 0,0376; \quad 0,864 + 2,110 \cdot 0,0376 \right] = \left[0,785; 0,943 \right]$$

für die marginale Konsumquote zum Konfidenzniveau von 95 %. ◆

Beispiel 2.2. 5: In der Geldpolitik werden zu verschiedenen Zwecken Bereiche für die Größenordnung monetärer Variablen verwendet. Hier soll eine Intervallschätzung der Einkommens- und Zinselastizität der Geldnachfrage vorgenommen werden. Die 95%-Konfidenzintervalle für die Einkommens- und Zinselastizität lauten

$$\hat{\beta}_2 \pm t_{0,975;n-3} \cdot \sqrt{\hat{\text{VAR}}(\hat{\beta}_2)}$$

und

$$\hat{\beta}_3 \pm t_{0,975;n-3} \cdot \sqrt{\hat{\text{VAR}}(\hat{\beta}_3)}.$$

Das 0,975-Quantil der t-Verteilung mit 17 Freiheitsgraden, die sich aus dem Stichprobenumfang von $n=20$ ergeben, beträgt 2,110. Mit den Schätzern

$$\hat{\beta}_2 = 1,432 \quad \text{und} \quad \hat{\beta}_3 = -0,076$$

und den Standardabweichungen

$$\sqrt{\hat{\text{VAR}}(\hat{\beta}_2)} = 0,057 \quad \text{und} \quad \sqrt{\hat{\text{VAR}}(\hat{\beta}_3)} = 0,019$$

erhält man für die Einkommenselastizität das Konfidenzintervall

$$\left[1,432 - 2,110 \cdot 0,057; \quad 1,432 + 2,110 \cdot 0,057 \right] = \left[1,312; \quad 1,501 \right]$$

und für die Zinselastizität das Konfidenzintervall

$$[-0,076 - 2,110 \cdot 0,019; -0,076 + 2,110 \cdot 0,019] = [-0,116; -0,036] .$$

Beide Konfidenzintervalle haben ein Vertrauensniveau von 95 %. ♦

Während bei Signifikanztests die Irrtumswahrscheinlichkeit α geeignet festzulegen ist, setzt die Bestimmung von Konfidenzintervallen eine geeignete Wahl des Konfidenzniveaus $1-\alpha$ voraus. Dabei ist zu beachten, dass eine größere Vertrauenswahrscheinlichkeit bei gegebenem Stichprobenumfang n stets zu Lasten der Genauigkeit geht, d.h. mit einem breiteren Konfidenzintervall verbunden ist. In der Mehrzahl der Anwendungen wird ein Konfidenzniveau von 0,95 gewählt, womit eine 95%ige Sicherheit gegeben ist, dass das Konfidenzintervall den unbekanntem Regressionskoeffizienten β_j überdeckt.

Wenn eine höhere Präzision erforderlich ist, müsste die Vertrauenswahrscheinlichkeit gesenkt werden. Eine Wahrscheinlichkeitsaussage über die Lokalisation von β_j könnte dann z.B. auf einem Konfidenzniveau von 0,90 erfolgen. Dagegen kann die Gefahr, dass das Konfidenzintervall den unbekanntem Parameter β_j nicht überdeckt, vermindert werden, indem das Konfidenzniveau z.B. auf 0,99 erhöht wird. In diesem Fall ist nur mit 1%iger Wahrscheinlichkeit damit zu rechnen, dass der Regressionskoeffizient β_j außerhalb des Konfidenzintervalls liegt.

Im Vergleich zu einem Signifikanztest der Hypothese $H_0: \beta_j = 0$ vermittelt das Konfidenzintervall mehr Informationen über einen Regressionskoeffizienten. Wenn die Irrtumswahrscheinlichkeit α mit dem Konfidenzniveau $1-\alpha$ konform ist, lässt sich das Ergebnis eines zweiseitigen Signifikanztests auch aus dem konkreten Konfidenzintervall erkennen. Ein Konfidenzintervall, das z.B. den Wert Null überdeckt, gibt an, dass der Regressionskoeffizient nicht signifikant, d.h. nicht gegen Null gesichert, ist. Sofern das Konfidenzintervall den Wert Null nicht überdeckt, führt der Signifikanztest zu einer Ablehnung der Nullhypothese, die $\beta_j = 0$ unterstellt. Im Unterschied zu Signifikanztests werden Konfidenzintervalle immer dort verwendet, wo konkrete Wahrscheinlichkeitsbereiche für die unbekanntem Regressionskoeffizienten gefragt sind.

2.2.5 Varianzanalyse und Signifikanz des Gesamtzusammenhangs

Der Determinationskoeffizient ist ein Maß für die Güte der Anpassung, das hauptsächlich deskriptiv interpretiert wird. Jedoch sind die hierin eingehenden Quadratsummen Zufallsvariablen, die eine varianzanalytische Prüfung zulassen. Bei einer Varianzanalyse (Analysis of Variance = ANOVA) eines ökonometrischen Modells wird die Summe der quadrierten Abweichungen der Werte der endogenen Variablen y von ihrem Mittel \bar{y} in die durch die Regression "erklärte" Quadratsumme und die Abweichungsquadratsumme der Residuen aufgeteilt. In einer ANOVA-Tabelle, die die Grundlage eines Signifikanztests über den Gesamtzusammenhang ist, werden zusätzlich die Freiheitsgrade und die mittleren quadratischen Abweichungen ausgewiesen.

Übersicht 2.2 1: ANOVA-Tabelle für eine Regressionsanalyse

Quelle der Streuung	Summe der Abweichungsquadrate	Freiheitsgrade	Mittlere Abweichungsquadratsumme
Regression (erklärt)	$\sum (\hat{y}_t - \bar{y})^2 = \hat{\mathbf{y}}' \hat{\mathbf{y}} - n\bar{y}^2$ $= \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n\bar{y}^2$	k-1	$(\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n\bar{y}^2) / (k-1)$
Residuen (unerklärt)	$\sum \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}}$	n-k	$\hat{\mathbf{u}}' \hat{\mathbf{u}} / (n-k)$
gesamt	$\sum (y_t - \bar{y})^2 = \mathbf{y}' \mathbf{y} - n\bar{y}^2$	n-1	

Wenn kein systematischer Zusammenhang zwischen der endogenen Variablen y und den Regressoren (ohne Scheinvariable) x_2, x_3, \dots, x_k besteht, müssen die "wahren" Regressionskoeffizienten für diese Variablen gleich null sein. Die Nullhypothese für einen Test des Gesamtzusammenhangs ist somit durch

$$(2.2.27) \quad H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$$

gegeben. Unter der Nullhypothese ist die Prüfgröße aus (2.1.46),

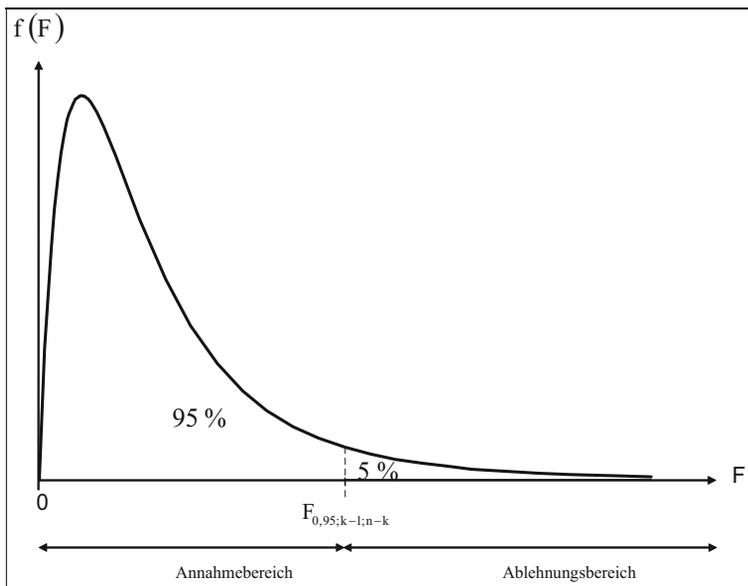
$$(2.2.28) \quad F = \frac{(\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n\bar{y}^2) / (k-1)}{\hat{\mathbf{u}}' \hat{\mathbf{u}} / (n-k)},$$

des Signifikanztests F-verteilt mit $k-1$ und $n-k$ Freiheitsgraden²⁷. Sie besteht aus dem Verhältnis der durch die Regression "erklärten" Streuung zur Residualstreuung, wobei die Streuungsmaße in Form der mittleren Abweichungsquadratsummen unter Berücksichtigung der Freiheitsgrade gebildet worden sind. Damit die Nullhypothese abgelehnt werden kann, d.h. der Gesamtzusammenhang als statistisch gesichert angesehen werden kann, muss die durch die Regression "erklärte" Streuung groß sein im Vergleich zur Residualstreuung. Genauer wird die Nullhypothese auf dem Signifikanzniveau α verworfen, wenn die Prüfgröße F das $(1-\alpha)$ -Quantil $F_{1-\alpha;k-1;n-k}$ der F-Verteilung mit $k-1$ und $n-k$ Freiheitsgraden übersteigt:

$$F > F_{1-\alpha;k-1;n-k} \rightarrow H_0 \text{ ablehnen.}$$

In diesem Fall können bei einer Irrtumswahrscheinlichkeit α nicht mehr alle Regressoren als irrelevant angesehen werden. Der F-Test, der in Abbildung 2.2. 2 für ein Signifikanzniveau von $\alpha=0,05$ wiedergegeben wird, ist demzufolge stets ein einseitiger Test.

Abbildung 2.2. 2: Annahme- und Ablehnungsbereich des F-Tests



²⁷ Bei Normalverteilung der Störgröße u sind die Quadratsummen $\hat{y}'\hat{y} - n\bar{y}^2$ und $\hat{u}'\hat{u}$ unabhängig χ^2 -verteilt mit $k-1$ bzw. $n-k$ Freiheitsgraden. Das Verhältnis zweier χ^2 -verteilter Größen ist F-verteilt mit der entsprechenden Anzahl von Freiheitsgraden. Vgl. Mood et al. (1974), S. 247.

Sofern der Determinationskoeffizient bekannt ist, lässt sich die Prüfgröße gleichwertig in der Form

$$(2.2.29) \quad F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}$$

berechnen. Gleichung (2.2.29) ergibt sich unmittelbar durch Erweitern von (2.2.28) um den Faktor $1 / (y'y - n\bar{y}^2)$, wie aus der Definition des Bestimmtheitsmaßes hervorgeht.

Beispiel 2.2. 6: Bei der ökonometrischen Analyse der gesamtwirtschaftlichen Konsumfunktion für die Bundesrepublik Deutschland im Stützzeitraum von 1974 bis 1992 ergibt sich folgende ANOVA-Tabelle:

Quelle der Streuung	Summe der Abweichungsquadrate	Freiheitsgrade	Mittlere Abweichungsquadratsumme
Regression (erklärt)	289082,3	1	289082,3
Residuen (unerklärt)	9276,3	17	545,7
Gesamt	293119,4	18	16284,4

Aus ihr lässt sich die Prüfgröße gemäß (2.2.28) berechnen:

$$F = \frac{289082,3}{545,7} = 529,7 .$$

Bei einem Signifikanztest auf dem 5 %-Signifikanzniveau ist dieser Wert mit dem 0,95-Quantil $F_{0,95;1;17}=4,45$ und auf dem 1 %-Signifikanzniveau mit dem 0,99-Quantil $F_{0,99;1;17}=8,40$ zu vergleichen. In beiden Fällen übersteigt der empirische F-Wert den kritischen Wert, so dass der durch die Konsumfunktion postulierte Zusammenhang statistisch gesichert ist. Alternativ hat man die Prüfgröße unter Verwendung des Determinationskoeffizienten $R^2=0,969$ bestimmen können:

$$F = \frac{0,969/1}{(1 - 0,969)/17} = 531,4 .$$

Die beiden Formeln (2.2.28) und (2.2.29) führen bis auf Rundungsungenauigkeiten zum selben Ergebnis. ♦

Man beachte, dass der F-Test Auskunft darüber gibt, ob der in einer Regressionsgleichung spezifizierte Zusammenhang zwischen den exogenen Variablen und der endogenen Variablen überhaupt statistisch gesichert ist. Der Erklärungsgehalt braucht bei einer Signifikanz des Tests keinesfalls besonders hoch zu sein. Die Forderung eines großen Wertes des Determinationskoeffizienten im deskriptiven Sinne stellt oft höhere Anforderungen an ein ökonometrisches Modell. Insbesondere bei neueren Entwicklungen in der ökonomischen Theorie, z.B. in der Finanzmarkttheorie und in der empirischen Kapitalmarktforschung, kommt jedoch dem Auffinden überhaupt statistisch gesicherter Beziehungen durchaus ein hoher Stellenwert zu.

Beispiel 2.2. 7: Bei einem Kapitalmarktmodell mit drei exogenen Variablen (einschl. der Scheinvariablen) beträgt der Determinationskoeffizient 0,303. Das Regressionsmodell ist aus 16 verbundenen Beobachtungen geschätzt worden. Es stellt sich nun die Frage, ob die exogenen Variablen insgesamt einen Einfluss auf die endogene Variable besitzen oder aber substantiell bedeutungslos sind.

Aufschluss hierüber kann ein F-Test auf Signifikanz des Determinationskoeffizienten vermitteln. Bei einem Signifikanzniveau von 5 % ist die Prüfgröße

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,303/(3-1)}{(1-0,303)/(16-3)} = 2,826$$

mit dem 0,95-Quantil $F_{0,95;3;15}$ der F-Verteilung mit 3 und 16 Freiheitsgraden zu vergleichen. Wegen

$$F = 2,826 < F_{0,95;3;15} = 3,287$$

kann die Nullhypothese (2.2.27) nicht verworfen werden, d.h. der Einfluss der exogenen Variablen auf die endogene Variable des Kapitalmarktmodells ist nicht statistisch gesichert. ♦

Aufgaben

- (2.2.1) Erläutern Sie die Ratio der Maximum-Likelihood-Methode!
- (2.2.2) Bestimmen Sie den Maximum-Likelihood-Schätzer für den Parameter λ der Zufallsvariable X , die die Wahrscheinlichkeitsfunktion

$$f(x | \lambda) = (\lambda^x / x!) \cdot e^{-\lambda}$$

(Poissonverteilung) besitzt!

- (2.2.3) Testen Sie die OLS-Schätzer $\hat{\beta}_1$ und $\hat{\beta}_2$ des Energiemodells I (s. Aufg. 2.1.4) auf Signifikanz ($\alpha=0,05$) und ermitteln Sie die zu einem Konfidenzniveau von 95% zugehörigen Konfidenzintervalle!
- (2.2.4) Wie groß ist der Standardfehler der Regression von GASV auf VEINKR (Energiemodell I, s. Aufg. 2.1.4)?
- (2.2.5) Geben Sie für das Energiemodell I (s. Aufg. 2.1.4) die beiden Residuen für das erste und letzte Jahr des Stützbereichs an! Welche Aussage über die Güte der Anpassung an den Rändern lässt sich aus ihnen gewinnen?
- (2.2.6) Zeigen Sie für das Energiemodell II (s. Aufg. 2.1.5) die Streuungszerlegung in Form einer ANOVA-Tabelle auf! Testen Sie die Nullhypothese $H_0 : \beta_2 = \beta_3 = 0$ und interpretieren Sie das Testergebnis ($\alpha=0,01$)!
- (2.2.7) Interpretieren Sie die im Aufgabe 2.1.5 für das Energiemodell II durchgeführte Modellschätzung unter Berücksichtigung der Informationen über die Varianzen $s^2(\hat{\beta}_2) = 6,071$ und $s^2(\hat{\beta}_3) = 27,082$ der Regressionskoeffizienten $\hat{\beta}_2$ und $\hat{\beta}_3$!

2.3 Multikollinearität

2.3.1 Begriff der Multikollinearität

Ökonomische Daten, die für die exogenen Variablen in einem Regressionsmodell verwendet werden, können nicht in einer für die ökonometrische Analyse adäquaten Form ausgewählt werden. Im Gegensatz zu den Naturwissenschaften, in denen der Experimentator die Werte der unabhängigen Variablen fixieren und ihren Einfluss auf die abhängige Variable studieren kann, müssen in den Wirtschaftswissenschaften in der Regel die unabhängigen und abhängigen Variablen beobachtet werden. Dadurch wird eine Messung des isolierten Einflusses einer Einflussgröße auf die zu erklärende Variable erschwert, da die exogenen Variablen häufig miteinander korreliert sind. In diesem Fall kann der Einfluss einer exogenen Variablen zumindest zum Teil durch einen anderen Regressor mit erfasst werden. Der Regressionskoeffizient dieser Variablen wird in einer multiplen Regression mit beiden Regressoren dann aber geringer sein als in einer einfachen Regression mit ersterer Einflussgröße. Tatsächlich wird der Gesamteinfluss der beiden korrelierten Regressoren bei einer Schätzung mit der gewöhnlichen Methode der kleinsten Quadrate korrekt erfasst. Die Zurechenbarkeit des Gesamteinflusses der beiden exogenen Variablen auf die endogene Variable wird jedoch verwischt.

Multikollinearität bedeutet nun eine gewisse Form der Abhängigkeit zwischen den exogenen Variablen. Eine exakte lineare Abhängigkeit zwischen den Regressoren würde dazu führen, dass der Rang der Beobachtungsmatrix \mathbf{X} kleiner als k wird, so dass die Annahme (2.1.16) verletzt ist. In diesem Extremfall würde der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ nicht mehr bestimmbar sein. Angenommen, das ökonometrische Modell lautet

$$(2.3.1) \quad y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

mit der Scheinvariablen $x_{1t}=1$ und den beiden Einflussgrößen x_2 und x_3 . Wenn nun die beiden Regressoren x_2 und x_3 exakt linear miteinander verknüpft sind,

$$(2.3.2) \quad x_{3t} = \gamma x_{2t} ,$$

dann sind die Koeffizienten β_1 , β_2 und β_3 nicht mehr eindeutig bestimmbar, da die Inverse der Produktionsmatrix $\mathbf{X}'\mathbf{X}$ nicht existiert. Da jedoch ein perfekter Zusammenhang zwischen x_2 und x_3 vorliegt, d.h. der Korrelationskoeffizient r_{23} gleich Eins ist,

reicht es aus, eine der beiden exogenen Variablen zur Erklärung von y heranzuziehen. Substituiert man (2.3.2) in (2.3.1), so erhält man die neue Regressionsgleichung

$$(2.3.3) \quad y_t = \beta_1 x_{1t} + (\beta_2 + \gamma\beta_3)x_{2t} + u_t ,$$

die nun ohne weiteres mit der OLS-Methode geschätzt werden kann. Dabei kann jedoch nur der Gesamteinfluss $\beta_2 + \gamma\beta_3$ der beiden Einflussgrößen x_2 und x_3 ermittelt werden. Er kann nicht in die Einzeleinflüsse β_2 und β_3 separiert werden, so dass man diese beiden Parameter als nicht identifizierbar bezeichnet. Schätztechnisch bedeutet dies, dass an Stelle einer Regressionsebene nur eine Regressionsgerade im dreidimensionalen Raum bestimmt werden kann. Wenn zwischen den beiden Regressoren x_2 und x_3 anstelle von (2.3.2) eine lineare Abhängigkeit der Form

$$(2.3.4) \quad x_{3t} = \delta + \gamma \cdot x_{2t}$$

bestehen würde, dann würde die schätzbare Regressionsgleichung

$$(2.3.5) \quad y_t = (\beta_1 + \delta\beta_3)x_{1t} + (\beta_2 + \gamma\beta_3)x_{2t} + u_t$$

lauten. In diesem Fall wäre auch der Achsenabschnitt β_1 nicht mehr identifizierbar.

Exakte lineare Abhängigkeiten zwischen den Regressoren können sich bei ökonometrischen Modellen z.B. aufgrund einer definitorischen Beziehung ergeben. Wenn man in einem Regressionsmodell z.B. als exogene Variablen das Gesamteinkommen Y , das Lohn- und Gehaltseinkommen L , die Gewinne G und das Transfereinkommen Tr verwendet, dann liegt aufgrund der Definitionsgleichung

$$Y_t = L_t + G_t + Tr_t$$

(exakte) Multikollinearität vor. Allerdings ist ein derartiges ökonometrisches Modell auch als fehlspezifiziert anzusehen, da das Einkommen praktisch doppelt verwendet wird, um die endogene Variable zu erklären. Ebenso kann eine fehlerhafte Definition von Dummy-Variablen z.B. zur Erfassung der Saisonkomponente Multikollinearität hervorrufen, wenn einige Spalten der Beobachtungsmatrix \mathbf{X} dadurch linear voneinander abhängig werden, was später ausführlicher diskutiert wird. In diesen Fällen ist die Produktmatrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar, so dass der OLS-Schätzer $\hat{\beta}$ nicht existiert. Diese Form der Multikollinearität ist jedoch fast immer durch eine sorgfältige Modellierung vermeidbar.

Denn exakte lineare Abhängigkeiten zwischen ökonomischen Variablen können ansonsten in der ökonometrischen Praxis vernachlässigt werden. Ökonomische Variablen sind

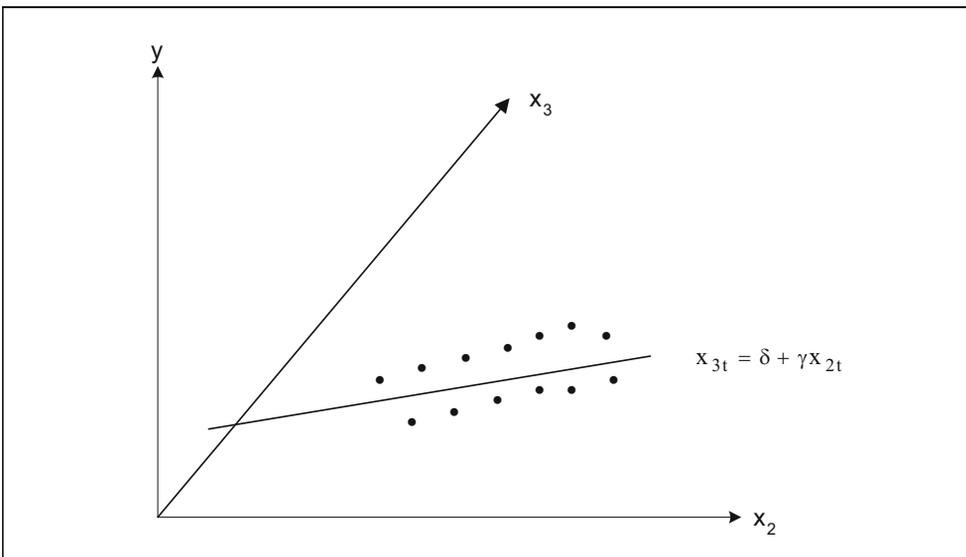
häufig durchaus hoch, jedoch nicht perfekt miteinander korreliert. Der OLS-Schätzer $\hat{\beta}$ ist dann zwar berechenbar, doch das Zurechenbarkeitsproblem verursacht Probleme bei der Interpretation einer ökonometrischen Analyse. Das Schätzproblem lässt sich graphisch für den Fall des Regressionsmodells (2.3.1) veranschaulichen, wenn man eine hohe Korrelation zwischen den beiden Regressoren x_2 und x_3 unterstellt. Anstatt der exakten linearen Abhängigkeit (2.3.4) liegt zwischen ihnen dann eine stochastische lineare Abhängigkeit der Form

$$(2.3.6) \quad x_{3t} = \delta + \gamma \cdot x_{2t} + v_t$$

vor, wobei v eine Störgröße bezeichnet. Die Projektionen der Beobachtungen (y_t, x_{2t}, x_{3t}) gruppieren sich in der x_2, x_3 -Ebene eng um die Gerade

$$(2.3.7) \quad x_{3t} = \delta + \gamma \cdot x_{2t} .$$

Abbildung 2.3. 1: Multikollinearität



Dadurch wird die Schätzung der Regressionsebene in dem von den Variablen y , x_2 und x_3 aufgespannten Raum unsicher. Denn mit zunehmender Nähe der Projektionen der Beobachtungswerte zur Geraden (2.3.7) geht die Regressionsebene, die sich der dreidimensionalen Punktwolke optimal anpasst, immer mehr in eine Regressionsgerade im dreidimensionalen Raum über. Wenn alle projizierten Beobachtungen auf der Geraden (2.3.7) liegen, ist die Lage der Regressionsebene vollkommen unbestimmt. In

diesem Fall der exakten Multikollinearität existiert statt einer Regressionsebene E_i nur noch die Regressionsgerade G . Je mehr die projizierten Beobachtungen dagegen in der x_2, x_3 -Ebene streuen, umso breiter ist die Basis zur Bestimmung einer Regressionsebene im dreidimensionalen Raum. Bei einer breiten Streuung der x_2 - und x_3 -Werte besteht aber eine Tendenz zur Unabhängigkeit beider Variablen.

Damit ist graphisch aufgezeigt, dass die Multikollinearität ein Problem ist, das bei linearen Abhängigkeiten zwischen den exogenen Variablen auftritt. Es handelt sich somit nicht um ein Problem bei der Spezifikation eines ökonometrischen Modells, sondern um ein Datenproblem. Wenn in zwei oder mehreren exogenen Variablen z.B. ein Trend wirksam ist, muss mit Multikollinearität gerechnet werden. Da viele ökonomische Variablen trendbehaftet sind, ist Multikollinearität bei einer multiplen Regression häufig vorzufinden. Um mit diesem Problem besser vertraut zu werden, befassen wir uns zunächst einmal mit den unmittelbaren Auswirkungen der Multikollinearität bei der Schätzung von ökonometrischen Modellen.

2.3.2 Auswirkungen der Multikollinearität

Bei Multikollinearität, die aus statistischen Abhängigkeiten zwischen den exogenen Variablen resultiert, lässt sich der OLS-Schätzer $\hat{\beta}$ weiterhin berechnen. Jedoch liegt die Determinante $|\mathbf{X}'\mathbf{X}|$, die zur Bestimmung der Inversen $(\mathbf{X}'\mathbf{X})^{-1}$ benötigt wird, dann nahe bei Null, so dass starke Rundungsfehler dadurch auftreten können, dass die Produktmatrix $\mathbf{X}'\mathbf{X}$ fast singular ist. Als Folge könnten sich die geschätzten Regressionskoeffizienten schon bei geringfügigen Veränderungen der Datenbasis beträchtlich verändern.

Dass die Regressionskoeffizienten $\hat{\beta}_j$ bei Multikollinearität unsicherer werden, lässt sich aus der Kovarianzmatrix des OLS-Schätzers $\hat{\beta}$ entnehmen, die entscheidend durch die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ determiniert ist. Wir wollen dies an Hand des Regressionsmodells (2.3.1) aufzeigen, das wir der Einfachheit halber für die Abweichungen der Beobachtungswerte von ihren Mittelwerten formulieren:

$$(2.3.8) \quad \tilde{y}_t = \beta_2 \tilde{x}_{2t} + \beta_3 \tilde{x}_{3t} + u_t$$

mit

$$\tilde{y}_t = y_t - \bar{y}, \quad \tilde{x}_{2t} = x_{2t} - \bar{x}_2 \quad \text{und} \quad \tilde{x}_{3t} = x_{3t} - \bar{x}_3 .$$

Vorteilhaft ist hier, dass das absolute Glied β_1 , das für die Multikollinearität ohne Bedeutung ist, entfällt, so dass sich mit der $n \times 2$ Beobachtungsmatrix

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{x}_{21} & \tilde{x}_{31} \\ \tilde{x}_{22} & \tilde{x}_{32} \\ \vdots & \vdots \\ \tilde{x}_{2n} & \tilde{x}_{3n} \end{bmatrix}$$

die 2×2 Produktmatrix

$$\tilde{\mathbf{X}}' \tilde{\mathbf{X}} = \begin{bmatrix} \sum \tilde{x}_{2t}^2 & \sum \tilde{x}_{2t} \tilde{x}_{3t} \\ \sum \tilde{x}_{2t} \tilde{x}_{3t} & \sum \tilde{x}_{3t}^2 \end{bmatrix}$$

ergibt, die leicht zu invertieren ist. Unter Verwendung der Determinante

$$|\tilde{\mathbf{X}}' \tilde{\mathbf{X}}| = \sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2$$

und der adjungierten Matrix

$$(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{\text{adj}} = \begin{bmatrix} \sum \tilde{x}_{3t}^2 & -\sum \tilde{x}_{2t} \tilde{x}_{3t} \\ -\sum \tilde{x}_{2t} \tilde{x}_{3t} & \sum \tilde{x}_{2t}^2 \end{bmatrix}$$

erhält man für den OLS-Schätzer

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}} \tilde{\mathbf{y}} \quad \text{mit } \tilde{\mathbf{y}} = (\tilde{y}_1 \tilde{y}_2 \dots \tilde{y}_n)$$

die Kovarianzmatrix

$$(2.3.9) \quad \text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} = \frac{\sigma^2}{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2} \begin{bmatrix} \sum \tilde{x}_{3t}^2 & -\sum \tilde{x}_{2t} \tilde{x}_{3t} \\ -\sum \tilde{x}_{2t} \tilde{x}_{3t} & \sum \tilde{x}_{2t}^2 \end{bmatrix}.$$

Die Varianzen von $\hat{\beta}_2$ und $\hat{\beta}_3$ lauten folglich

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2 \sum \tilde{x}_{3t}^2}{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2}$$

und

$$\text{Var}(\hat{\beta}_3) = \frac{\sigma^2 \sum \tilde{x}_{2t}^2}{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 - (\sum \tilde{x}_{2t} \tilde{x}_{3t})^2}.$$

Man erhält eine geeignete Interpretation der Multikollinearität, wenn man hierin den Korrelationskoeffizienten r_{23} zwischen x_2 und x_3 einführt, der durch

$$r_{23} = \frac{\sum \tilde{x}_{2t} \tilde{x}_{3t}}{\sqrt{\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2}}$$

definiert ist. Nach Quadrierung ergibt sich hieraus

$$(2.3.10) \quad \left(\sum \tilde{x}_{2t} \tilde{x}_{3t} \right)^2 = r_{23}^2 \left(\sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2 \right),$$

so dass die Varianzen von $\hat{\beta}_2$ und $\hat{\beta}_3$ in der Form

$$(2.3.11) \quad \text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{(1 - r_{23}^2) \sum \tilde{x}_{2t}^2}$$

und

$$(2.3.12) \quad \text{Var}(\hat{\beta}_3) = \frac{\sigma^2}{(1 - r_{23}^2) \sum \tilde{x}_{3t}^2}$$

darstellbar sind. Bei multikollinearen Regressoren x_2 und x_3 geht r_{23} tendenziell gegen Eins, was die Nenner von (2.3.11) und (2.3.12) klein werden lässt. Multikollinearität führt mithin zu hohen Varianzen der geschätzten Regressionskoeffizienten, deren Ausprägungen umso unsicherer werden. Eine Signifikanz der Regressionskoeffizienten lässt sich daher schwieriger ermitteln. Der Gesamtzusammenhang kann statistisch gesichert sein, obwohl die einzelnen Regressionskoeffizienten insignifikant sind.

Aus der Kovarianzmatrix (2.3.9) lässt sich ebenso der Zusammenhang zwischen den beiden Schätzern $\hat{\beta}_2$ und $\hat{\beta}_3$ auffindig machen. Unter Verwendung von (2.3.10) erhält man für die Kovarianz zwischen $\hat{\beta}_2$ und $\hat{\beta}_3$ den Ausdruck

$$\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-\sigma^2 \sum \tilde{x}_{2t} \tilde{x}_{3t}}{(1 - r_{23}^2) \sum \tilde{x}_{2t}^2 \sum \tilde{x}_{3t}^2},$$

der anzeigt, dass die Regressionskoeffizienten $\hat{\beta}_2$ und $\hat{\beta}_3$ bei Multikollinearität negativ korreliert sind. Dies ergibt sich daraus, dass das Kreuzprodukt $\sum \tilde{x}_{2t} \tilde{x}_{3t}$, das die Richtung des Zusammenhangs zwischen x_2 und x_3 angibt, bei Multikollinearität ein positives Vorzeichen haben muss. Genauer lässt sich zeigen, dass die Korrelation zwischen $\hat{\beta}_2$ und $\hat{\beta}_3$ bei umgekehrten Vorzeichen absolut so hoch ist wie die Korrelation zwischen x_2 und x_3 :

$$r_{\hat{\beta}_2, \hat{\beta}_3} = -r_{23} .$$

Das bedeutet, dass eine Überschätzung von $\hat{\beta}_2$ mit einer Unterschätzung von $\hat{\beta}_3$ einhergeht und umgekehrt. Hieran wird noch einmal die Schwierigkeit einer Separierung des Gesamteinflusses der beiden exogenen Variablen x_2 und x_3 verdeutlicht.

Was hier beispielhaft anhand zweier exogener Variablen aufgezeigt worden ist, gilt im Prinzip auch bei mehr als zwei Einflussgrößen. Damit kann das Problem der Multikollinearität allgemein gravierende Auswirkungen bei einer ökonomischen Analyse haben. Eine Einschätzung des Einflusses einer bestimmten exogenen Variablen wird mit zunehmender Ausprägtheit der Multikollinearität immer unsicherer. Dies ist insbesondere unbefriedigend, wenn der empirische Gehalt einer ökonomischen Hypothese oder die Wirkungen wirtschaftspolitischer Maßnahmen untersucht werden sollen. Weniger gravierend ist das Multikollinearitätsproblem dagegen bei einer Wirtschaftsprognose, wenn die Beziehungen zwischen den exogenen Größen im Prognosezeitraum weitgehend erhalten bleibt.

2.3.3 Aufdeckung von Multikollinearität

Zur Aufdeckung von Multikollinearität gibt es verschiedene Techniken, die von einer einfachen Korrelationsanalyse der exogenen Variablen bis hin zu einer Hauptkomponentenanalyse (principal components) reichen²⁸. Wir beschränken uns hier auf die Darstellung einiger einfacherer Techniken, da letzteres Verfahren Kenntnisse der multivariaten Statistik voraussetzt. Wie die Hauptkomponentenanalyse zur Aufdeckung der Anwesenheit und Form von Multikollinearität eingesetzt werden kann, wird in fortgeschrittenen ökonomischen Lehrbüchern aufgezeigt (s. z.B. Fomby, Hill und Johnson, 1988, S. 293-296).

Wenn ein Regressionsmodell zwei Regressoren (außer der Scheinvariablen) x_2 und x_3 enthält, reicht eine **einfache Korrelationsanalyse** aus. Multikollinearität wird dann durch einen hohen Korrelationskoeffizienten r_{23} angezeigt. Als Faustregel kann hierbei ein Korrelationskoeffizient von absolut 0,8 und größer gelten. Nun könnte man bei mehr

²⁸ Ein Verfahren zur Aufdeckung von Multikollinearität ist erstmals in Form einer "Büschelkartenanalyse" (confluence analysis) von Frisch (1934) entwickelt worden, das jedoch recht aufwendig und von subjektiven Urteilen abhängig ist. Vgl. hierzu Menges (1961), S. 146-154.

als zwei exogenen Variablen entsprechend alle paarweisen Korrelationskoeffizienten $r_{23}, r_{24}, \dots, r_{k-1,k}$ zur Beurteilung der Multikollinearität heranziehen. Mit dieser Methode lassen sich jedoch nur paarweise Abhängigkeiten aufdecken, nicht jedoch wechselseitige Beziehungen bei Beteiligung mehrerer Einflussgrößen. So könnte sich z.B. im Falle von drei exogenen Variablen durchaus Multikollinearität aus einer wechselseitigen Abhängigkeit aller drei Regressoren ergeben, obwohl die paarweisen Korrelationskoeffizienten absolut niedrig sind. Außerdem lässt sich allgemein durch eine einfache Korrelationsanalyse nicht die Auswirkung der Multikollinearität auf die Genauigkeit der Regressionskoeffizienten einschätzen.

Die **einfache Korrelationsanalyse** wird **erweitert**, wenn man bei der Beurteilung der Multikollinearität den multiplen Korrelationskoeffizienten $R_{yx_2, \dots, k}$ mit einbezieht. Nach Klein (1969, S. 101) wird die Multikollinearität zu einem Problem, wenn die Korrelation r_{ij} zwischen zwei Regressoren x_i und x_j , $i \neq j$, absolut den multiplen Korrelationskoeffizienten $R_{yx_2, \dots, k}$ übersteigt:

$$|r_{ij}| > R_{yx_2, \dots, k}, \quad i, j = 2, 3, \dots, k, \quad i \neq j .$$

Die Einschränkungen, die gegenüber der einfachen Korrelationsanalyse gemacht worden sind, werden dadurch jedoch nicht aufgehoben. Allein die Beurteilung der kritischen Grenze der Korrelationskoeffizienten, von der ab die Multikollinearität als wirksam angesehen wird, ist bei diesem Verfahren objektiviert worden.

Als Diagnosemaß zur Entdeckung von Multikollinearität wird gelegentlich die **Determinante der Korrelationsmatrix \mathbf{R}** der exogenen Variablen verwendet. Die Korrelationsmatrix \mathbf{R} ist regulär, wenn die Datenmatrix \mathbf{X} vollen Rang hat. Ansonsten wird \mathbf{R} singulär, so dass ihre Determinante gleich null ist. Wenn exogene Variablen nun stark korreliert sind, besteht zwischen zwei oder mehreren Spalten von \mathbf{X} näherungsweise eine lineare Beziehung, was zur Folge hat, dass die Korrelationsmatrix \mathbf{R} fast singulär wird. Bei Multikollinearität wird die Determinante von \mathbf{R} mithin einen Wert nahe bei null annehmen:

$$(2.3.13) \quad |\mathbf{R}| \approx 0 .$$

Die Beurteilung wird jedoch dadurch erschwert, dass es keinen objektiven kritischen Wert für $|\mathbf{R}|$ gibt, von dem ab die Multikollinearität als gravierend eingeschätzt werden kann. Darüber hinaus ist zu beachten, dass es sich hierbei um ein globales Maß handelt, das keinerlei Einsicht über die Art der Multikollinearität vermittelt. Insofern hat dieses

sicht darüber vermittelt, durch welche Variablen die Multikollinearität hervorgerufen wird. Dies kann bei Maßnahmen zur Beseitigung der Multikollinearität von Bedeutung sein. Farrar und Glauber (1967, S. 92ff.) beabsichtigen allgemein, durch diese Tests die Datenbasis einzugrenzen, durch die die Multikollinearität hervorgerufen wird. Jedoch ist zu beachten, dass es bei ökonomischen Variablen häufig zumindest einige signifikante F-Werte geben wird. Dann unterliegen die Hilfsregressionen jedoch selbst der Multikollinearität, wodurch es schwierig wird, die spezifischen Abhängigkeiten zwischen den exogenen Variablen herauszufiltern.

Bei einer deskriptiven Ausrichtung des Verfahrens der Hilfsregressionen kann man sich auf eine Betrachtung der Größenordnung der Bestimmtheitsmaße R_j^2 beschränken. Anstelle der Bestimmtheitsmaße zieht man bei einer Diagnose von Multikollinearität dann jedoch häufig die sog. Toleranzkoeffizienten tol_j heran, die durch

$$(2.3.15) \quad tol_j = 1 - R_j^2, \quad j=2,3,\dots,k$$

definiert sind. Niedrige Toleranzkoeffizienten indizieren eine Präsenz von Multikollinearität. Als Faustregel ist das Ausmaß der Multikollinearität als kritisch zu beurteilen, wenn die Toleranzkoeffizienten die Schranke von 0,05 unterschreiten. Danach ist man mithin bereit, im Falle von $tol_j \geq 0,05$ kollineare Beziehungen zwischen den Regressoren zu tolerieren.

Unter Verwendung der Toleranzkoeffizienten tol_j lassen sich die Varianzinflationsfaktoren

$$(2.3.16) \quad VIF_j = \frac{1}{tol_j}, \quad j=2,3,\dots,k,$$

bilden, aus denen die durch Multikollinearität bedingte Inflationierung der Varianzen der Regressionskoeffizienten $\hat{\beta}_j$ hervorgeht. Genauer gibt die Kenngröße VIF_j die Erhöhung der Varianz $Var(\hat{\beta}_j)$ an, die darauf zurückgeführt werden kann, dass die Variable x_j mit den übrigen exogenen Variablen des Regressionsmodells nicht unkorreliert ist, sondern in einer durch das Bestimmtheitsmaß R_j^2 gemessenen kollinearen Beziehung steht.

Beispiel 2.3.1: In der Geldnachfragefunktion (2.1.29) wird die logarithmierte Geldnachfrage ($\ln m$) durch die auf einer logarithmischen Skala gemessenen exogenen

Variablen Einkommen ($\ln y$) und Zinssatz ($\ln r$) erklärt. Die einfache und erweiterte Korrelationsanalyse zur Aufdeckung von Multikollinearität stützen sich hier allein auf den Korrelationskoeffizienten r_{23} zwischen $\ln y$ und $\ln r$, der einen Wert von $-0,278$ annimmt. Da $|r_{23}| = 0,278$ sowohl kleiner als die durch die Faustregel der einfachen Korrelationsanalyse gegebene Schranke von $0,8$ als auch kleiner als der multiple Korrelationskoeffizient $R_{y \times 23}$ von $0,989$ ist, geben beide Verfahren keinen Hinweis auf die Präsenz von Multikollinearität.

Dementsprechend weicht die Determinante der Korrelationsmatrix \mathbf{R} der exogenen Variablen beträchtlich von null ab:

$$|\mathbf{R}| = \det \begin{bmatrix} 1 & 0,278 \\ 0,278 & 1 \end{bmatrix} = 1 - 0,278^2 = 0,923$$

In dem Verfahren der Hilfsregressionen ist allein *eine* Regression durchzuführen, da es im Hinblick auf die Bestimmtheit gleichgültig ist, ob $\ln y$ auf $\ln r$ oder $\ln r$ auf $\ln y$ regressiert wird. Die OLS-Schätzung der ersteren Regression führt zu der Regressionsfunktion

$$\hat{\ln y} = 7,585 - 0,092 \cdot \ln r,$$

die eine Bestimmtheit von $0,077$ hat. Will man dieses Bestimmtheitsmaß ($R^2_2 = 0,077$) mit dem F-Test auf Signifikanz testen, ist der empirische F-Wert (2.3.14) mit dem $(1-\alpha)$ -Quantil der F-Verteilung mit $k-2=3-2=1$ und $n-k+1=20-3+1=18$ Freiheitsgraden zu vergleichen. Der Wert der Prüfgröße (2.3.14) von $1,502$ liegt bei einem Signifikanzniveau von 5% (1%) deutlich unterhalb des kritischen Wert von $4,43$ ($8,33$). Bei deskriptiver Ausrichtung des Verfahrens wäre der Toleranzkoeffizient tol_2 von $0,923$ ($=1-0,077$) mit der durch die Faustregel gegebene Schranke von $0,05$ zu vergleichen, die klar überschritten wird. Beide Vorgehensweisen führen daher zu dem Ergebnis, dass das Multikollinearitätsproblem bei der ökonometrischen Schätzung der Geldnachfragegleichung nicht gravierend ist.

Man wird aus diesem Grund davon ausgehen können, dass die Isolierung des Einflusses der beiden exogenen Variablen auf die Geldnachfrage hier zuverlässig geschätzt worden ist. Der Varianzinflationsfaktor VIF_2 von $1,084$ ($=1/0,923$) zeigt aber an, dass Multikollinearität nicht völlig abwesend ist: Sie bläht die Varianz des OLS-Schätzers $\hat{\beta}_2$ (und damit auch die Varianz von $\hat{\beta}_3$) um $8,4\%$ auf. ♦

2.3.4 Überwindung von Multikollinearität

Wenn ein ökonometrisches Modell mit Multikollinearität behaftet ist, stellt sich für den empirischen Forscher die Frage, wie diesem Problem am besten Rechnung getragen werden kann. Während das Multikollinearitätsproblem für Prognosezwecke eventuell nicht berücksichtigt werden braucht, ist es bei anderen ökonometrischen Analysen häufig wünschenswert, es auszuschalten oder zumindest zu vermindern. Bei Querschnittsanalysen lässt sich das Problem unter Umständen dadurch lösen, dass die Datenbasis erweitert wird. Voraussetzung hierfür ist jedoch, dass die Beobachtungswerte der exogenen Variablen zusätzlich einbezogener statistischer Einheiten einen breiteren Streubereich aufweisen als das vorhandene Datenmaterial. Die Möglichkeit einer **Erweiterung der Datenbasis** ist bei Zeitreihendaten dagegen häufig nicht gegeben. Zwar könnten möglicherweise zusätzliche Daten aus der weiter zurückliegenden Vergangenheit beschafft werden. Doch ist in diesem Fall die Gefahr eines Strukturbruchs zu berücksichtigen, der eine einheitliche Behandlung des Datenmaterials verhindert. Die Verwendung von Paneldaten (→ Abschn. 2.12), d.h. die Kombination von Zeitreihen- und Querschnittsdaten, kann in empirisch-ökonometrischen Analysen jedoch entscheidend zu einer Entschärfung des Multikollinearitätsproblems beitragen.

Wenn sich die Datenbasis nicht auf diese Weise erweitern lässt, könnte es jedoch möglich sein, **externe Informationen** in das ökonometrische Modell zu integrieren, die die Multikollinearität beseitigen könnten. Angenommen, es wird die Nachfrage nach einer Gütergruppe in Abhängigkeit von den Konsumausgaben und dem Preisniveau unter Verwendung von Zeitreihendaten ökonometrisch untersucht. Da die Konsumausgaben häufig mit dem Preisniveau korreliert sein werden, ist mit Multikollinearität zu rechnen. Wenn nun die Größenordnung des Einflusses der Konsumausgaben auf die Nachfrage der Gütergruppe z.B. aus einer Querschnittserhebung bekannt ist, stellt sich die Frage, inwieweit der Regressionskoeffizient dieser exogenen Variablen bei der Schätzung der Nachfragefunktion berücksichtigt werden kann. Falls sich dies als sinnvoll erweisen sollte, wäre die Multikollinearität beseitigt, da nur noch der Einfluss des Preisniveaus auf die Güternachfrage geschätzt zu werden bräuchte. Allerdings ist eine Verknüpfung von Querschnitts- und Zeitreihendaten nicht nur aufgrund des Aggregationsproblems problematisch. Denn möglicherweise wird ein langfristiger Reaktionskoeffizient benötigt, während der vorliegende Regressionskoeffizient eine kurzfristige Wirkung wiedergibt. Aus der ökonomischen Theorie selbst werden sich dagegen nur in Ausnahmefällen Aussagen über die Größenordnung eines Parameters ableiten lassen.

Sofern sich eine Panelanalyse als nicht gangbar erweist, wird man im Regelfall die Multikollinearität nicht durch eine Erweiterung der Datenbasis oder die Verwendung externer Informationen vermindern können. Bei Beobachtungsdaten, d.h. nicht-experimentellen Daten wird das Problem der Multikollinearität mehr oder minder stark stets präsent sein. In jedem Fall ist eine festgestellte Multikollinearität bei der Interpretation von Modellschätzungen zu berücksichtigen. Grundsätzlich kann man zwar auf Bereinigungsverfahren zurückgreifen, die auf die vorhandene Datenbasis anwendbar sind. Die sich hierbei ergebenden Vorteile müssen aber mit Nachteilen erkaufte werden, die durchaus gravierend sein können. Die Vorschläge reichen hier von einer Variablenunterdrückung bis hin zu einer Hauptkomponentenregression. Wir beschränken uns abschließend an dieser Stelle auf die Vorstellung zweier einfacher Bereinigungsverfahren. Die Hauptkomponentenregression (principal components regression) erfordert Kenntnisse der multivariaten Statistik, die hier nicht behandelt werden, so dass auf eine Diskussion dieses Verfahrens verzichtet wird. Sie ist z.B. in der "Advanced Econometrics Methods" von Fomby, Hill und Johnson (1988) vorzufinden.

Das **Verfahren der Variablenunterdrückung** versucht, das Multikollinearitätsproblem dadurch zu lösen, indem eine oder mehrere hochkorrelierende Variablen aus dem Regressionsmodell ausgeschlossen werden. Wenn z.B. zwei Variablen hoch korreliert sind, erfasst eine der beiden Variablen zugleich den größten Teil des Einflusses der anderen Variablen, so dass der "Erklärungsgehalt" bei der Unterdrückung einer Variablen im Großen und Ganzen erhalten bleibt. Die Multikollinearität wird dadurch nicht nur vermindert, sondern sogar völlig ausgeschaltet. Gleichwohl geht damit ein unerwünschter Effekt einher. Wenn das ökonometrische Modell ursprünglich korrekt war, ist es nun fehlspezifiziert. Die unterdrückte exogene Variable ist in der Störvariablen enthalten, die nun mit dem verbleibenden Regressor korreliert ist. Der Erwartungswert der Störvariablen kann zudem nicht mehr unbedingt als null angenommen werden, denn er kann ebenfalls einen systematischen Fehler enthalten. Die Optimalitätseigenschaften der OLS-Methode gehen dadurch verloren. Auf keinen Fall kann durch Variablenunterdrückung aber das Ziel einer Separierung des Einflusses der exogenen Variablen erreicht werden.

Beim **Verfahren der Differenzbildung** wird davon ausgegangen, dass die Multikollinearität in ökonometrischen Modellen durch einen Trend in den exogenen Variablen auftritt. Wenn man näherungsweise einen linearen Trend unterstellen kann, ist seine

Ausschaltung durch eine einfache Differenzenbildung gegeben. Bei zwei exogenen Variablen (ohne Scheinvariable) lautet die Regressionsgleichung für die Periode t

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t .$$

Subtrahiert man hiervon die Regressionsgleichung für die Vorperiode $t-1$,

$$y_{t-1} = \beta_1 + \beta_2 x_{2,t-1} + \beta_3 x_{3,t-1} + u_{t-1} ,$$

so erhält man das Regressionsmodell

$$(2.3.17) \quad \Delta y_t = \beta_2 \Delta x_{2t} + \beta_3 \Delta x_{3t} + \Delta u_t ,$$

wobei z.B. $\Delta y_t = y_t - y_{t-1}$ bedeutet. In dem Maße, wie es gelingt, den Trend aus den exogenen Variablen x_2 und x_3 auszuschalten, vermindert sich die Multikollinearität gegenüber dem ursprünglichen Modell mit den Niveauvariablen. Gleichung (2.3.17) bezeichnet man als homogenes Regressionsmodell, da die Scheinvariable x_1 eliminiert ist. Ein absolutes Glied würde nur dann erhalten bleiben, wenn eine exogene Variable in dem ursprünglichen Modell eine Trendvariable wäre. Dann würde die Einführung eines absoluten Gliedes in (2.3.17) bei einer ökonometrischen Schätzung den Trendeinfluss in dem ursprünglichen Regressionsmodell messen.

Allerdings ist das Verfahren der Differenzenbildung nicht unproblematisch. Zunächst einmal lässt sich eine Verminderung der Multikollinearität allein bei einem näherungsweise linearen Trend in den exogenen Variablen theoretisch begründen. Darüber hinaus gibt es keine Gewähr, dass eine Ausschaltung der Multikollinearität durch dieses Verfahren die Unsicherheit der Parameterschätzung vermindert. Denn durch die Differenzenbildung wird der Streubereich der exogenen Variablen erheblich eingeschränkt, was zu einer Vergrößerung der Varianzen der geschätzten Regressionskoeffizienten führt. Dadurch kann es dazu kommen, dass sich der Einfluss einer exogenen Variablen auf die endogene Variablen als nicht signifikant erweist, obwohl sie tatsächlich vielleicht bedeutsam ist. Außerdem sind die Störvariablen Δu_t autokorreliert, wenn die Störvariablen u_t die klassischen Eigenschaften aus (2.1.7) erfüllen. Der OLS Schätzer hat dann nicht mehr die BLUE-Eigenschaft.

Aufgaben

- (2.3.1)** Worin besteht das Problem der Multikollinearität?
- (2.3.2)** Welche Interpretationsprobleme ergeben sich im Falle von Multikollinearität bei ökonomischen Modellschätzungen?
- (2.3.3)** Nennen Sie Verfahren zur Aufdeckung von Multikollinearität.
- (2.3.4)** Welche Probleme bringen die Verfahren der Variablenunterdrückung und Differenzenbildung zur Überwindung von Multikollinearität mit sich?
- (2.3.5)** Beurteilen Sie das Ausmaß der Multikollinearität anhand einer erweiterten einfachen Korrelationsanalyse und der Determinante der Korrelationsmatrix der exogenen Variablen für das Energiemodell III (Nachfrage nach Erdgas (GASV) in Abhängigkeit von den Preisen für Erdgas (GASPR) und Fernwärme (FERNWPR) sowie dem verfügbaren Einkommen (VEINKR), Daten s. Anhang B, Tab. B1)!
- (2.3.6)** Führen Sie für das Energiemodell III (s. Aufgabe (2.3.5)) das Verfahren zu Hilfsregressionen zur Aufdeckung von Multikollinearität durch!

2.4 Heteroskedastizität und Autokorrelation

2.4.1 Form und Auswirkungen der Modelldefekte

Nun werden mit der Heteroskedastizität und der Autokorrelation diejenigen Modelldefekte diskutiert, die sich beide auf die Varianz-Kovarianz-Matrix der Störterme beziehen. In den bisher verwendeten Regressionsmodellen wurde stets die Annahme

$$(2.4.1) \quad \text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma^2 \mathbf{I}$$

einer skalaren Varianz-Kovarianz-Matrix unterstellt, die bei einer näheren Betrachtung zwei unterschiedliche Implikationen beinhaltet: Erstens haben die Störterme eine konstante und endliche Varianz, d.h. sie sind homoskedastisch. Diese Hypothese ist jedoch insbesondere bei der Analyse von zeitabhängigen Daten nicht unproblematisch. Folgen etwa die exogenen Variablen einem aufsteigenden Trend, dann kann auch eine im Zeitablauf steigende Störgrößenvarianz vorliegen. Sofern allgemeiner zugelassen wird, dass sich die Varianzen der Störterme unterschiedlicher Beobachtungspunkte voneinander unterscheiden, spricht man von **Heteroskedastizität**. In diesem Fall ist die Varianz-Kovarianz-Matrix der Störterme durch

$$(2.4.2) \quad \text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

gegeben. Die Kovarianzmatrix $\text{Cov}(\mathbf{u})$ ist somit eine $n \times n$ -Diagonalmatrix, deren Hauptdiagonalelemente σ_t^2 die Varianz der Störterme bezeichnet, der bei der t -ten Beobachtung auftritt. Bei Heteroskedastizität gilt allgemein $\sigma_t^2 \neq \sigma_s^2$ für $t \neq s$, so dass sich die Varianzen der Störterme unterschiedlicher Beobachtungen voneinander unterscheiden. Somit sind in der Varianz-Kovarianz-Matrix (2.4.2) n Parameter unbekannt. Zusammen mit den k Parametern im Vektor $\boldsymbol{\beta}$ der Regressionskoeffizienten wären daher in einem allgemeineren Modell insgesamt $k+n$ Parameter zu schätzen, was bei n Beobachtungen jedoch an den fehlenden Freiheitsgraden der Schätzung scheitern muss.

Allgemein ist bei einer Verletzung der Annahme (2.4.1) die $n \times n$ -Einheitsmatrix I durch eine $n \times n$ -Matrix Ω zu ersetzen, die die Varianzen und Kovarianzen der Störvariablen unterschiedlicher Beobachtungspunkte bis auf den sog. Skalenfaktor σ^2 wiedergibt:

$$(2.4.3) \quad \text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma^2 \Omega.$$

Im Falle der Heteroskedastizität nimmt die Matrix Ω Diagonalgestalt an:

$$(2.4.4) \quad \Omega = \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_1 \end{bmatrix}.$$

Die Definition der Kovarianzmatrix (2.4.2) ist hiermit verträglich, wenn man für die Varianzen der Störgrößen die Beziehung

$$(2.4.5) \quad \sigma_t^2 = \sigma^2 \cdot \omega_t$$

unterstellt. Die Größe ω lässt sich hierin als eine Skalierungsvariable auffassen. Gleichung (2.4.5) stellt eine wichtige Relation dar, um Heteroskedastizität bei einer Modellschätzung trotz der fehlenden Freiheitsgrade einfangen zu können.

Die zweite Implikation der Annahme einer skalaren Kovarianzmatrix (2.4.1) ist, dass sich Störterme unterschiedlicher Beobachtungen nicht beeinflussen, also unkorreliert sind. In den bisher diskutierten Regressionsmodellen wurde somit die Abwesenheit von **Autokorrelation** unterstellt. Diese Hypothese ist etwa bei konjunkturellen Daten kritisch, wenn durch den Regressionsansatz eine lineare Beziehung zwischen der endogenen und den exogenen Variablen unterstellt wird. Allgemein können sich autokorrelierte Störterme jedoch auch dann ergeben, wenn im zugrunde gelegten Regressionsmodell relevante Erklärungsvariablen nicht berücksichtigt werden, z.B. um dem Problem der Multikollinearität zu entgehen oder um die Anzahl der Freiheitsgrade bei der Schätzung nicht zu reduzieren. Sofern die Realisationen der ausgeschlossenen Variablen in der Beobachtungsperiode variieren, was speziell bei zeitabhängigen Daten nahezu stets der Fall sein dürfte, wird eine signifikante Korrelation der Störterme unterschiedlicher Zeiteinheiten hervorgerufen.

Bei autokorrelierten Störtermen ist die Varianz-Kovarianz-Matrix anstelle von (2.4.1) durch

$$(2.4.6) \quad \text{Cov}(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma^2 \end{bmatrix}$$

gegeben. In (2.4.6) wurde erneut die Annahme der Homoskedastizität vorausgesetzt, so dass die Diagonalelemente der Matrix identisch sind. Diese Vorgehensweise wird hier deshalb gewählt, um die beiden unabhängigen Modelldefekte separat analysieren zu können.

Das Nebendiagonalelement $\sigma_{ts} = E(u_t \cdot u_s)$, $t \neq s$, $t, s = 1, 2, \dots, n$, der Varianz-Kovarianz-Matrix (2.4.6) bezeichnet die Autokovarianz zwischen den Störtermen der t -ten und der s -ten Beobachtung. Die Autokovarianzen sind bei Vorliegen einer Autokorrelationsstruktur im Allgemeinen verschieden. Jede der n Zeilen der Kovarianzmatrix enthält somit $n-1$ unterschiedliche Autokovarianzen. Da die Matrix außerdem symmetrisch ist ($\sigma_{ts} = \sigma_{st}$), sind zusätzlich zu den $k+1$ Parametern im multiplen Regressionsmodell (k Parameter im Vektor $\boldsymbol{\beta}$ sowie die Störvarianz σ^2) $n(n-1)/2$ Autokovarianzen σ_{ts} zu schätzen. Wie im Fall der Heteroskedastizität ist bei Autokorrelation eine Schätzung des allgemeineren Modells ohne weitere Annahmen aufgrund der fehlenden Freiheitsgrade nicht möglich.

Wie die Heteroskedastizität lässt damit auch die Autokorrelation bei einer Modellschätzung nur durch Auferlegung bestimmter Restriktionen berücksichtigen. Wie wir später sehen werden (\rightarrow Abschn. 2.4.4.2.2), spezifiziert sich die Matrix $\boldsymbol{\Omega}$ z.B. bei einer Autokorrelation erster Ordnung zu

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix}.$$

Unter dieser Restriktion ist neben der Störvarianz σ^2 allein der sog. autoregressive Parameter ϕ zu schätzen. Die Anzahl der unbekannt Parameter hat sich dadurch erheblich reduziert.

Grundsätzlich empfiehlt es sich, vorab statistische Tests auf Heteroskedastizität und Autokorrelation durchzuführen, um Hinweise auf die Art und Form des Modelldefektes

zu erhalten. Abhängig von den Ergebnissen dieser Tests lässt sich dann die Varianz-Kovarianz-Matrix der Störterme spezifizieren, die in einem verallgemeinerten multiplen Regressionsmodell bei der Schätzung der Regressionskoeffizienten berücksichtigt wird.

Für den Fall der Heteroskedastizität soll hier genauer gezeigt werden, welche Auswirkungen dieses Modelleffekt auf die Schätzeigenschaften des Kleinst-Quadrate-Schätzers $\hat{\beta}$ in unmodifizierter Form nach sich ziehen würde. Der Anschaulichkeit halber gehen wir hierzu vom einfachen Regressionsmodell

$$y_t = \beta_1 + \beta_2 x_t + u_t$$

aus und beschränken uns auf den OLS-Schätzer $\hat{\beta}_2$, der sich wegen (2.1.21) in der Form

$$(2.4.7) \quad \hat{\beta}_2 = \frac{s_{xy}}{s_x^2} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} = \frac{\sum(x_t - \bar{x})y_t - \sum(x_t - \bar{x})\bar{y}}{\sum(x_t - \bar{x})^2}$$

darstellen lässt. Gleichung (2.4.7) lässt sich unter Verwendung von (2.1.17) zu

$$(2.4.8) \quad \hat{\beta}_2 = \frac{\sum(x_t - \bar{x})y_t}{\sum(x_t - \bar{x})^2} = \beta_2 + \frac{\sum(x_t - \bar{x})}{\sum(x_t - \bar{x})^2} \cdot u_t$$

vereinfachen, woraus für den Erwartungswert

$$E(\hat{\beta}_2) = E\left[\beta_2 + \frac{\sum(x_t - \bar{x})}{\sum(x_t - \bar{x})^2} \cdot u_t\right] = \beta_2 + \frac{\sum(x_t - \bar{x})}{\sum(x_t - \bar{x})^2} \cdot E(u_t)$$

und mit der Annahme (2.1.2), $E(u_t) = 0$,

$$E(\hat{\beta}_2) = \beta_2$$

folgt. Hierbei ist von der Eigenschaft vorgegebener fester x -Werte Gebrauch gemacht worden. Wie sich zeigt, ist der OLS-Schätzer für β_2 erwartungstreu unabhängig davon, ob die Störvariable u_t homoskedastisch oder heteroskedastisch ist.

Dies trifft jedoch für die Varianz von $\hat{\beta}_2$ nicht zu. Im Gegenteil kann die Verzerrung hier beträchtlich sein. Unter Verwendung von (2.4.8) erhält man

$$\text{Var}(\hat{\beta}_2) = \text{Var}\left[\beta_2 + \frac{\sum(x_t - \bar{x})}{\sum(x_t - \bar{x})^2} \cdot u_t\right] = \text{Var}\left[\frac{\sum(x_t - \bar{x})}{\sum(x_t - \bar{x})^2} \cdot u_t\right],$$

da β_2 als additive Konstante mit einer Varianz von null vernachlässigt werden kann. Während man bei einer homoskedastischen Störvariablen u_t den gesamten Bruch wegen $\text{Var}(u_t) = \sigma^2$ herausziehen kann,

$$(2.4.9) \quad \text{Var}(\hat{\beta}_2) = \frac{\sum (x_t - \bar{x})^2}{\left[\sum (x_t - \bar{x})^2 \right]^2} \cdot \text{Var}(u_t) = \frac{1}{\sum (x_t - \bar{x})^2} \cdot \sigma^2$$

ist dies bei einer heteroskedastischen Störvariablen u_t wegen $\text{Var}(u_t) = \sigma_t^2$ nicht möglich. Hierfür erhält man

$$(2.4.10) \quad \text{Var}(\hat{\beta}_2) = \frac{\sum (x_t - \bar{x})^2}{\left[\sum (x_t - \bar{x})^2 \right]^2} \cdot \text{Var}(u_t) = \frac{\sum (x_t - \bar{x})^2 \sigma_t^2}{\left[\sum (x_t - \bar{x})^2 \right]^2},$$

was für $\sigma_t^2 \neq \sigma_s^2, t \neq s$, die Problematik der OLS-Schätzung bei Heteroskedastizität transparent werden lässt. Im Falle einer heteroskedastischen Störvariablen u_t ist die Varianz von $\hat{\beta}_2$ durch (2.4.10) gegeben. Gleichwohl wird im Falle einer OLS-Schätzung (2.4.9) zur Bestimmung eines Varianzschätzers für $\hat{\beta}_2$ herangezogen. Sofern jedoch $\sigma_t^2 \neq \sigma_s^2$ für $t \neq s$ gilt, differieren die in (2.4.9) und (2.4.10) wiedergegebenen Varianzen voneinander, so dass auf der Basis von (2.4.9) ein verzerrter Varianzschätzer für $\hat{\beta}_2$ resultiert. Damit ist aber zugleich auch die t-Statistik als Prüfgröße des Signifikanztests für $\hat{\beta}_2$ verzerrt. Die Anwendung des t-Tests in der üblichen Form vermittelt daher nicht mehr notwendig zuverlässige Aussagen über die Signifikanz einer Einflussgröße.

2.4.2 Tests auf Heteroskedastizität

2.4.2.1 Goldfeld-Quandt-Test

Mit dem Goldfeld-Quandt-Test (Goldfeld und Quandt, 1965) stellen wir hier ein eher traditionelles Verfahren zur Überprüfung der Nullhypothese homoskedastischer Störterme vor, das sich jedoch in der Praxis großer Beliebtheit erfreut. Für die Störgrößen gelten die bereits diskutierten Annahmen,

$$(2.4.11) \quad E(u_t) = 0, \quad t = 1, \dots, n$$

$$(2.4.12) \quad E(u_t u_s) = 0, \quad t \neq s, \quad t = 1, \dots, n$$

d.h., die Störterme haben einen Erwartungswert von 0 und sind nicht miteinander korreliert. Jedoch wird die Hypothese der Varianzhomogenität in Zweifel gezogen. Beim Goldfeld-Quandt-Test werden normalverteilte Störterme unterstellt.

Die Grundidee des Tests besteht darin, dass sich die Varianz der Störterme von den Regressoren abhängig entwickelt. Variiert die Störvarianz etwa mit steigenden Werten der j -ten exogenen Variablen in der Form

$$\sigma_t^2 = \sigma \cdot x_{jt}^2, \quad t=1,2,\dots,n,$$

dann fungiert die Variable x_j als Skalierungsvariable, d.h. es gilt $\omega_t = x_{jt}^2$. Um diese Hypothese in einen statistischen Test umzusetzen, werden in einem ersten Schritt die vorliegenden n Beobachtungen derart umgruppiert, dass ihre neue Reihenfolge den aufsteigenden Werten des j -ten Regressors entspricht. Da die Ordnung der Daten aber nur zur Bildung zweier Gruppen „niedriger“ und „großer“ Regressorwerte dient, empfiehlt es sich zur Erhöhung der Trennschärfe des Tests, in der Mitte der geordneten Stichprobe c Beobachtungen wegzulassen. Man erhält somit zwei Stichproben, die jeweils $(n-c)/2$ Beobachtungen umfassen. Der Goldfeld-Quandt-Test überprüft nun, ob eine **gruppenweise Heteroskedastizität** vorliegt, d.h. ob sich die Varianzen der Störterme in den beiden Stichproben, σ_1^2 und σ_2^2 , signifikant voneinander unterscheiden. Im Falle eines signifikanten Unterschieds wird die Nullhypothese

$$(2.4.13) \quad H_0 : \sigma_1 = \sigma_2 = \sigma$$

zugunsten der Alternativhypothese

$$(2.4.14) \quad H_1 : \sigma_1 \neq \sigma_2$$

verworfen. Innerhalb jeder Stichprobe gilt dabei weiterhin die Annahme der Homoskedastizität, so dass bei einer Ablehnung der Nullhypothese nur auf Heteroskedastizität der Störterme zwischen den beiden unterschiedlicher Stichproben geschlossen werden kann. ($\sigma_1^2 \neq \sigma_2^2$). Der Ausschluss der mittleren c Beobachtungen soll verhindern, dass eventuell vorhandene Unterschiede zwischen σ_1^2 und σ_2^2 aufgrund annähernd identischer Skalierungsfaktoren ω_t unentdeckt bleiben.

In jeder der beiden Stichproben wird nun eine Regression

$$(2.4.15) \quad y_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{u}_i, \quad i = 1, 2$$

nach der OLS-Methode durchgeführt, wobei der Index i die jeweilige Stichprobe bezeichnet. Neben der Ungleichung $(n - c)/2 > k$ muss dabei natürlich auch die Rangbedingung $\text{rg}(\mathbf{X}_i) = k$ erfüllt sein. Aus den geschätzten Regressionen lassen sich dann die OLS-Residuen wie üblich durch

$$(2.4.16) \quad \hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i \quad i = 1, 2$$

berechnen. Für den Goldfeld-Quandt-Test ist nun der Wert der Teststatistik

$$(2.4.17) \quad \text{GQ} = \hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 / \hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1$$

zu bestimmen, also der Quotient der Residuenquadratsummen, die sich aus den getrennt durchgeführten OLS-Schätzungen ergeben. Da die Schätzung der Varianz der Störterme in der i -ten Stichprobe mit

$$(2.4.18) \quad \hat{\sigma}_i^2 = \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i / [(n-c)/2 - k], \quad i = 1, 2$$

erfolgt, lässt sich die Teststatistik (2.4.17) alternativ als

$$(2.4.19) \quad \text{GQ} = \hat{\sigma}_2^2 / \hat{\sigma}_1^2$$

schreiben. Aus dieser Formulierung folgt unmittelbar, dass die Größe GQ zum Test der Nullhypothese homoskedastischer Störterme eingesetzt werden kann. Da ein Anstieg der Varianz der Störterme mit wachsenden Werten des j -ten Regressors vermutet wird und die zweite Stichprobe qua Konstruktion die hohen Werte dieses Regressors enthält, spezifiziert sich die Alternativhypothese (2.4.14) zu

$$(2.4.20) \quad H_1: \sigma_1^2 < \sigma_2^2.$$

Bei dieser Form der Heteroskedastizität nimmt die Prüfgröße GQ tendenziell Werte an, die größer als eins sind. Wird umgekehrt unterstellt, dass die Varianz der Störterme mit steigenden Werten des j -ten Regressors sinkt, wird

$$(2.4.21) \quad H_1: \sigma_1^2 > \sigma_2^2$$

als Alternativhypothese verwendet. Um auch hier zu gewährleisten, dass die Prüfgröße bei Heteroskedastizität hohe Werte annimmt, ist anstelle von GQ die inverse Teststatistik $1/\text{GQ}$ zu berechnen. In jedem Fall wird jedoch nur getestet, ob Störterme unterschiedlicher Stichproben heteroskedastisch sind.

Die Nullhypothese homoskedastischer Störterme wird abgelehnt, wenn die Prüfgröße GQ (bzw. $1/\text{GQ}$) zu große Werte annimmt. Um genauer beurteilen zu können, wann dabei die Werte der Teststatistik so gravierend sind, dass die Nullhypothese verworfen

werden muss, ist noch die Verteilung der Testgröße unter H_0 zu spezifizieren. Aus der Darstellung des multiplen Regressionsmodells ist bekannt, dass die durch die Störtermvarianz dividierte Quadratsumme der OLS-Residuen,

$$\hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma^2,$$

bei normalverteilten Störtermen χ^2 -verteilt ist mit $n-k$ Freiheitsgraden. Analog sind hier die Größen

$$\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1/\sigma^2 \quad \text{und} \quad \hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2/\sigma^2$$

jeweils χ^2 -verteilt mit $[(n-c)/2]-k$ Freiheitsgraden. Die Quadratsummen sind dabei unabhängig voneinander χ^2 -verteilt, weil sie aus zwei getrennt durchgeführten Regressionen entstammen. Sofern nun zwei unabhängig voneinander χ^2 -verteilte Zufallsvariablen zunächst durch ihre Freiheitsgrade dividiert und dann aufeinander bezogen werden, ist der Quotient F-verteilt mit den Freiheitsgraden der beiden χ^2 -verteilten Zufallsvariablen (siehe hierzu Eckey, Kosfeld und Dreger (2002), S. 509ff.). Hier entspricht der Quotient genau der Prüfgröße GQ,

$$GQ = \frac{\hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2/\sigma^2 / [(n-c)/2 - k]}{\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1/\sigma^2 / [(n-c)/2 - k]} = \frac{\hat{\mathbf{u}}_2'\hat{\mathbf{u}}_2}{\hat{\mathbf{u}}_1'\hat{\mathbf{u}}_1},$$

die damit F-verteilt mit $(n-c)/2 - k$ und $(n-c)/2 - k$ Freiheitsgraden ist. Sofern danach der Wert der Goldfeld-Quandt-Statistik den kritischen Wert der F-Verteilung bei diesen Freiheitsgraden übersteigt, ist die Nullhypothese mit der gewählten Irrtumswahrscheinlichkeit abzulehnen.

Beispiel 2.4. 1: Bei der Anwendung des Goldfeld-Quandt-Tests auf die Geldnachfragefunktion (2.1.29) setzen wir $c=2$, so dass die beiden mittleren Beobachtungen ausgeschlossen werden. Aufgrund der hohen Korrelation zwischen der Geldmenge und des Brutto sozialprodukts ($r=0,978$ für die realen Größen auf einer logarithmischen Skala) bietet es sich an, die Beobachtungen unter Verwendung des Regressors $\ln y_t$ in aufsteigender Reihenfolge zu ordnen. Die beiden Stichproben der „kleineren“ und „größeren“ Werte sind dann wie folgt abgegrenzt:

1. Stichprobe		2. Stichprobe	
$\ln y_t$	Jahr	$\ln y_t$	Jahr
7,187506	1970	7,458071	1980
7,217517	1971	7,459166	1981
7,260031	1972	7,466456	1983
7,307270	1973	7,496653	1984
7,307806	1974	7,514527	1985
7,295056	1975	7,536044	1986
7,349038	1976	7,550819	1987
7,374253	1977	7,586702	1988
7,408167	1978	7,624033	1989

Die 1. Stichprobe umfasst damit den Zeitraum von 1970 bis 1978; die 2. Stichprobe enthält die Beobachtungen der Jahre 1980 und 1981 sowie des Zeitraums von 1983 bis 1989. In den Jahren 1979 und 1982 nimmt der Regressor $\ln y_t$ die Werte 7,447693 bzw. 7,448042 an, die genau zwischen den beiden Gruppierungen liegen und daher ausgeschlossen werden.

Eine OLS-Schätzung der Geldnachfragefunktion (2.1.28) mit der 1. Stichprobe ergibt

$$\hat{\ln m}_t = -2,501 + 1,123 \cdot \ln y_t - 0,099 \cdot \ln r_t,$$

$$R_1^2 = 0,956, \quad \hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 = 0,004482,$$

während die 2. Stichprobe die Schätzergebnisse

$$\hat{\ln m}_t = -8,757 + 1,949 \cdot \ln y_t - 0,042 \cdot \ln r_t,$$

$$R_2^2 = 0,982, \quad \hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 = 0,002203$$

hervorbringt. Da die Quadratsumme der Residuen in der 1. Stichprobe diejenige der 2. Stichprobe übersteigt, wird die Teststatistik des Goldfeld-Quandt-Tests in der Form

$$1/\text{GQ} = \hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1 / \hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2 = 0,004482 / 0,002203 = 2,0345$$

gebildet. Die F-verteilte Prüfgröße $1/\text{GQ}$ hat im Zähler und Nenner $(n - c)/2 - k = (20 - 2)/2 - 3 = 6$ Freiheitsgrade. Hiermit ergibt sich bei einem Signifikanzniveau von 5 % ein kritischer Wert von $F_{6,6;0,95} = 4,28$, der den Wert der Prüfgröße übersteigt. Auf dem 5%-Signifikanzniveau kann die Nullhypothese der Homoskedastizität der Störgröße der Geldnachfragefunktion auf der Basis des Goldfeld-Quandt-Tests daher nicht abgelehnt werden. ♦

Problematisch für das Testverfahren erscheint zunächst die Wahl des j -ten Regressors, nach dessen Werten die Beobachtungen im ersten Schritt geordnet werden. Dieser Nachteil stellt angesichts der heutigen Verfügbarkeit ökonometrischer Programmpakete auf den EDV-Anlagen jedoch keine echte Beschränkung mehr dar. Sofern in der Anwendung nicht eindeutig klar ist, welcher der $k-1$ (echten) Regressoren die Varianz der Störterme determiniert, kann der Goldfeld-Quandt-Test relativ schnell für alternative Regressoren berechnet werden. Das Testverfahren ist darüber hinaus auch dann anwendbar, wenn der Vektor der Regressionskoeffizienten in den beiden Stichproben nicht identisch ist, also z.B. Strukturbrüche im Untersuchungszeitraum aufgetreten sind.

Kritisch muss jedoch angemerkt werden, dass der vorgestellte Test extrem konservativ ist, in vielen Fällen also die Nullhypothese homoskedastischer Störterme eindeutig bevorzugt. Diese Aussage hängt mit den c Beobachtungen zusammen, die in der Mitte der geordneten Stichprobe weggelassen werden. Genauer werden bei zu kleinen Werten von c die Unterschiede zwischen den Residuenquadratsummen $\hat{\mathbf{u}}_1' \hat{\mathbf{u}}_1$ und $\hat{\mathbf{u}}_2' \hat{\mathbf{u}}_2$ eher geringer, so dass sich insgesamt kleinere Werte der GQ-Statistik ergeben. Werden dagegen zu viele Beobachtungen weggelassen, hat man für den Test eine kleinere Anzahl von Freiheitsgraden zur Verfügung. Da sinkende Freiheitsgrade jedoch mit wachsenden kritischen Werten der F-Verteilung einhergehen, kann auch in diesem Fall die Nullhypothese nur schwer verworfen werden. Aus diesem Grund ist der Einsatz alternativer Testverfahren zur Aufdeckung möglicher Heteroskedastizität in Erwägung zu ziehen.

2.4.2.2 Breusch-Pagan-Test

Eine Alternative zum Goldfeld-Quandt-Test bietet der Breusch-Pagan-Test (Breusch und Pagan, 1980). Er unterstellt, dass sich die Varianzen der Störvariablen u_t des multiplen Regressionsmodells

$$(2.4.22) \quad y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

unter der Alternativhypothese H_1 als Funktion der beobachtbaren Variablen z_2, z_3, \dots, z_p in der Form

$$(2.4.23) \quad \sigma_t^2 = h(\alpha_1 + \alpha_2 \cdot z_{2t} + \dots + \alpha_p \cdot z_{pt})$$

oder mit

$$\boldsymbol{\alpha}' = (\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_p)$$

und

$$\mathbf{z}'_t = (z_{1t} = 1 \quad z_{2t} \quad \cdots \quad z_{pt})$$

$$(2.4.24) \quad \sigma_t^2 = h(\mathbf{z}'_t \boldsymbol{\alpha})$$

darstellen lassen. Die Funktion h braucht hierbei nicht näher spezifiziert zu werden. Mit dem Ansatz lassen sich somit verschiedene Fälle möglicher Heteroskedastizität wie z.B.

$$\sigma_t^2 = \exp(\mathbf{z}'_t \boldsymbol{\alpha}) \quad \text{oder} \quad \sigma_t^2 = \mathbf{z}'_t \boldsymbol{\alpha}$$

abdecken. Dabei kann die Varianz der Störterme über den gesamten Zeitraum variieren. Gefordert wird allerdings, dass diese Variation vollständig auf die Variablen z_2, z_3, \dots, z_p zurückführbar ist.

Für $p=2$ ist die Varianz σ_t^2 der Störvariablen u_t allein von einer Variablen $z=z_2$ abhängig:

$$\sigma_t^2 = h(\alpha_1 + \alpha_2 z_t).$$

Speziell ist die Störvarianz σ_t^2 bei Linearität in diesem Fall durch

$$\sigma_t^2 = \alpha_1 + \alpha_2 z_t$$

gegeben. Als z -Variable kommt hierbei eine bestimmte exogene Variable z.B. x_2 in Betracht, die auch in logarithmischer Form ($\ln x_2$) oder quadratischer Form (x_2^2) verwendet werden kann.

Getestet wird die Nullhypothese H_0 der Homoskedastizität, unter der die Koeffizienten $\alpha_2, \alpha_3, \dots, \alpha_p$ gleich null sein müssen:

$$(2.4.25) \quad H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_p = 0 .$$

In diesem Fall gilt $\sigma_t^2 = \sigma^2$ für alle $t=1, 2, \dots, n$. Formal lässt sich der Breusch-Pagan-Test als Lagrange-Multiplier-Test (LM-Test) betrachten, da das restringierte Regressionsmodell unter der Nullhypothese mit dem unrestringierten Regressionsmodell unter der Alternativhypothese auf der Basis einer Maximum-Likelihood-Schätzung ver-

glichen wird.²⁹ Speziell werden den $k+p$ zu schätzenden Parametern in α und β unter H_0 $p-1$ Restriktionen auferlegt, die durch (2.4.25) gegeben sind.

Praktisch lässt sich der Breusch-Pagan-Test wie folgt durchführen. Das Regressionsmodell (2.4.22) wird zunächst OLS-geschätzt, woraus die Residuen $\hat{u}_2, \hat{u}_3, \dots, \hat{u}_n$ hervorgehen. Unter Verwendung des Maximum-Likelihood-Schätzer (ML Schätzer) der Störvarianz

$$\hat{\sigma}_{ML}^2 = \sum_{t=1}^n \hat{u}_t^2 / n$$

werden die Größen $\hat{u}_t^2 / \hat{\sigma}_{ML}^2$ konstruiert, die die abhängige Variable der Hilfsregression

$$(2.4.26) \quad \hat{u}_t^2 / \hat{\sigma}_{ML}^2 = \mathbf{z}_t' \alpha + v_t$$

bilden, in der v_t die Störgröße bezeichnet. Nach Durchführung der Hilfsregression (2.4.26) lässt sich die durch die Regression erklärte Summe der Abweichungsquadrate der abhängigen Variablen (SS_E) als Differenz der gesamten Abweichungsquadratsumme (SS_T) und der Summe der quadrierten Residuen (SS_R) bestimmen:

$$SS_E = SS_T - SS_R.$$

Es lässt sich zeigen, dass die halbe Differenzgröße,

$$BP_1 = \frac{1}{2} SS_E,$$

bei einer Normalverteilung der Störgröße u_t unter der Nullhypothese asymptotisch χ^2 -verteilt ist mit $p-1$ Freiheitsgraden:

$$BP_1 \overset{a}{\sim} \chi_{p-1}^2.$$

Die Nullhypothese (2.4.25) der Homoskedastizität ist daher auf einem Signifikanzniveau α zu verwerfen, wenn die Prüfgröße BP das $(1-\alpha)$ -Quantil einer χ^2 -Verteilung mit $p-1$ Freiheitsgraden übersteigt. Dieser Fall wird um so eher eintreten, je stärker die gewählten z -Variablen geeignet sind, vorhandene Variationen in den quadrierten OLS-Residuen zu erklären.

Asymptotisch äquivalent ist der Test unter Verwendung der Prüfgröße

²⁹ Zum Konstruktionsprinzip des Lagrange-Multiplier-Tests (LM-Test) siehe z.B. Darnell (1994), S. 215ff.

$$BP_2 = n \cdot R^2,$$

in der sich das Bestimmtheitsmaß R^2 aus einer Regression der quadrierten OLS-Residuen \hat{u}_t^2 auf z_1, z_2, \dots, z_p ergibt:

$$BP_2 \stackrel{a}{\sim} \chi_{p-1}^2.$$

Hierbei ist die Normalverteilungsannahme entbehrlich. Darnell (1994, S. 178) weist allerdings darauf hin, dass der Test bei endlichen Stichproben die Nullhypothese der Homoskedastizität zu oft verwirft. Der Fehler 1. Art ist mithin größer als es das vorgegebene Signifikanzniveau α indiziert.

Beispiel 2.4. 2: Alternativ zum Goldfeld-Quantt-Test soll hier die Homoskedastizität der Störgröße mit dem Breusch-Pagan-Test geprüft werden. Hierbei wird unterstellt, dass die Störvarianz unter der Alternativhypothese der Heteroskedastizität, σ_t^2 , allein eine Funktion des (logarithmierten) realen Bruttozsozialprodukts $\ln y_t$ ist:

$$\sigma_t^2 = h(\alpha_0 + \alpha_1 \cdot \ln y_t) .$$

Die logarithmierte Form von y ist hier in Übereinstimmung mit der Geldnachfragefunktion (2.1.29) gewählt worden.

Zur Bestimmung des ML-Schätzers $\hat{\sigma}_{ML}^2$ für die Varianz σ^2 der Störvariablen u_t des ökonometrischen Geldnachfragemodells (2.1.29) übernehmen wir die in Beispiel 2.2.3 ermittelte Residualquadratsumme

$$\sum \hat{u}_t^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} = 0,014855 ,$$

mit der sich

$$\hat{\sigma}_{ML}^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{\mathbf{n}} = \frac{0,014855}{20} = 0,00074275$$

ergibt. Hiermit lassen sich die Größen $\hat{u}_t^2 / \hat{\sigma}_{ML}^2$ berechnen, die auf $\ln y_t$ regressiert werden. Aus der OLS-Schätzung

$$\left(\hat{u}_t^2 / \hat{\sigma}_{ML}^2 \right) = \begin{matrix} -13,490 + 1,954 \\ (-0,766) \quad (0,823) \end{matrix}$$

lässt sich die Residualquadratsumme

$$SS_R = 29,41327$$

ermitteln. Mit der Abweichungssumme der abhängigen Variablen $\hat{u}_t^2 / \hat{\sigma}_{ML}^2$ von

$$SS_T = 30,51990$$

erhält man eine durch die Hilfsregression erklärte Summe der Abweichungsquadrate in Höhe von

$$SS_E = SS_T - SS_R = 1,1066.$$

Die Prüfgröße BP_1 des Breusch-Pagan-Tests nimmt mithin den Wert

$$BP_1 = \frac{1}{2} SS_E = 0,55332$$

an, der bei einem Signifikanzniveau von 5% mit den $(1-\alpha)$ -Quantil einer χ^2 -Verteilung mit 1 Freiheitsgrad zu vergleichen ist. Wegen

$$(BP_1 = 0,553) < (\chi_{1,0,95}^2 = 3,84)$$

kann die Nullhypothese homoskedastischer Störterme der Geldnachfragefunktion nicht verworfen werden. ♦

2.4.2.3 White-Test

Ein weiterer Test auf Heteroskedastizität ist durch das Verfahren von White (1980) gegeben, das auf einem Vergleich der geschätzten Varianz-Kovarianzmatrizen des OLS-Schätzers für den Parametervektor β beruht. Dabei wird die Kovarianzmatrix des klassischen Modells, die nur bei homoskedastischen Störprozessen konsistent ist, einer Schätzung gegenübergestellt, die sowohl bei hetero- als auch bei homoskedastischen Störtermen konsistent ist. Der Test wird implementiert, in dem die Residuen des ursprünglichen Modells quadriert und anschließend gemäß

$$(2.4.27) \quad \hat{u}_t^2 = \alpha_0 + \sum_{i=2}^k \alpha_i x_{it} + \sum_{i=2}^k \sum_{j \geq i}^k \alpha_{ij} x_{it} x_{jt} + v_t$$

auf die $k-1$ ursprünglichen Regressoren, ihre Quadrate und ihre Kreuzprodukte regressiert werden. Die Größe v_t ist in der Hilfsregression (2.4.27) die Störvariable. Die Prüf-

größe ergibt sich aus dem Produkt aus dem unkorrigierten Bestimmtheitsmaß der Hilfsregression und der Anzahl der Beobachtungen:

$$W = n \cdot R^2 .$$

Die Teststatistik ist unter der Nullhypothese homoskedastischer Störterme asymptotisch χ^2 -verteilt,

$$W \overset{a}{\sim} \chi_r^2 ,$$

wobei die Zahl der Freiheitsgrade, r , der Anzahl der Regressoren in Gleichung (2.4.27) ohne Berücksichtigung der Konstanten entspricht:

$$r = (k-1)(k+2)/2.$$

Bei einer großen Anzahl von Regressoren kann der White-Test unter Vernachlässigung der quadratischen Terme und Kreuzprodukte durchgeführt werden. In diesem Fall besitzt die Prüfgröße $k-1$ Freiheitsgrade.

Beispiel 2.4. 3: Mit dem White-Test soll überprüft werden, ob die Residuen der Geldnachfragefunktion (2.1.29) heteroskedastisch sind. Dazu werden die quadrierten Residuen auf ursprünglichen Regressoren, ihre Quadrate und ihre Kreuzprodukte regressiert. Geschätzt wird also die Hilfsregression

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 \cdot \ln y_t + \alpha_2 \cdot \ln r_t + \alpha_3 \cdot (\ln y_t)^2 + \alpha_4 \cdot (\ln r_t)^2 + \alpha_5 \cdot \ln y_t \cdot \ln r_t + v_t ,$$

für die sich ein Bestimmtheitsmaß von 0,442 ergibt. Bei 20 Beobachtungen resultiert ein Wert der Prüfgröße von 8,84, der unterhalb des kritischen Werts der Chi-Quadrat-Verteilung $\chi_{5;0,95}^2$ von 11,1 liegt. Bei einem Signifikanzniveau von 5% ergibt sich mithin keine Evidenz für heteroskedastische Störterme. ◆

2.4.3 Tests auf Autokorrelation

2.4.3.1 Durbin-Watson-Test

Bei Vorliegen einer Autokorrelation sind Störterme unterschiedlicher Beobachtungen stochastisch abhängig. Der Autokorrelationskoeffizient j -ter Ordnung für die unbekannt und homoskedastischen Störterme ist mit

$$(2.4.28) \quad \rho_j = \text{Cov}(u_t, u_{t-j}) / \text{Var}(u_t) = \sigma_{t,t-j} / \sigma^2, \quad j = 1, 2, \dots$$

definiert. Er ergibt sich also, indem die Autokovarianz zwischen Störtermen, die im Beobachtungszeitraum j Perioden voneinander entfernt sind, durch die konstante Störtermvarianz dividiert wird. Mit der Annahme $E(u_t) = 0$ für $t = 1, \dots, n$ lässt sich ρ_j einfacher durch

$$(2.4.29) \quad \rho_j = E(u_t \cdot u_{t-j}) / E(u_t^2), \quad j = 1, 2, \dots$$

angeben. Sofern die Störterme unkorreliert sind, gilt $E(u_t \cdot u_{t-j}) = 0$, so dass man für die Autokorrelationskoeffizienten in diesem Fall

$$(2.4.30) \quad \rho_j = 0, \quad j = 1, 2, \dots$$

erhält. Andernfalls heißen die Störterme autokorreliert.

Da die Störvariable eine nicht beobachtbare Größe ist, sind die theoretischen Autokorrelationskoeffizienten ρ_j allerdings im Allgemeinen unbekannt. Bei einem Test auf Autokorrelation werden in die empirischen Autokorrelationskoeffizienten der OLS-Residuen

$$(2.4.31) \quad \hat{\rho}_j = \sum_{t=j+1}^n (\hat{u}_t \cdot \hat{u}_{t-j}) / \sum_{t=1}^n \hat{u}_t^2, \quad j = 1, 2, \dots,$$

die sich als Quotient aus der empirischen Autokovarianz und der im Beobachtungszeitraum identischen Varianz ergeben, herangezogen. Ein signifikanter Unterschied der $\hat{\rho}_j$ vom Wert 0 ist dann als Indiz für autokorrelierte Störterme zu werten. Auf diesem Prinzip basieren die im Folgenden darzustellenden Tests auf Autokorrelation.

Ein möglicher Ansatz zur Modellierung von Autokorrelation ist durch einen so genannten autoregressiven Prozess erster Ordnung (Markov-Prozess)

$$(2.4.32) \quad u_t = \phi \cdot u_{t-1} + v_t, \quad |\phi| < 1, \quad t = 2, \dots, n$$

gegeben. Da der Störterm u_{t-1} der vergangenen Periode den Störterm u_t der gegenwärtigen Periode beeinflusst, spricht man hier von einer Autokorrelation erster Ordnung. Die Bedeutung der Annahme, dass der Parameter ϕ betragsmäßig kleiner als 1 ist, wird später verständlicher. Hier sei zunächst nur angemerkt, dass $|\phi| < 1$ eine endliche, konstante Varianz der u_t garantiert.

Die Größen v_t bezeichnen die Störterme in der Beziehung (2.4.32), für die die klassischen Annahmen

$$(2.4.33) \quad E(v_t) = 0 \quad \text{für } t = 2, \dots, n$$

$$(2.4.34) \quad E(v_t \cdot v_s) = \begin{cases} \sigma_v^2 & \text{für } t = s \\ 0 & \text{sonst} \end{cases}$$

gelten soll. Die Störgrößen v_t haben danach einen Erwartungswert von 0, sind homoskedastisch mit einer Varianz von σ_v^2 und darüber hinaus nicht miteinander korreliert.

Falls der Parameter ϕ in (2.4.32) positiv ist, genauer zwischen 0 und 1 liegt, besteht eine positive Autokorrelation erster Ordnung. Störterme aufeinander folgender Beobachtungszeitpunkte haben dann tendenziell das gleiche Vorzeichen. Bei $-1 < \phi < 0$ spricht man dagegen von negativer Autokorrelation erster Ordnung, so dass Störterme benachbarter Perioden wechselnde Vorzeichen aufweisen. Werden nun die nicht beobachtbaren Störterme in (2.4.32) durch die OLS-Residuen ersetzt, hat man

$$(2.4.35) \quad \hat{u}_t = \phi \cdot \hat{u}_{t-1} + v_t, \quad |\phi| < 1$$

und damit eine Beziehung, aus der der Parameter ϕ geschätzt werden kann. Durch Anwendung der OLS-Methode ergibt sich für ϕ der Schätzer

$$\hat{\phi} = \frac{\frac{1}{n-1} \sum_{t=2}^n \hat{u}_t \hat{u}_{t-1} - \bar{\hat{u}}_0 \bar{\hat{u}}_1}{\frac{1}{n-1} \sum_{t=2}^n \hat{u}_{t-1}^2 - \bar{\hat{u}}_1^2}$$

$$\text{mit} \quad \bar{\hat{u}}_0 = \frac{1}{n-1} \sum_{t=2}^n \hat{u}_t \quad \text{und} \quad \bar{\hat{u}}_1 = \frac{1}{n-1} \sum_{t=2}^n \hat{u}_{t-1}$$

Wegen

$$\bar{\hat{u}}_0 \approx \bar{\hat{u}}_1 \approx 0$$

und

$$\sum_{t=2}^n \hat{u}_{t-1}^2 = \sum_{t=1}^{n-1} \hat{u}_t^2 \approx \sum_{t=2}^n \hat{u}_t^2$$

erhält man

$$(2.4.36) \quad \hat{\phi} \approx \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2}$$

was der Definition (2.4.31) des empirischen Autokorrelationskoeffizienten erster Ordnung, $\hat{\rho}_1$, entspricht.

Äquivalent zur Nullhypothese fehlender Autokorrelation erster Ordnung ($\rho_1=0$) ist daher die Formulierung

$$(2.4.37) \quad H_0 : \phi = 0 .$$

Sofern die Nullhypothese zutrifft, folgt aus dem Ansatz (2.4.32), dass $u_t=v_t$ ist, so dass die Störterme u_t die klassischen Eigenschaften im multiplen Regressionsmodell erfüllen.

Für den **Test auf Autokorrelation erster Ordnung** wird hier zunächst der Durbin-Watson-Test (Durbin und Watson 1950, 1951) vorgestellt, der inzwischen zum Standardinstrumentarium der Ökonometrie gehört. Die als Durbin-Watson-Statistik DW bekannte Prüfgröße basiert auf den OLS-Residuen und ist durch

$$(2.4.38) \quad DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}$$

gegeben. Da im Zähler und Nenner jeweils quadrierte Summanden stehen, ist die Durbin-Watson-Statistik stets nicht negativ ($DW \geq 0$). Der Zähler in (2.4.38) lässt sich noch umformen, was sich insbesondere im Hinblick auf die Interpretation der Testgröße empfiehlt. Durch Auflösen des Binoms erhält man zunächst

$$\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2 = \sum_{t=2}^n (\hat{u}_t^2 - 2\hat{u}_t \hat{u}_{t-1} + \hat{u}_{t-1}^2) = \sum_{t=2}^n \hat{u}_t^2 + \sum_{t=2}^n \hat{u}_{t-1}^2 - 2 \sum_{t=2}^n \hat{u}_t \hat{u}_{t-1} .$$

Sofern die Anzahl der Beobachtungen n groß genug ist, gelten die Approximationen

$$\sum_{t=2}^n \hat{u}_t^2 \approx \sum_{t=2}^n \hat{u}_{t-1}^2 \approx \sum_{t=1}^n \hat{u}_t^2 ,$$

so dass sich

$$\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2 \approx 2 \sum_{t=1}^n \hat{u}_t^2 - 2 \sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}$$

ergibt. Durch Einsetzen dieser Beziehung in (2.4.38) erhält man

$$DW \approx \left(2 \sum_{t=1}^n \hat{u}_t^2 - 2 \sum_{t=2}^n \hat{u}_t \hat{u}_{t-1} \right) / \sum_{t=1}^n \hat{u}_t^2 = 2 \left(1 - \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2} \right)$$

und unter Berücksichtigung von (2.4.36) schließlich

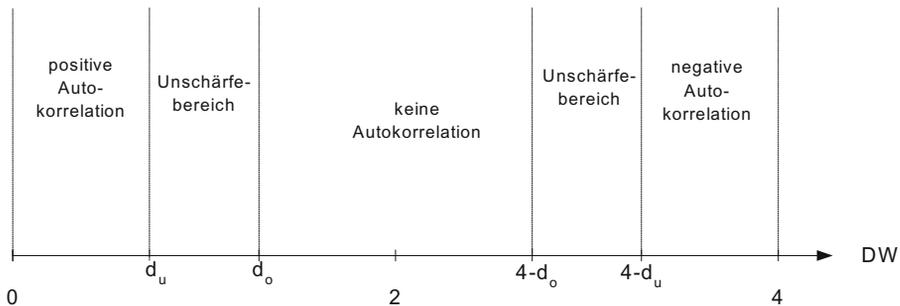
$$(2.4.39) \quad DW \approx 2(1 - \hat{\phi}) .$$

Damit ist der Zusammenhang zwischen der Durbin-Watson-Statistik und dem empirischen Autokorrelationskoeffizienten erster Ordnung der Residuen, $\hat{\rho}_1$, hergestellt. Wegen $|\hat{\rho}_1| < 1$ nimmt die Durbin-Watson-Statistik stets Werte im Intervall $(0, 4)$ an.

Bei Abwesenheit einer Autokorrelation erster Ordnung ($\rho_1 = 0$) werden der empirische Autokorrelationskoeffizient $\hat{\rho}_1$ und mithin der OLS-Schätzer $\hat{\phi}$ in der Nähe von 0 liegen, so dass die Durbin-Watson-Statistik in diesem Fall Werte annimmt, die sich nicht wesentlich vom theoretischen Wert 2 unterscheiden. Im Grenzfall einer perfekt positiven Autokorrelation erster Ordnung ($\hat{\rho}_1 = 1$), wird der berechnete Wert der DW-Statistik dagegen approximativ gleich 0 sein, während sich bei perfekt negativer Autokorrelation erster Ordnung ($\hat{\rho}_1 = -1$) wegen (2.4.39) ein Wert von 4 ergibt.

Was im Rahmen des Durbin-Watson-Tests noch zu tun bleibt, ist, die An- und Ablehnbereiche für die Hypothese $H_0 : \rho_1 = 0$ genauer festzulegen, mithin die Verteilung der Prüfgröße zu spezifizieren. Dazu wird vorausgesetzt, dass die Störvariablen u_t unter der Nullhypothese unabhängige, identisch normalverteilte Zufallsvariablen sind. Die erklärenden Variablen sind dagegen feste gegebene Größen. Die Verteilung der Durbin-Watson-Statistik hängt bei unterstellter Gültigkeit der Nullhypothese jedoch entscheidend von der Beobachtungsmatrix der Regressoren \mathbf{X} ab. Sie wird nicht nur von der Anzahl der exogenen Variablen und vom Stichprobenumfang, sondern auch von den realisierten Variablenwerten beeinflusst, so dass sich für jede Anwendung eine andere Verteilung ergibt. Um dieses Problem zu vermeiden, wird eine unter den angegebenen Vorausset-

zungen universell einsetzbare Verteilung angegeben. Sie ist jedoch mit dem Nachteil einer gewissen Unschärfe bei der Tabellierung der kritischen Werte verbunden, da diese unabhängig von den Variableninhalten festgelegt werden müssen. Daher ergeben sich Unbestimmtheitsbereiche, die zwischen einem unteren und einem oberen kritischen Wert (d_u und d_o) liegen. Die für die Testentscheidung relevanten Bereiche der Durbin-Watson-Statistik lassen sich genauer auf einer Skala anordnen:



Die Nullhypothese fehlender Autokorrelation erster Ordnung lässt sich danach nicht verwerfen, wenn der berechnete Wert der Durbin-Watson-Statistik im Intervall $(d_o, 4-d_o)$ liegt. Sofern die Prüfgröße Werte im Intervall $[0, d_u)$ annimmt, besteht eine positive, bei Werten im Intervall $[4-d_u, 4)$ eine negative Autokorrelation erster Ordnung. In den Unschärfebereichen $[d_u, d_o]$ und $[4-d_o, 4-d_u]$ lässt sich dagegen keine Entscheidung treffen. Die kritischen Werte d_u und d_o sind in Abhängigkeit von der Anzahl der Beobachtungen und der exogenen Variablen sowie von der gewählten Irrtumswahrscheinlichkeit α ($\alpha=0,05$ oder $\alpha=0,01$) im Anhang tabelliert.

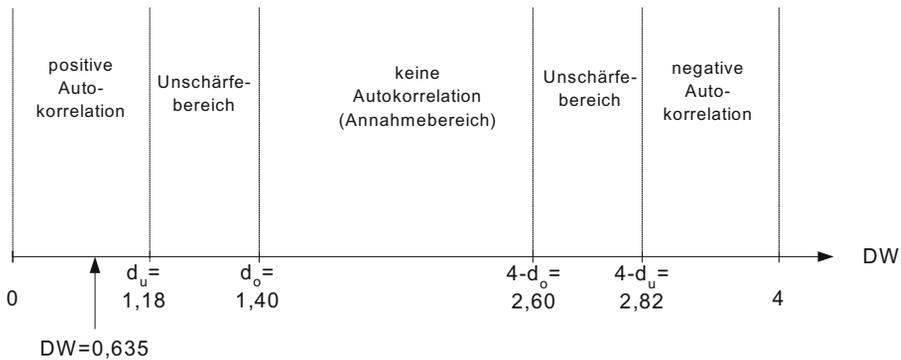
Beispiel 2.4. 4: Der Durbin-Watson-Test soll unter Verwendung der Konsumfunktion (2.1.28) illustriert werden. Damit lässt sich beurteilen, ob die Annahme einer fehlenden Autokorrelation der Störvariablen der Konsumfunktion berechtigt ist. Genauer wird die Störvariable u_t mit dem Durbin-Watson-Test auf Autokorrelation erster Ordnung getestet. Als Prüfgröße dieses Tests wird die Durbin-Watson-Statistik DW unter Verwendung der OLS-Residuen \hat{u}_t bestimmt. Die hierzu benötigten Abweichungsquadratsummen werden in der folgenden Arbeitstabelle berechnet:

t	\hat{u}_t	\hat{u}_{t-1}	\hat{u}_t^2	$(\hat{u}_t - \hat{u}_{t-1})^2$
1	-31,72180	–	1006,272	–
2	-28,69755	-31,72180	823,5494	9,14607
3	-19,49951	-28,69755	380,2308	84,60397
4	-3,82272	-19,49951	14,6132	245,7618
5	-10,76985	-3,82272	115,9897	48,26270
6	-11,67601	-10,76985	136,3291	0,82111
7	11,30838	-11,67601	127,8795	528,2820
8	22,72315	11,30838	516,3413	130,2969
9	23,51412	22,72315	552,9139	0,625642
10	30,00490	23,51412	900,2942	41,13026
11	33,57936	30,00490	1127,573	12,77671
12	31,34366	33,57936	982,4251	4,99833
13	-7,07607	31,34366	50,0708	1476,076
14	-1,48596	-7,07607	2,2081	31,24939
15	-23,56504	-1,48596	555,3112	487,4860
16	-13,83269	-23,56504	191,3434	94,71860
17	-34,29186	-13,83269	1175,931	418,5773
18	12,56193	-34,29186	157,8022	2195,278
19	21,40355	12,56193	458,1121	78,17424
Σ			9275,191	5889,265

Die Durbin-Watson-Statistik nimmt somit den Wert

$$DW = \frac{\sum_{t=2}^{19} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{19} \hat{u}_t^2} = \frac{5889,265}{9275,191} = 0,635$$

an. Die kritischen Werte des Durbin-Watson-Testes für ein Signifikanzniveau von 5 % bei einem Stichprobenumfang von $n=19$ und $k=2$ Regressoren (einschl. der Scheinvariablen) gehen aus der folgenden Skizze hervor:



In unserem Beispiel fällt die Durbin-Watson-Statistik DW in den linken Bereich,

$$DW = 0,635 \leftarrow [0; d_u = 1,18],$$

was bedeutet, dass aus dem Testergebnis auf eine Autokorrelation der OLS-Residuen der Keynes'schen Konsumfunktion geschlossen werden kann. Der Durbin-Watson-Test indizierte, dass der private Konsum trotz des hohen Erklärungsgehalts nicht allein auf das verfügbare Einkommen zurückzuführen ist. In den OLS-Residuen bleibt eine systematische Variation bestehen, die auf zusätzliche ökonomische Einflüsse hinweisen, die z.B. durch Stimmungsindikatoren erfasst werden könnten.³⁰ ♦

2.4.3.2 Breusch-Godfrey-Test und Ljung-Box-Test

Ein Test auf Autokorrelation nicht nur erster, sondern beliebiger Ordnung ist z.B. der Breusch-Godfrey-Test (Breusch, 1978; Godfrey, 1978). Der Test ist auch bei stochastischen Regressoren einsetzbar. Ausgangspunkt ist das multiple Regressionsmodell

$$(2.4.40) \quad y_t = \mathbf{x}'_t \cdot \boldsymbol{\beta} + u_t,$$

das um m verzögerte endogene Variablen $y_{t-1}, y_{t-2}, \dots, y_{t-m}$ erweitert werden kann:

$$(2.4.41) \quad y_t = \mathbf{y}'_{-t} \cdot \boldsymbol{\gamma} + \mathbf{x}'_t \cdot \boldsymbol{\beta} + u_t$$

mit $\mathbf{y}'_{-t} = (y_{t-1} \ y_{t-2} \ \dots \ y_{t-m})$ und $\boldsymbol{\gamma}' = (\gamma_1 \ \gamma_2 \ \dots \ \gamma_m)$.

³⁰ Zur Operationalisierung von Stimmungsindikatoren in Form eines Konsumklimaindex und ihrer Verwendung in einer makroökonomischen Konsumfunktion siehe Kosfeld (1996).

Das erweiterte Regressionsmodell (2.4.41) enthält neben den k nicht-stochastischen Erklärungsgrößen (einschl. der Scheinvariablen) x_1, x_2, \dots, x_k m stochastische Regressoren in Form der verzögerten endogenen Variablen. Für die Störvariable u_t in (2.4.40) bzw. (2.4.41) wird ein autoregressiver Prozess p -ter Ordnung zugelassen,

$$(2.4.42) \quad u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + \alpha_p u_{t-p} + v_t,$$

worin v_t einem White-Noise-Prozess folgt. Hierdurch kann eine komplexe Autokorrelationsstruktur der Störgröße eingefangen werden. In der Nullhypothese wird die Abwesenheit von Autokorrelation unterstellt, was die Geltung der Parameterrestriktionen $\alpha_1 = \dots = \alpha_p = 0$ impliziert:

$$H_0: \alpha_1 = \dots = \alpha_p = 0.$$

Dann gilt $u_t = v_t$, so dass die Störvariable in (2.4.40) bzw. (2.4.41) die Eigenschaft eines reinen Zufallsprozesses erfüllt.

Der Breusch-Godfrey-Test auf Autokorrelation basiert auf dem allgemeinen Lagrange-Multiplier-Prinzip, das nur eine Schätzung des restringierten Modells erforderlich macht. Dies hat zur Folge, dass sich daraus eine einfache Teststatistik ableiten lässt. Danach wird im Anschluss an die Schätzung des Modells (2.4.40) bzw. (2.4.41) die Regression

$$(2.4.43) \quad \hat{u}_t = \mathbf{z}_t' \boldsymbol{\delta} + \alpha_1 \hat{u}_{t-1} + \dots + \alpha_p \hat{u}_{t-p} + w_t$$

mit w_t als Störterm durchgeführt. Die Residuen \hat{u}_t werden mithin durch die bis zu einem Lag p zeitlich verzögerten Residuen und einen Variablenvektor \mathbf{z}_t erklärt. Während die Vektoren \mathbf{z}_t und $\boldsymbol{\delta}$ bei Verwendung des Regressionsmodells (2.4.40) mit den Vektoren \mathbf{x}_t und $\boldsymbol{\beta}$ identisch sind, enthalten sie bei einer ökonometrischen Schätzung auf der Basis des Modells (2.4.41) die nicht-stochastischen und stochastischen Regressoren bzw. deren Parameter:

$$\mathbf{z}_t' = \begin{pmatrix} \mathbf{y}'_{-t} & \mathbf{x}'_t \end{pmatrix} \quad \text{und} \quad \boldsymbol{\delta}' = \begin{pmatrix} \boldsymbol{\gamma}' & \boldsymbol{\beta}' \end{pmatrix}.$$

Die Teststatistik basiert auf dem unkorrigierten Bestimmtheitsmaß der Hilfsregression (2.4.43), das mit der Anzahl der Beobachtungen multipliziert wird. Die Prüfgröße

$$(2.4.44) \quad BG = n \cdot R^2$$

ist unter der Nullhypothese fehlender Autokorrelation p -ter Ordnung asymptotisch χ^2 -verteilt mit p Freiheitsgraden.

Der Breusch-Godfrey-Test ist analog einsetzbar, wenn die Autokorrelationsstruktur nicht durch einen autoregressiven Prozess p -ter Ordnung, sondern durch einen so genannten Moving-Average-Prozess der Ordnung q beschreibbar ist. In diesem Fall wird anstelle von (2.4.42) die Spezifikation

$$(2.4.45) \quad u_t = v_t - \lambda_1 v_{t-1} - \lambda_2 v_{t-2} - \dots - \lambda_q v_{t-q}$$

unterstellt. Dagegen ist bei Störtermen, die über einen autoregressiven Moving-Average-Prozess der Ordnung p und q korreliert sind,

$$(2.4.46) \quad u_t - \alpha_1 u_{t-1} - \alpha_2 u_{t-2} - \dots - \alpha_p u_{t-p} = v_t - \lambda_1 v_{t-1} - \lambda_2 v_{t-2} - \dots - \lambda_q v_{t-q}$$

der Ljung-Box-Test (Ljung und Box, 1978) anwendbar. Beim Test auf Autokorrelation der Ordnung $k = \max\{p, q\}$ wird die Prüfgröße

$$(2.4.47) \quad Q_k = n \sum_{j=1}^k \frac{n+2}{n-j} r_j^2$$

verwendet, die auf den empirischen Autokorrelationskoeffizienten r_j der Residuen des Modells (2.4.40) bzw. (2.4.41) basiert. Die Prüfgröße ist unter der Nullhypothese fehlender Autokorrelation k -ter Ordnung χ^2 -verteilt mit k Freiheitsgraden. Die Q -Statistik kann analog bei autoregressiven oder Moving-Average-Störprozessen eingesetzt werden. Bei einer Ablehnung der Nullhypothese ist allerdings nicht klar, über welche Art des Prozesses sich die Autokorrelation vollzieht.

Beispiel 2.4. 5: Die Nullhypothese der Abwesenheit einer Autokorrelation erster Ordnung der Störgröße der Keynes'schen Konsumfunktion (2.1.28) lässt sich auch mit dem Breusch-Godfrey-Test untersuchen. Dazu wird die Hilfsregression

$$\hat{u}_t = \alpha_0 + \alpha_1 \cdot \hat{u}_{t-1} + \alpha_2 \cdot Y_t^V$$

durchgeführt, die ein Bestimmtheitsmaß von 0,390 aufweist. Bei 19 Beobachtungen ergibt sich ein Wert der Prüfstatistik BG von 7,410, der nicht nur das 95%-Quantil $\chi_{1;0,95}^2 = 3,84$, sondern ebenfalls das 99%-Quantil $\chi_{1;0,99}^2 = 6,63$ überschreitet. Der Breusch-Godfrey-Test bestätigt damit das Ergebnis des Durbin-Watson-Tests einer Autokorrelation erster Ordnung der Störgröße der Keynes'schen Konsumfunktion. ♦

2.4.4 Verallgemeinerte Methode der kleinsten Quadrate (Generalized Least Squares)

2.4.4.1 GLS-Schätzung bei bekannter Kovarianzmatrix der Störterme

Bei einer Anwendung der verallgemeinerten Methode der kleinsten Quadrate (generalized least squares, GLS) ³¹ gehen wir von der Form

$$\text{Cov}(\mathbf{u}) = \sigma^2 \mathbf{\Omega}$$

der Kovarianzmatrix der Störterme aus. Um die Grundstruktur des verallgemeinerten Kleinst-Quadrate-Schätzers im linearen Regressionsmodell transparenter zu machen, wird hier zunächst einmal unterstellt, dass die $n \times n$ -Matrix $\mathbf{\Omega}$ bekannt ist.

Wenn $\mathbf{\Omega}$ symmetrisch und positiv definit ist ³², gilt dies ebenfalls für ihre Inverse $\mathbf{\Omega}^{-1}$, die sich dann wie folgt faktorisieren lässt:

$$(2.4.43) \quad \mathbf{\Omega}^{-1} = \mathbf{T}'\mathbf{T}$$

\mathbf{T} ist darin eine reguläre Matrix der Ordnung n . Wird nun das ökonometrische Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

von links mit der Matrix \mathbf{T} , die zunächst als bekannt unterstellt wird, multipliziert, folgt

$$(2.4.44) \quad \mathbf{T}\mathbf{y} = \mathbf{T}\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\mathbf{u} ,$$

woraus man mit den Definitionen

$$(2.4.45) \quad \mathbf{T}\mathbf{y} = \mathbf{y}^* , \quad \mathbf{T}\mathbf{X} = \mathbf{X}^* \quad \text{und} \quad \mathbf{T}\mathbf{u} = \mathbf{u}^*$$

das transformierte Regressionsmodell

$$(2.4.46) \quad \mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^*$$

erhält. In dem Modell (2.4.46) erfüllen die Störvariablen erneut die klassischen Eigenschaften

$$(2.4.47) \quad E(\mathbf{u}^*) = \mathbf{0}$$

$$(2.4.48) \quad E(\mathbf{u}^* \mathbf{u}^{*\prime}) = \sigma^2 \mathbf{I} ,$$

³¹ Die verallgemeinerte Methode der kleinsten Quadrate geht auf Aitken (1935) zurück.

³² Eine $n \times n$ -Matrix \mathbf{A} ist positiv definit, wenn ihre quadratische Form positiv ist, d.h., wenn für alle $n \times 1$ -Vektoren $\mathbf{x} \neq \mathbf{0}$ ist, muss $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ gelten.

was leicht gezeigt werden kann. Dass der Erwartungswert der transformierten Störgröße \mathbf{u}^* in jeder Beobachtungsperiode gleich Null ist, ergibt sich aus

$$E(\mathbf{u}^*) = E(\mathbf{T}\mathbf{u}) = \mathbf{T}E(\mathbf{u}) = \mathbf{0} .$$

Die Linearität des Erwartungsoperators kann hier ausgenutzt werden, weil die Matrix \mathbf{T} als bekannt vorausgesetzt wird, also keine stochastischen Größen enthält. Die Kovarianzmatrix, die unter Berücksichtigung von (2.4.47) durch

$$\text{Cov}(\mathbf{u}^*) = E(\mathbf{u}^* \mathbf{u}^{*'})$$

gegeben ist, hat wegen $\mathbf{u}^* = \mathbf{T}\mathbf{u}$ die Form

$$\text{Cov}(\mathbf{u}^*) = E(\mathbf{T}\mathbf{u}(\mathbf{T}\mathbf{u})') = E(\mathbf{T}\mathbf{u}\mathbf{u}'\mathbf{T}) = \mathbf{T}E(\mathbf{u}\mathbf{u}')\mathbf{T} .$$

Da die Varianz-Kovarianz-Matrix der originären Störgröße \mathbf{u} durch $\sigma^2\mathbf{\Omega}$ gegeben ist, folgt

$$\text{Cov}(\mathbf{u}^*) = \mathbf{T}\sigma^2\mathbf{\Omega}\mathbf{T}' = \sigma^2\mathbf{T}\mathbf{\Omega}\mathbf{\Omega}' .$$

Wegen (2.4.43) gilt nun ³³

$$\mathbf{\Omega} = (\mathbf{T}'\mathbf{T})^{-1} = \mathbf{T}^{-1}(\mathbf{T}')^{-1}$$

womit man

$$\text{Cov}(\mathbf{u}^*) = \sigma^2\mathbf{T}\mathbf{T}^{-1}(\mathbf{T}')^{-1}\mathbf{T}'$$

erhält. Wegen

$$\mathbf{T}\mathbf{T}^{-1} = \mathbf{I} \quad \text{und} \quad (\mathbf{T}')^{-1}\mathbf{T}' = \mathbf{I}$$

ergibt sich schließlich

$$\text{Cov}(\mathbf{u}^*) = \sigma^2\mathbf{I} .$$

Wie man sieht, wird die Matrix \mathbf{T} so festgelegt, dass $\mathbf{T}\mathbf{\Omega}\mathbf{T}' = \mathbf{I}$ gilt. Damit ist gezeigt, dass die transformierte Störgröße \mathbf{u}^* die klassischen Eigenschaften besitzt.

Folglich lässt sich der Parametervektor $\boldsymbol{\beta}$ im transformierten Modell (2.4.46) durch die OLS-Methode schätzen. Es handelt sich dabei dann jedoch um einen verallgemeinerten Kleinst-Quadrate-Schätzer (generalized least squares=GLS), der von der Form

$$(2.4.49) \quad \hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^*$$

³³ Die Berechnung der Inversen erfolgt nach der Regel $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

ist. Da die \mathbf{X}^* -Matrix nur feste Größen enthält, ist er linear in den transformierten Werten der endogenen Variablen.

Bei der Bestimmung des GLS-Schätzers $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ braucht jedoch keinesfalls auf die Transformationsmatrix \mathbf{T} zurückgegriffen werden. Vielmehr lässt er sich unmittelbar aus

$$(2.4.50) \quad \hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

bestimmen, worin allein die Inverse der bis auf den Skalar σ^2 gegebenen Kovarianzmatrix $\boldsymbol{\Omega}$ eingeht. Die Gestalt des GLS-Schätzers (2.4.50) erhält man durch Anwendung elementarer Matrizenoperationen aus der Schätzfunktion (2.4.49):

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{GLS}} &= (\mathbf{X}^*{}'\mathbf{X}^*)^{-1} \mathbf{X}^*{}'\mathbf{y}^* = [(\mathbf{TX})'(\mathbf{TX})]^{-1} (\mathbf{TX})'\mathbf{T}\mathbf{y} \\ &= (\mathbf{X}'\mathbf{T}'\mathbf{TX})^{-1} \mathbf{X}'\mathbf{T}'\mathbf{T}\mathbf{y} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}. \end{aligned}$$

Für inferenzstatistische Auswertungen benötigt man noch die Kovarianzmatrix von $\hat{\boldsymbol{\beta}}_{\text{GLS}}$. Ausgangspunkt ist dabei die analog zum klassischen multiplen Regressionsmodell geltende Beziehung,

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma^2 (\mathbf{X}^*{}'\mathbf{X}^*)^{-1},$$

für die transformierten Modellvariablen. Setzt man hierin \mathbf{TX} für \mathbf{X}^* ein, so erhält man

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma^2 (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1},$$

worin der Parameter σ^2 allerdings unbekannt ist. Er kann jedoch im verallgemeinerten Regressionsmodell analog wie beim klassischen Regressionsmodell durch

$$\hat{\sigma}_{\text{GLS}}^2 = \frac{\hat{\mathbf{u}}^*{}'\hat{\mathbf{u}}^*}{n-k}$$

erwartungstreu geschätzt werden. Die Größen $\hat{\mathbf{u}}^*$, die durch

$$\hat{\mathbf{u}}^* = \mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}_{\text{GLS}}$$

gegeben sind, heißen GLS-Residuen. Der GLS-Schätzer für σ^2 ergibt sich mithin aus

$$\begin{aligned}\hat{\sigma}_{\text{GLS}}^2 &= \frac{(\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}_{\text{GLS}})' (\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}_{\text{GLS}})}{n - k} \\ &= \frac{(\mathbf{T}\mathbf{y} - \mathbf{T}\mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})' (\mathbf{T}\mathbf{y} - \mathbf{T}\mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})}{n - k} \\ &= \frac{(\bar{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})' \mathbf{T}'\mathbf{T}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})}{n - k},\end{aligned}$$

woraus sich wegen $\mathbf{T}'\mathbf{T} = \boldsymbol{\Omega}^{-1}$

$$\hat{\sigma}_{\text{GLS}}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})}{n - k}$$

folgt. Damit wird klar, dass die Transformationsmatrix \mathbf{T} bei der GLS-Schätzung selbst nicht bestimmt zu werden braucht. Vielmehr hat sie die Funktion einer Hilfsgröße. Die GLS-Schätzung kann daher unter Verwendung der originären Beobachtungswerte durchgeführt werden, die hierzu nicht transformiert zu werden brauchen. In der ökonomischen Praxis ist jedoch zusätzlich die Kovarianzmatrix $\boldsymbol{\Omega}$ zu schätzen, womit wir uns nun beschäftigen wollen.

2.4.4.2 GLS-Schätzung bei unbekannter Kovarianzmatrix der Störterme

Bei empirischen Anwendungen ist die $\boldsymbol{\Omega}$ -Matrix im Allgemeinen unbekannt. Damit ist vor der GLS-Schätzung des Modells ein konsistenter Schätzer $\hat{\boldsymbol{\Omega}}$ zu entwickeln. Hat man $\hat{\boldsymbol{\Omega}}$ gefunden, sind alle interessierenden Größen durch

$$(2.4.51) \quad \hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y},$$

$$(2.4.52) \quad \text{Cov}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma_{\text{GLS}}^2 (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1} \text{ und}$$

$$(2.4.53) \quad \sigma_{\text{GLS}}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})' \hat{\boldsymbol{\Omega}}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}})}{n - k}$$

berechenbar. Er ist mit den Gleichungen (2.4.51), (2.4.52) und (2.4.53) ist die GLS-Schätzung durchführbar (flexible GLS estimators). Als Konsequenz einer unbekannt Matrix $\boldsymbol{\Omega}$ ergibt sich zunächst, dass $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ kein linearer Schätzer mehr ist. Da $\hat{\boldsymbol{\Omega}}$ im Allgemeinen eine Zufallsgröße ist, sind auch die Gewichte der Linearkombination Zufallsvariablen. Aus dem gleichen Grund existiert nun ein Erwartungswert von $\hat{\boldsymbol{\Omega}}$, so

dass $\hat{\beta}_{GLS}$ auch die Eigenschaft der Unverzerrtheit verliert. Allgemein bleibt so lediglich die Güteeigenschaft der Konsistenz erhalten.

Die Ω -Matrix enthält bei Heteroskedastizität und Autokorrelation bis zu $n(n-1)/2$ verschiedene Parameter, was bereits oben angedeutet wurde. Bei nur n Beobachtungen lässt sich Ω aufgrund der fehlenden Freiheitsgrade nicht schätzen. Daher ist hier zu versuchen, durch die Anwendung von Autokorrelations- und Heteroskedastizitätstests die Struktur der Kovarianzmatrix der Störterme aufzudecken, um so die Anzahl der zu schätzenden Parameter zu reduzieren. Das bedeutet auf der anderen Seite, dass kein allgemeines Verfahren zur Schätzung der Ω -Matrix angegeben werden kann. Vielmehr wird man sich von der konkreten Form des Modelldefektes leiten lassen, um so einen Schätzer für Ω zu erhalten. Die anzuwendende Vorgehensweise wird daher im Folgenden separat für Autokorrelation und Heteroskedastizität diskutiert.

2.4.4.2.1 Kovarianzmatrix bei Heteroskedastizität

Bei Heteroskedastizität, aber fehlender Autokorrelation ist $\text{Cov}(\mathbf{u})$ allgemein durch

$$\text{Cov}(\mathbf{u}) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

gegeben. Es handelt sich um eine Diagonalmatrix, deren Diagonalelemente sich allgemein voneinander unterscheiden.

Ansatz 1: Gruppenweise Heteroskedastizität

Das Modell wird nun unter Umständen schätzbar, wenn einige der Varianzen σ_t^2 , $t = 1, \dots, n$, identisch sind. Sofern nur m Varianzen verschieden sind, $m < n$, hat man

$$\text{Cov}(\mathbf{u}) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \sigma_m^2 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_m^2 \end{pmatrix},$$

wobei für $m=2$ der Goldfeld-Quandt-Test Hinweise auf eine derartige Spezifikation liefern kann. Dieser Ansatz zur Parameterreduktion lässt sich z.B. bei Zeitreihendaten begründen, wenn die Varianz der Störterme etwa bei wachsenden Werten der erklärenden Variablen steigt, in bestimmten Teilintervallen des Beobachtungszeitraums jedoch näherungsweise konstant ist. Die m verschiedenen Varianzen sind unter der vorgeschlagenen Strategie aus den Residuenquadratsummen von m separat durchgeführten Regressionen zu ermitteln, wodurch sich ein Schätzer für die unbekannte Ω -Matrix ergibt. Ein klarer Nachteil dieses Verfahrens besteht allerdings darin, dass relativ große Beobachtungsumfänge benötigt werden. So muss in jeder der m Regressionen die Anzahl der Beobachtungen den Wert $k+1$ übersteigen, damit die Durchführung der Schätzung gewährleistet ist. Dieser Einwand macht die Strategie nur für relativ kleine Werte von m , etwa $m=2$, praktikabel.

Ansatz 2: Heteroskedastizität durch exogene Variable

Alternativ lässt sich z.B. ein Modell zur Bestimmung der n Varianzen σ_t^2 formulieren. Ein derartiger Ansatz ist in

$$(2.4.54) \quad \sigma_t^2 = \sigma^2 z_t^\delta, \quad t = 1, \dots, n$$

gegeben. Die z_t sind darin die n Beobachtungen eines Regressors oder alternativ eine Linearkombination der Werte mehrerer Variablen, in jedem Fall aber exogen gegeben. Damit sind nur noch die Parameter σ^2 und δ unbekannt. Da die Störtermvarianzen jedoch nicht-negativ sein müssen und $z_t \leq 0$ ist, wird hier $\delta=2$ gesetzt. Dann ist in der Spezifikation

$$(2.4.55) \quad \sigma_t^2 = \sigma^2 z_t^2, \quad t = 1, \dots, n$$

nur noch σ^2 zu bestimmen. Die Validität dieses Ansatzes kann empirisch etwa mit dem Breusch-Pagan-Test auf Heteroskedastizität überprüft werden. Unter der Annahme (2.4.55) ist die Kovarianzmatrix der Störvariablen $\text{Cov}(\mathbf{u})$ durch

$$\text{Cov}(\mathbf{u}) = \sigma^2 \mathbf{\Omega} = \begin{pmatrix} \sigma^2 z_1^2 & 0 & \dots & 0 \\ 0 & \sigma^2 z_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma^2 z_n^2 \end{pmatrix} = \sigma^2 \cdot \begin{pmatrix} z_1^2 & 0 & \dots & 0 \\ 0 & z_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_n^2 \end{pmatrix}$$

gegeben. Aus

$$\mathbf{\Omega} = \begin{pmatrix} z_1^2 & 0 & \cdots & 0 \\ 0 & z_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_n^2 \end{pmatrix}$$

kann auf einfache Art und Weise die Transformationsmatrix \mathbf{T} bestimmt werden. Da die Inverse von $\mathbf{\Omega}$,

$$\mathbf{\Omega}^{-1} = \begin{pmatrix} 1/z_1^2 & 0 & \cdots & 0 \\ 0 & 1/z_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/z_n^2 \end{pmatrix}$$

aufgrund von (2.4.43) durch die Produktmatrix $\mathbf{T}'\mathbf{T}$ gegeben ist, folgt unmittelbar

$$\mathbf{T} = \begin{pmatrix} 1/z_1 & 0 & \cdots & 0 \\ 0 & 1/z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/z_n \end{pmatrix}.$$

Sofern die z -Werte durch die Beobachtungen einer einzigen exogenen Variablen z.B. x_2 gegeben sind, spezifiziert sich die Transformationsmatrix \mathbf{T} zu

$$\mathbf{T} = \begin{pmatrix} 1/x_{21} & 0 & \cdots & 0 \\ 0 & 1/x_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/x_{2n} \end{pmatrix}.$$

Man erhält dann einen GLS-Schätzer für $\boldsymbol{\beta}$, indem man die Variablen der Modellgleichung

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t$$

mit den inversen Beobachtungswerten $1/x_{2t}$ multipliziert und das transformierte Modell

$$(2.4.56) \quad \frac{y_t}{x_{2t}} = \beta_1 \frac{1}{x_{2t}} + \beta_2 \frac{x_{1t}}{x_{2t}} + \beta_3 \frac{x_{3t}}{x_{2t}} + \dots + \beta_k \frac{x_{kt}}{x_{2t}} + \frac{u_t}{x_{2t}}$$

$$\begin{array}{ccccccc} \uparrow & \uparrow & \uparrow & \uparrow & & \uparrow & \uparrow \\ y_t^* & x_{1t}^* & x_{2t}^* = 1 & x_{3t}^* & & x_{kt}^* & u_t^* \end{array}$$

mit der OLS-Methode schätzt. Dass die transformierte Störvariable u_t^* tatsächlich homoskedastisch ist, lässt sich nun leicht zeigen. Aufgrund von (2.4.56) ist die Varianz von u_t^* durch

$$\text{Var}(u_t^*) = \text{Var}\left(\frac{u_t}{x_{2t}}\right) = \frac{1}{x_{2t}^2} \text{Var}(u_t) = \frac{1}{x_{2t}^2} \sigma^2 z_t^2$$

gegeben, so dass man mit $z_t = x_{1t}$

$$\text{Var}(u_t^*) = \sigma^2$$

erhält.

Die Heteroskedastizität wird nach dieser Verfahrensweise mithin durch eine Gewichtung der Modellvariablen beseitigt. Die Gewichte sind hierbei allgemein durch den Faktor $1/z_t$ gegeben. Eine Schätzung des Parametervektors β unter Verwendung der transformierten Modellvariablen

$$y_t^* = \frac{y_t}{z_t}, \quad x_{1t}^* = \frac{x_{1t}}{z_t}, \quad x_{2t}^* = \frac{x_{2t}}{z_t}, \quad \dots, \quad x_{kt}^* = \frac{x_{kt}}{z_t}, \quad u_t^* = \frac{u_t}{z_t}$$

mit der OLS-Methode wird aus diesem Grund auch als **gewichtete Kleinst-Quadrate-Schätzung** (weighted least-squares (WLS)) bezeichnet, die eine spezielle Variante der GLS-Schätzung darstellt.

Beispiel 2.4. 5: Obwohl die Nullhypothese homoskedastischer Störterme der Geldnachfragefunktion (2.1.29) im Beobachtungszeitraum 1970–1989 nicht verworfen werden konnte, soll GLS-Schätzung für den Fall der Heteroskedastizität hieran beispielhaft aufgezeigt werden. Ausgangspunkt ist hierbei die OLS-Schätzung der Geldnachfragefunktion:

$$\ln m_t = -4,741 + 1,424 \cdot \ln y_t - 0,077 \cdot \ln r_t, \quad R^2 = 0,978.$$

$$\begin{array}{ccc} (-10,900) & (24,903) & (-4,076) \end{array}$$

Zur Modellierung der Heteroskedastizität verwenden wir den Ansatz 2, nach dem die Heteroskedastizität auf eine exogene Variable zurückzuführen ist. Wählt man hierzu den Regressors $\ln y_t$, dann lässt sich die Störvarianz σ_t^2 in der Form

$$\sigma_t^2 = \sigma^2 \cdot (\ln y_t)^2$$

spezifizieren. Das Regressionsmodell der gewichteten Methode der kleinsten Quadrate (WLS-Methode) als Spezialfall der GLS-Methode lautet hier

$$(2.4.57) \quad \left(\frac{\hat{\ln m}_t}{\ln y_t} \right) = \beta_1 \cdot \frac{1}{\ln y_t} + \beta_2 + \beta_3 \cdot \frac{\ln r_t}{\ln y_t} + u_t^*$$

mit $u_t^* = u_t / \ln y_t$. OLS-Schätzung des ökonometrischen Modells (2.4.57) erbringt die GLS-Schätzer

$$\hat{\beta}_{1, \text{GLS}} = -4,071 \quad (t = -10,882)$$

$$\hat{\beta}_{2, \text{GLS}} = 1,418 \quad (t = 24,974)$$

und

$$\hat{\beta}_{3, \text{GLS}} = -0,077 \quad (t = -4,106),$$

deren Inferenzstatistik valide ist. In unserer speziellen Anwendung weichen die GLS-Schätzer allerdings nur geringfügig von den originären OLS-Schätzern ab, so dass sich hier im Hinblick auf die Interpretation der ökonometrischen Schätzergebnisse im Kern keine Unterschiede ergeben.

Zu beachten ist allerdings, dass der Determinationskoeffizient der OLS-Schätzung von $R_{\text{OLS}}^2 = 0,978$ nicht unmittelbar mit dem Bestimmtheitsmaß der WLS-Schätzung (2.4.57) von $R_{\text{GLS}^*}^2 = 0,918$ ³⁴ vergleichbar ist, da sich letzteres auf die transformierten Variablen bezieht. Man erhält ein vergleichbares Bestimmtheitsmaß der $\hat{\ln m}_{t, \text{GLS}}$ -Schätzung, indem man GLS-Regressionswerte unter Verwendung der GLS-Schätzer für die Regressionskoeffizienten ermittelt und die sich daraus ergebende Varianz der Regressionswerte auf die Gesamtvarianz der abhängigen Variablen bezieht.

So ergibt sich z.B. der GLS-Regressionswert $\hat{\ln m}_{t, \text{GLS}}$ unter Verwendung der auf sechs Dezimalstellen exakten GLS-Koeffizientenschätzer

³⁴ Der Stern (*) soll darauf hinweisen, dass das Bestimmtheitsmaß aus den transformierten Beobachtungswerten resultiert.

$$\begin{aligned}\ln \hat{m}_{t, \text{GLS}} &= -4,700680 + 1,418280 \cdot \ln y_1 - 0,077200 \cdot \ln r_1 \\ &= 5,320150.\end{aligned}$$

Führt man die entsprechenden Berechnungen für den Stützzeitraum der Regression durch, dann erhält man die Zeitreihe der Regressionswerte $\ln \hat{m}_{t, \text{GLS}}$:

t	1	2	3	4	5	6
$\ln \hat{m}_{t, \text{GLS}}$	5,320150	5,383918	5,462940	5,470343	5,486850	5,522122
t	7	8	9	10	11	12
$\ln \hat{m}_{t, \text{GLS}}$	5,610610	5,644223	5,705171	5,715505	5,702828	5,685965
t	13	14	15	16	17	18
$\ln \hat{m}_{t, \text{GLS}}$	5,694138	5,753404	5,793477	5,826263	5,869060	5,900896
t	19	20				
$\ln \hat{m}_{t, \text{GLS}}$	5,946245	5,960796				

Während die Varianz $s_{\ln m}^2$ der logarithmierten Geldmenge 0,033568 beträgt, nimmt die Varianz $s_{\ln \hat{m}, \text{GLS}}^2$ den Wert 0,032604 an. Mithin lautet das Bestimmtheitsmaß der

GLS-Schätzung für die Variablen auf originärem Messniveau

$$R_{\text{GLS}}^2 = \frac{s_{\ln \hat{m}, \text{GLS}}^2}{s_{\ln m}^2} = \frac{0,032604}{0,033568} = 0,971,$$

so dass die Güte der Anpassung der GLS-geschätzten Geldnachfragefunktion hier gegenüber R_{OLS}^2 nicht merklich reduziert wird.

2.4.4.2 Kovarianzmatrix bei Autokorrelation

Wenn nicht alle relevanten Einflussgrößen in einem ökonometrischen Eingleichungsmodell einbezogen worden sind, folgt die Störgröße keinem White-Noise-Prozess, da sie die unerklärten systematischen Variationen „auffängt“. Man spricht dann von einer autokorrelierten Störgröße, die bei einer Nichtberücksichtigung zum Fall der Heteroskedastizität vergleichbare inferenzstatistische Konsequenzen hat. Autokorrelation der Störgröße deutet auf eine Fehlspezifikation des ökonometrischen Modells hin, die sich

durch eine Einbeziehung geeigneter Einflussgrößen beheben ließe, was wo immer möglich unter wirtschaftstheoretischen Aspekten anzustreben sein sollte. Hier wird als Alternative eine Beseitigung der Autokorrelation unter Verwendung schätztechnischer Modifikationen aufgezeigt. Dabei konzentrieren wir uns auf eine Untersuchung der Autokorrelation erster Ordnung, die sich speziell durch den autoregressiven (AR)-Prozess

$$(2.4.58) \quad u_t = \phi u_{t-1} + v_t, |\phi| < 1$$

ergibt (Markov-Prozess). Die Störterme der Beziehung sind die v_t , für die die klassischen Annahmen (2.1.6) und (2.1.7) gelten. Im Spezialfall $\phi=0$ folgt $u_t=v_t$ und somit die Abwesenheit der hypothetischen Autokorrelationsstruktur.

Für die Berechnung der Momente des Prozesses u_t empfiehlt sich bei $\phi \neq 0$ von einer zu (2.4.58) äquivalenten Darstellung auszugehen, die im Folgenden zunächst abgeleitet wird. Wird (2.4.58) um eine Periode in die Vergangenheit verschoben, ergibt sich

$$u_{t-1} = \phi u_{t-2} + v_{t-1} \quad ,$$

so dass nach Einsetzen

$$u_t = \phi^2 u_{t-2} + \phi v_{t-1} + v_t$$

folgt. Es gilt nun analog

$$u_{t-2} = \phi u_{t-3} + v_{t-2} \quad ,$$

so dass man nach erneuter Substitution

$$u_t = \phi^3 u_{t-3} + \phi^2 v_{t-2} + \phi v_{t-1} + v_t$$

erhält. Allgemein folgt damit nach i Ersetzungen

$$(2.4.59) \quad u_t = \phi^{i+1} u_{t-i-1} + \phi^i v_{t-i} + \dots + \phi v_{t-1} + v_t \quad .$$

Da annahmegemäß $|\phi| < 1$ ist, geht bei $i \rightarrow \infty$ der Ausdruck $\phi^{i+1} u_{t-i-1}$ gegen 0, so dass u_t allein als Funktion von v_t und unendlicher Verzögerungen dieses Prozesses darstellbar ist:

$$u_t = v_t + \phi v_{t-1} + \phi^2 v_{t-2} + \dots$$

oder

$$(2.4.60) \quad u_t = \sum_{i=0}^{\infty} \phi^i v_{t-i} \quad .$$

Somit sind die autokorrelierten Größen u_t wegen $|\phi| < 1$ auf zwei äquivalente Weisen darstellbar: zum einen über (2.4.58) als autoregressiver Prozess erster Ordnung, d.h. durch ein AR(1)-Modell. Zweitens nach (2.4.60) durch einen Moving-Average-Prozess unendlicher Ordnung, MA(∞)-Modell, der allgemein als Linearkombination von White-Noise-Größen³⁵ definiert ist. Die Gewichte der Linearkombination werden hier durch die unbekannt Parameter ϕ^i angegeben. Man spricht in diesem Zusammenhang auch von der Invertierbarkeit des AR-Prozesses: mit $|\phi| < 1$ lässt sich der AR(1)-Prozess (2.4.58) als MA(∞)-Prozess schreiben und ist daher invertierbar.

Die Momente der u_t werden nun aus der Darstellung (2.4.60) abgeleitet. Für den Erwartungswert erhält man

$$E(u_t) = E\left(\sum_{i=0}^{\infty} \phi^i v_{t-i}\right) = \sum_{i=0}^{\infty} \phi^i E(v_{t-i}).$$

Der Erwartungswert der v_t ist jedoch wegen (2.1.6) gleich 0, so dass

$$(2.4.61) \quad E(u_t) = 0, \quad t = 1, \dots, n$$

folgt. Damit ist die Varianz der u_t gleich $E(u_t^2)$. Aufgrund von (2.4.60) erhält man

$$\text{Var}(u_t) = E\left[\left(\sum_{i=0}^{\infty} \phi^i v_{t-i}\right)^2\right] = E\left[\sum_{i=0}^{\infty} (\phi^i)^2 (v_{t-i})^2 + \sum_{i=0}^{\infty} \sum_{j=0, j \neq i}^{\infty} \phi^i \phi^j v_{t-i} v_{t-j}\right]$$

wobei im zweiten Ausdruck die Produkte der v_t mit ungleichem Index stehen. Wegen der Linearität des Erwartungsoperators folgt nun

$$\text{Var}(u_t) = \sum_{i=0}^{\infty} \phi^{2i} E(v_{t-i}^2) + \sum_{i=0}^{\infty} \sum_{j=0, j \neq i}^{\infty} \phi^i \phi^j E(v_{t-i} v_{t-j}).$$

Da die v_t die klassischen Eigenschaften erfüllen, also insbesondere aufgrund von (2.4.30) homoskedastisch mit der Varianz σ_v^2 und unkorreliert sind, hat man

$$\text{Var}(u_t) = \sum_{i=0}^{\infty} \phi^{2i} \sigma_v^2 = \sigma_v^2 \sum_{i=0}^{\infty} \phi^{2i} = \sigma_v^2 (1 + \phi^2 + \phi^4 + \dots).$$

In der Klammer steht nun eine unendliche geometrische Reihe mit dem Quotienten ϕ^2 . Da $|\phi| < 1$ ist, ist $\phi^2 < 1$, so dass der Grenzwert dieser Reihe existiert und gleich $1/(1 - \phi^2)$ ist. Damit ergibt sich

³⁵ Eine Zufallsvariable folgt einem White-Noise-Prozess (reiner Zufallsprozess), wenn sie bei konstantem Erwartungswert und konstanter Varianz frei von Autokorrelation ist.

$$(2.4.62) \quad \text{Var}(u_t) = \frac{\sigma_v^2}{1-\phi^2}; t = 1, \dots, n,$$

so dass die u_t homoskedastisch sind. Man sieht, dass die Annahme $|\phi| < 1$ hier eine endliche Varianz impliziert. Ohne diese Annahme wäre die Varianz unendlich, würde also nicht existieren.

Die Autokovarianzen der u_t berechnen sich wegen (2.4.61) durch $\text{Cov}(u_t, u_{t-j}) = E(u_t \cdot u_{t-j})$. Damit erhält man z.B. für die Autokovarianz zwischen den Störtermen im verallgemeinerten Modell, die im Beobachtungszeitraum unmittelbar aufeinander folgen,

$$\text{Cov}(u_t, u_{t-1}) = E(u_t u_{t-1}).$$

Durch Einsetzen von (2.4.60) ergibt sich daraus

$$\begin{aligned} \text{Cov}(u_t, u_{t-1}) &= E \left[\left(\sum_{i=0}^{\infty} \phi^i v_{t-i} \right) \left(\sum_{i=0}^{\infty} \phi^i v_{t-1-i} \right) \right] \\ &= E \left[v_t + \phi \cdot (v_{t-1} + \phi \cdot v_{t-2} + \dots) \right] \left[v_{t-1} + \phi \cdot v_{t-2} + \dots \right] \\ &= E \left[v_t (v_{t-1} + \phi \cdot v_{t-2} + \dots) + \phi \cdot (v_{t-1} + \phi \cdot v_{t-2} + \dots)^2 \right] \\ &= E \left[v_t (v_{t-1} + \phi \cdot v_{t-2} + \dots) \right] + \phi \cdot E \left[(v_{t-1} + \phi \cdot v_{t-2} + \dots)^2 \right] \end{aligned}$$

Offensichtlich ist $E[v_t (v_{t-1} + \phi v_{t-2} + \dots)] = 0$, was nach Ausmultiplizieren und der Berechnung des Erwartungswerts für die Terme $v_t v_s$, $s \neq t$, unmittelbar aus (2.4.30) folgt.

Damit erhält man

$$\text{Cov}(u_t, u_{t-1}) = \phi E \left[(v_{t-1} + \phi v_{t-2} + \dots)^2 \right] = \phi E \left(\sum_{i=0}^{\infty} \phi^i v_{t-i-1} \right)^2.$$

Wird die Quadrierung durchgeführt und der Erwartungsoperator angewendet, sind alle $E(v_t v_s)$ mit $t \neq s$ gleich 0, da die v_t wegen (2.4.30) unkorreliert sind. Damit ergibt sich

$$\text{Cov}(u_t, u_{t-1}) = \phi \sum_{i=0}^{\infty} \phi^{2i} E(v_{t-i-1}^2).$$

Wegen (2.1.7) ist nun $E(v_t^2) = \sigma_v^2$ für alle t , so dass

$$\text{Cov}(u_t, u_{t-1}) = \phi \sum_{i=0}^{\infty} \phi^{2i} \sigma_v^2 = \phi \sigma_v^2 \sum_{i=0}^{\infty} \phi^{2i}$$

folgt. Analog zur Herleitung von (2.4.62) erhält man daraus

$$(2.4.63) \quad \text{Cov}(u_t, u_{t-1}) = \phi \cdot \frac{\sigma_v^2}{1 - \phi^2} = \phi \text{Var}(u_t).$$

In (2.4.63) ist die Autokovarianz zwischen Störtermen unmittelbar benachbarter Beobachtungen im verallgemeinerten Modell gegeben. Wird dagegen die Autokovarianz zwischen den Störvariablen berechnet, die im Beobachtungszeitraum genau j Perioden voneinander entfernt sind, ergibt sich analog

$$(2.4.64) \quad \text{Cov}(u_t, u_{t-j}) = \phi^j \cdot \frac{\sigma_v^2}{1 - \phi^2} = \phi^j \text{Var}(u_t).$$

Da $|\phi| < 1$ ist, nimmt somit die Kovarianz und daher auch die Korrelation zwischen den Störvariablen mit wachsender Laglänge j ab. Der betragsmäßig größte Autokorrelationskoeffizient

$$(2.4.65) \quad \rho_j = \frac{\text{Cov}(u_t, u_{t-j})}{\text{Var}(u_t)} = \frac{\sigma_{t,t-j}}{\sigma^2} = \phi^j; \quad j = 1, 2, \dots$$

ergibt sich also für die Störterme, die unmittelbar aufeinander folgen. Je größer der Parameter ϕ ist, d.h. je näher er betragsmäßig an 1 liegt, desto schwächer ist wegen (2.4.65) die Abnahme der Autokorrelationskoeffizienten, wenn die Länge des Lags j erhöht wird.

Für die im Folgenden zu diskutierenden Schätzstrategie im verallgemeinerten Modell bei Autokorrelation erster Ordnung wird nun $\text{Var}(u_t) = \sigma_v^2 / (1 - \phi^2) = \sigma^2$ gesetzt. Dann folgt für die Varianz-Kovarianz-Matrix der Störterme unter Beachtung von Gleichung (2.4.64)

$$\text{Cov}(\mathbf{u}) = \sigma^2 \mathbf{\Omega} = \begin{pmatrix} \sigma^2 & \phi\sigma^2 & \dots & \phi^{n-1}\sigma^2 \\ \phi\sigma^2 & \sigma^2 & \dots & \phi^{n-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1}\sigma^2 & \phi^{n-2}\sigma^2 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \cdot \begin{pmatrix} 1 & \phi & \dots & \phi^{n-1} \\ \phi & 1 & \dots & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \dots & 1 \end{pmatrix}$$

so dass zur Bestimmung der $\mathbf{\Omega}$ -Matrix lediglich der unbekannt Parameter ϕ zu berechnen ist. Damit lässt sich nun ein Algorithmus angeben, der unmittelbar zu dem GLS-Schätzer $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ führt.

Im ersten Schritt ist $\boldsymbol{\beta}$ im Modell

$$(2.4.66) \quad y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

mit

$$(2.4.67) \quad u_t = \phi \cdot u_{t-1} + v_t,$$

in dem \mathbf{x}'_t als Zeilenvektor die t-te Beobachtung der k exogenen Variablen enthält, nach OLS zu schätzen. Aus $y_t - \mathbf{x}'_t \hat{\boldsymbol{\beta}} = \hat{u}_t$ lassen sich die Residuen berechnen, die im zweiten Schritt in (2.4.67) für die unbekanntenen Störterme einzusetzen sind. Durch erneute Anwendung der OLS-Methode lässt sich so ein Schätzer $\hat{\phi}$ ableiten.

Alternativ kann auf dieser zweiten Stufe natürlich auch die Durbin-Watson-Statistik der Regression (2.4.66) berechnet werden. Aus dem ermittelten Wert DW dieser Größe ergibt sich sofort ein Schätzer für ϕ , da bekanntlich $DW \approx 2(1 - \phi)$ und daher $\hat{\phi} = 1 - DW/2$ gelten muss.

Im dritten Schritt ist mit $\hat{\phi}$ die $\hat{\boldsymbol{\Omega}}$ -Matrix abzuleiten, die aufgrund der bisherigen Überlegungen hier als

$$\hat{\boldsymbol{\Omega}} = \begin{pmatrix} 1 & \hat{\phi} & \dots & \hat{\phi}^{n-1} \\ \hat{\phi} & 1 & \dots & \hat{\phi}^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\phi}^{n-1} & \hat{\phi}^{n-2} & \dots & 1 \end{pmatrix}$$

gegeben ist. Durch Invertierung ergibt sich daraus $\hat{\boldsymbol{\Omega}}^{-1}$. Viertens berechnet sich der GLS-Schätzer $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ dann aus der Beziehung (2.4.51)

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y},$$

in der nun alle Größen gegeben sind. Analog erhält man mit (2.4.51)

$$\hat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = \sigma^2 (\mathbf{X}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1}$$

die Varianz-Kovarianz-Matrix des GLS-Schätzers, in der noch der Parameter

$$\sigma^2 = \text{Var}(u_t) = \sigma_v^2 / (1 - \hat{\phi}^2)$$

unbekannt ist. Man erhält einen Schätzer für σ^2 aus der Beziehung (2.4.52)

$$\hat{\sigma}_{\text{GLS}}^2 = \hat{\mathbf{u}}' \hat{\boldsymbol{\Omega}}^{-1} \hat{\mathbf{u}} / (n - k)$$

oder alternativ durch

$$(2.4.68) \quad \hat{\sigma}_{\text{GLS}}^2 = \hat{\sigma}_v^2 / (1 - \hat{\phi}^2)$$

wobei $\hat{\phi}$ der OLS-Schätzer für ϕ und $\hat{\sigma}_v^2$ die geschätzte Varianz der Störgrößen v_t in der Beziehung (2.4.60) ist, die sich analog zu (2.2.10) ergibt. Damit lässt sich die Kovarianzmatrix der GLS-Schätzer berechnen, auf deren Grundlage die üblichen Parametertests wieder durchführbar sind.

2.4.4.2.3 Alternative Strategien bei autokorrelierten Störtermen

Im Folgenden werden hier einige Strategien angegeben, die eine Schätzung der Ω -Matrix nicht benötigen und daher einfacher handhabbar sind. Ausgangspunkt ist jeweils das Modell

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

mit

$$u_t = \phi u_{t-1} + v_t,$$

in dem \mathbf{x}_t' als $(1 \times k)$ -Vektor die t -te Beobachtung der k exogenen Größen enthält. Wird die Beziehung für y_t um eine Periode verzögert und mit $\phi \neq 0$ multipliziert, hat man

$$\phi y_{t-1} = \phi(\mathbf{x}_{t-1}' \boldsymbol{\beta}) + \phi u_{t-1}.$$

Durch Subtraktion vom Ausgangsmodell ergibt sich

$$y_t - \phi y_{t-1} = \mathbf{x}_t' \boldsymbol{\beta} - \phi \mathbf{x}_{t-1}' \boldsymbol{\beta} + u_t - \phi u_{t-1}$$

oder

$$(2.4.69) \quad y_t - \phi y_{t-1} = (\mathbf{x}_t' - \phi \mathbf{x}_{t-1}') \boldsymbol{\beta} + v_t$$

und damit ein Regressionsansatz, in dem die Störgrößen die klassischen Eigenschaften erfüllen. Gleichung (2.4.69) kann daher durch Anwendung der OLS-Methode geschätzt werden, wenn der Parameter ϕ bekannt ist.

Da ϕ unbekannt ist, muss vorab ein Schätzer $\hat{\phi}$ entwickelt werden, der dann in (2.4.69) einzusetzen ist. Hierfür existieren nun verschiedene Alternativen. Nach der **Methode von Cochran und Orcutt** (1949) ist im ersten Schritt das Modell

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

durch OLS zu schätzen, woraus man die Residuen \hat{u}_t erhält. Mit diesen kann dann aus der Regression

$$\hat{u}_t = \phi \hat{u}_{t-1} + v_t$$

ein Schätzer für $\hat{\phi}$ entwickelt werden, der dann in (2.4.69) zu verwenden ist, so dass abschließend

$$(2.4.70) \quad y_t - \hat{\phi}y_{t-1} = (\mathbf{x}'_t - \hat{\phi}\mathbf{x}'_{t-1})\boldsymbol{\beta} + v_t$$

nur noch nach OLS zu schätzen ist. Man sieht, dass das Cochran-Orcutt-Verfahren mit den im letzten Abschnitt diskutierten ersten beiden Stufen im GLS-Algorithmus identisch ist.

Alternativ ist der **Ansatz von Hildreth und Lu** (1960). Wie bei der Cochran-Orcutt-Methode ist hier im ersten Schritt das Modell $y_t = \mathbf{x}'_t\boldsymbol{\beta} + u_t$ nach OLS zu schätzen. Mit der Durbin-Watson-Statistik erhält man dann Rückschlüsse auf die Art der möglichen Autokorrelation erster Ordnung.

Bei positiver Autokorrelation ist $0 < \phi \leq 1$. Für jeden ϕ -Wert, z.B. $\phi=0,1$, $\phi=0,2$, ..., $\phi=0,9$, $\phi=1,0$ wird nun $\boldsymbol{\beta}$ im Modell (2.4.69) nach OLS geschätzt und der Wert für ϕ angenommen, bei dem die Residuenquadratsumme $\sum \hat{v}_t^2$ minimal wird. Daraus ergeben sich die mit dieser Methode gefundenen Regressionsparameter.

Als **weitere mögliche Alternative** lässt sich der Störprozess auch direkt in den Regressionsansatz einsetzen. Dann folgt mit

$$(2.4.71) \quad y_t = \mathbf{x}'_t\boldsymbol{\beta} + \phi\hat{u}_{t-1} + v_t$$

ein Modell, in dem \hat{u}_{t-1} als erklärende Variable behandelt wird³⁶. Da für die v_t die klassischen Eigenschaften gelten, ist die OLS-Methode auch hier anwendbar. Zu beachten ist jedoch, dass stets eine Schätzung des Modells $y_t = \mathbf{x}'_t\boldsymbol{\beta} + u_t$ erforderlich ist, da sich nur auf diese Weise die Residuen \hat{u}_t und somit die Werte des Regressors \hat{u}_{t-1} erhalten lassen.

Beispiel 2.4. 6: Die Autokorrelation der Störterme der Keynes'schen Konsumfunktion weist auf systematische Fluktuationen des privaten Verbrauchs hin, die nicht durch die

³⁶ Der Regressor u_{t-1} ist mit dem Störprozess unkorreliert, $E(u_{t-1} \cdot v_t) = 0$, weil die Störgröße v_t erst in Periode t auftreten kann. Einen formalen Beweis kann man ableiten, indem die Korrelation berechnet wird, wobei u_{t-1} zunächst in die $MA(\infty)$ -Darstellung zu transformieren ist.

absolute Einkommenshypothese erklärt werden können. Zusätzliche erklärende Variablen und/oder alternative Einkommenskonzepte könnten hier gegebenenfalls Verbesserungen nach sich ziehen. Sofern die Keynes'sche Konsumfunktion Gegenstand der ökonomischen Analyse bleiben soll, kann alternativ mittels ökonometrischer Verfahren versucht werden, die Autokorrelation der Störgröße zu bereinigen. Hier wird die Bereinigung der Autokorrelation mittels der Cochran-Orcutt-Methode vorgestellt.

Hierzu werden die Residuen \hat{u}_t der OLS-geschätzten Konsumfunktion

$$\hat{C}_t = 37,317 + 0,864 \cdot Y_t^v, \\ \quad \quad \quad (0,844) \quad (22,974)$$

$$R^2 = 0,969, \quad DW = 0,635,$$

auf ihre um eine Periode verzögerten Werte \hat{u}_{t-1} regressiert:

$$\hat{u}_t = 0,635 \cdot \hat{u}_{t-1}, \\ \quad \quad \quad (3,580)$$

$$R^2 = 0,462, \quad DW = 1,979.$$

Man erhält hieraus einen OLS-Schätzer für den autoregressiven Parameter ϕ , der zugleich ein Schätzer für den Autokorrelationskoeffizienten 1. Ordnung, ρ_1 , ist:

$$\hat{\rho}_1 = \hat{\phi} = 0,635.$$

Unter Verwendung dieses Schätzers lassen sich die verallgemeinerten Differenzen

$$C_t^* = C_t - 0,635 \cdot C_{t-1}$$

und

$$Y_t^{*v} = Y_t^v - 0,635 \cdot Y_t^v$$

der Modellvariablen C_t und Y_t^v bilden. Eine OLS-Schätzung der verallgemeinerten Differenzengleichung

$$\hat{C}_t^* = \beta_1 \cdot (1 - 0,635) + \beta_2 \cdot Y_t^{*v}$$

ergibt dann die GLS-Schätzer

$$\hat{\beta}_1^* = \hat{\beta}_1 \cdot (1 - 0,635) = 59,285 \quad (t = 2,081)$$

und

$$\hat{\beta}_2^* = \hat{\beta}_2 = 0,768 \quad (t = 2,081)$$

nach der Cochran-Orcutt-Strategie. Während die marginale Konsumquote unmittelbar mit dem Schätzer $\hat{\beta}_2^*$ der transformierten Regressionsgleichung übereinstimmt, ergibt sich der Schätzer für den autonomen Konsum aus der Transformation

$$\hat{\beta}_1 = \hat{\beta}_1^* / (1 - 0,635) = 162,425 .$$

Für den autonomen Konsum (=absolutes Glied) erhält man bei Berücksichtigung der Autokorrelation der Residuen mithin einen erheblich höheren Wert als bei der einfachen LS-Schätzung; dagegen sinkt die marginale Konsumquote (=Streuungsmaß) um knapp 10 Prozentpunkte.

Der Durbin-Watson-Test weist aus, dass mit Hilfe des Cochran-Orcutt-Verfahrens eine Bereinigung der Autokorrelation gelungen ist, da die sich ergebende DW-Statistik von 1,804 nun im Annahmehbereich liegt ($\alpha=0,05$; $n=18$; $k=2$):

$$DW = 1,804 \varepsilon [d_0 = 1,39; 4 - d_0 = 2,61] .$$

Der Determinationskoeffizient $R_{GLS^*}^2$ von 0,902 bezieht sich hier ähnlich wie bei der Anwendung der gewichteten Kleinst-Quadrate-Methode im Falle der Heteroskedastizität allein auf die Güte der Anpassung der verallgemeinerten Differenzgleichung. Hier lässt sich ganz analog wie im Beispiel 2.4.3 ein Determinationskoeffizient R_{GLS}^2 für die GLS-geschätzte Konsumfunktion bestimmen, der bei einem Wert von 0,764 wesentlich von demjenigen der OLS-Schätzung abweicht. ♦

Aufgaben

- (2.4.1) Was versteht man unter Heteroskedastizität und wodurch kann dieser Modelldefekt verursacht werden?
- (2.4.2) Welche inferenzstatistischen Konsequenzen ergeben sich für die OLS-Parameterschätzer eines multiplen Regressionsmodells im Falle von Heteroskedastizität?
- (2.4.3) Erläutern Sie die Grundidee der verallgemeinerten Methode der kleinsten Quadrate (GLS-Methode) zur Beseitigung der Heteroskedastizität!
- (2.4.4) Testen Sie die Störterme der durch das verfügbare Einkommen erklärten Erdgasnachfragefunktion ($\hat{=}$ Energiemodell I, s. Aufgabe 2.1.4) mit Hilfe des Goldfeld-Quandt-Tests auf Heteroskedastizität ($\alpha=0,10$)!
- (2.4.5) Lässt sich das in Aufgabe (2.4.4) erhaltene Testergebnis durch den White-Test bestätigen?
- (2.4.6) Schätzen Sie die in Aufgabe (2.4.4) betrachtete Nachfragefunktion nach Erdgas unter Verwendung der Spezifikation $\sigma_t^2 = \sigma^2 \cdot \text{VEINKR}_t^2$ der Störtermvarianz mit der gewichteten Methode der kleinsten Quadrate (WLS-Methode)!
- (2.4.7) Wie hoch ist die Güte der Anpassung der originären Erdgasnachfragefunktion bei Verwendung der in Aufgabe (2.4.5) ermittelten WLS-Schätzer?
- (2.4.8) Welche Problematik ist mit einer Autokorrelation der Störterme eines ökonomischen Modells verbunden?
- (2.4.9) Wodurch ist ein Markov-Prozess gekennzeichnet?
- (2.4.10) Prüfen Sie die Annahme fehlender Autokorrelation der Störterme der Gasnachfragefunktion mit dem Erdgaspreis als erklärender Variable ($\hat{=}$ Energiemodell IV, Daten s. Anhang B, Tab. B1) auf der Basis des Durbin-Watson-Tests ($\alpha=0,05$)!
- (2.4.11) Bestimmen Sie GLS-Parameterschätzer für die in Aufgabe (2.4.10) überprüfte Nachfragefunktion unter Verwendung der Cochran-Orcutt-Methode! Nehmen Sie Stellung zum Erfolg der GLS-Schätzung im Hinblick auf eine Beseitigung der Autokorrelation ($DW=1,445$; $\alpha=0,05$)!

2.5 Ökonometrische Modelle mit verteilten Verzögerungen

2.5.1 Begriff der verteilten Verzögerungen

In der ökonomischen Theorie wird häufig davon ausgegangen, dass die Wirtschaftseinheiten unverzüglich auf veränderte Marktbedingungen reagieren. Die Variablen des Modells beziehen sich dann jeweils auf denselben Zeitpunkt oder Zeitraum, so dass man auch von einem statischen Ansatz spricht. Ein statisches Modell ist dann adäquat, wenn sich die Reaktionen der Wirtschaftseinheiten innerhalb der gewählten Zeiteinheit vollziehen. So kann z.B. bei Jahresdaten ein statisches Modell gerechtfertigt sein, während bei Monats- oder Quartalsdaten das Verhalten der Wirtschaftssubjekte nur durch ein dynamisches Modell erfasst werden kann. Mithin kommt dem zeitlichen Aggregationsgrad eine nicht unbedeutende Rolle bei der Modellkonstruktion zu.

Einer sofortigen Anpassung an unveränderte Marktbedingungen können verschiedenartige Gründe entgegenstehen. So treten z.B. durch die ökonomischen Entscheidungen vorgelagerte Informationsbeschaffung und -verarbeitung Verzögerungen (Lags) auf. Die Wirkung wirtschaftspolitischer Maßnahmen auf bestimmte Zielgrößen tritt nicht unmittelbar ein, da sie eine Reaktion von Individuen oder Institutionen voraussetzt, die unter Umständen ihre Erwartungen über die wirtschaftliche Entwicklung zu modifizieren haben. Verzögerungen können jedoch auch natürliche, technische oder institutionelle Gründe haben. Während natürliche Ursachen vorwiegend in der Landwirtschaft z.B. durch die Zeit von der Saat bis zur Ernte vorzufinden sind, treten technisch bedingte Verzögerungen in der Industrie z.B. durch die erforderliche Produktionszeit zur Herstellung von Gütern auf. Institutionelle Lags entstehen beispielsweise durch vorgegebene Zahlungstermine oder behördlich erlassene Bestimmungen.

Verteilte Verzögerungen (distributed lags) werden sich immer dann ergeben, wenn die Wirtschaftssubjekte mit Zeitverzögerungen verschiedener Länge auf veränderte Bedingungen reagieren. Dies ist im Allgemeinen dann gegeben, wenn die Individuen heterogene Erwartungen über die wirtschaftliche Entwicklung haben. Jedoch schließen homogene Erwartungen verteilte Verzögerungen keinesfalls aus. Oftmals bilden sich nämlich die Erwartungen der Wirtschaftseinheiten aus der zeitlichen Entwicklung der Vergangenheitswerte einer Variablen heraus. Ökonomische Verhaltenshypothesen wie die Hypothese der partiellen Anpassung und die Hypothese der adaptiven Erwartung

führen zu ganz bestimmten Lag-Modellen. Als Beispiele für eine Verwendung des Modells verteilter Verzögerungen in der ökonomischen Theorie seien hier die Erklärung des Preiserwartungseffekts in der Zinstheorie (Siebke und Willms, 1972) und die Permanente-Einkommens-Hypothese in der Konsumtheorie (Frieman, 1957) genannt. Eines der wichtigsten Anwendungsgebiete von Distributed-Lag-Modellen ist die Erklärung der privaten Investitionstätigkeit in einer Volkswirtschaft (Koyck, 1954; Westerhoff, 1976).

2.5.2 Das allgemeine Modell verteilter Verzögerungen

Das allgemeine Distributed-Lag-Modell geht davon aus, dass eine oder mehrere erklärende Variablen mit unendlich großem Zeithorizont auf die zu erklärende Variable einwirken. Eine lineare Beziehung zwischen der endogenen Variablen y und einer exogenen Variablen x lässt sich dann durch das exogen-dynamische Regressionsmodell

$$(2.5.1) \quad y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + u_t$$

wiedergeben, wobei u eine Störvariable mit dem Erwartungswert Null und konstanter Varianz ist.³⁷ Aufgrund ihrer Abhängigkeit von der Zufallsvariablen u ist die endogene Variable y selbst auch eine stochastische Größe. Dagegen wird die exogene Variable x im Allgemeinen als deterministisch betrachtet. Alternativ kann man die x -Werte als Realisationen eines Zufallsprozesses ansehen, der einigen Mindestvoraussetzungen genügen muss.³⁸

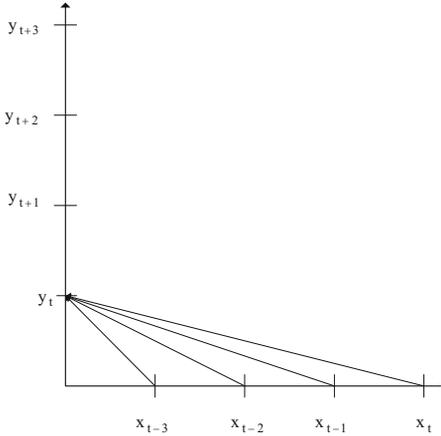
Die Größen $\alpha, \beta_0, \beta_1, \beta_2, \dots$ sind festliegende unbekannte Parameter. Während das absolute Glied α ausschließlich schätztechnischen Zwecken dient, beschreibt die Folge (β_i) die Verzögerungsstruktur des Modells. Ihre Glieder, die man auch als Reaktionskoeffizienten bezeichnet, können in zweifacher Weise interpretiert werden:

³⁷ Im Gegensatz zum endogen-dynamischen Regressionsmodell enthält das exogen-dynamische Regressionsmodell keine zeitlich verzögerten endogenen Variablen. Letzteres lässt sich jedoch formal in ersteres überführen, worauf später noch eingegangen wird.

³⁸ Insbesondere müssen u und x unkorreliert sein, damit die OLS-Methode weiter zu konsistenten Schätzungen führt.

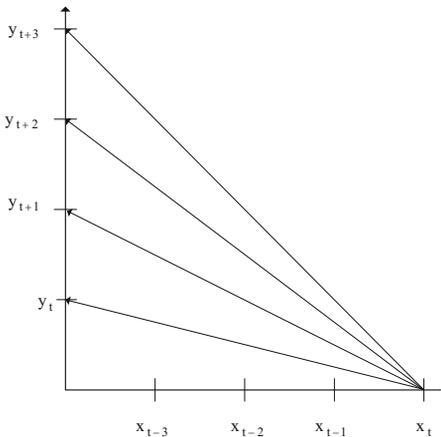
- Zum einen geben die Koeffizienten β_i , $i = 0, 1, 2, \dots$, an, welchen Einfluss die Werte der exogenen Variablen x der Perioden t , $t-1$, $t-2, \dots$ auf die endogene Variable y in der Periode t haben (Abbildung 2.5. 1).

Abbildung 2.5. 1: Grafische Interpretation der Reaktionskoeffizienten (I)



- Zum anderen geben die Koeffizienten β_i an, welchen Einfluss die exogene Variable x z.B. der Periode t auf die Werte der endogenen Variablen y in den Perioden t , $t-1$, $t-2, \dots$ ausübt (Abbildung 2.5. 2).

Abbildung 2.5. 2: Grafische Interpretation der Reaktionskoeffizienten (II)



Die Äquivalenz beider Betrachtungsweisen folgt unmittelbar aus der zeitlichen Konstanz der Verzögerungsstruktur.

Um das Distributed-Lag-Modell (2.5.1) einer ökonomischen Analyse zugänglich zu machen, unterwirft man die Reaktionskoeffizienten β_i in der Regel bestimmten Restriktionen. So ist es häufig sinnvoll zu unterstellen, dass alle β_i das gleiche Vorzeichen besitzen:

$$(2.5.2) \quad \beta_i \cdot \beta_j \geq 0, \quad i, j = 0, 1, 2, \dots$$

Während β_0 die kurzfristige Reaktion (impact multiplier) der endogenen Variablen y_t auf einen Impuls der exogenen Variablen x_t wiedergibt, misst die Reihe

$$\beta_0 + \beta_1 + \beta_2 + \dots = \sum_{i=0}^{\infty} \beta_i = \beta$$

den gesamten Einfluss (total multiplier) aller x -Werte der Perioden $t, t-1, t-2, \dots$ auf y_t . Damit der Gesamteinfluss überhaupt messbar ist, fordern wir, dass die Summe der β_i endlich ist:

$$(2.5.3) \quad \left| \sum_{i=0}^{\infty} \beta_i \right| = |\beta| < \infty.$$

Diese Bedingung impliziert, dass die Folge (β_i) für $i \rightarrow \infty$ gegen Null konvergiert:

$$(2.5.4) \quad \lim_{i \rightarrow \infty} \beta_i = 0,$$

was aus ökonomischer Sicht eine recht plausible Restriktion ist, bedeutet sie doch, dass der Einfluss von x auf y bei genügend großem Zeithorizont verschwindend gering wird.

Vielfach interessiert nicht nur die absolute, sondern auch die relative Wirkung der einzelnen x_{t-i} , $i=0,1,2,\dots$ auf y_t . Man erhält die relativen Verzögerungskoeffizienten w_i , auch Lag-Gewichte genannt, indem man die β_i auf den langfristigen Multiplikator β bezieht:

$$(2.5.5) \quad w_i = \frac{\beta_i}{\beta}, \quad i = 0, 1, 2, \dots$$

Unter Benutzung von Gleichung (2.5.5) formulieren wir das normierte Verzögerungsmodell

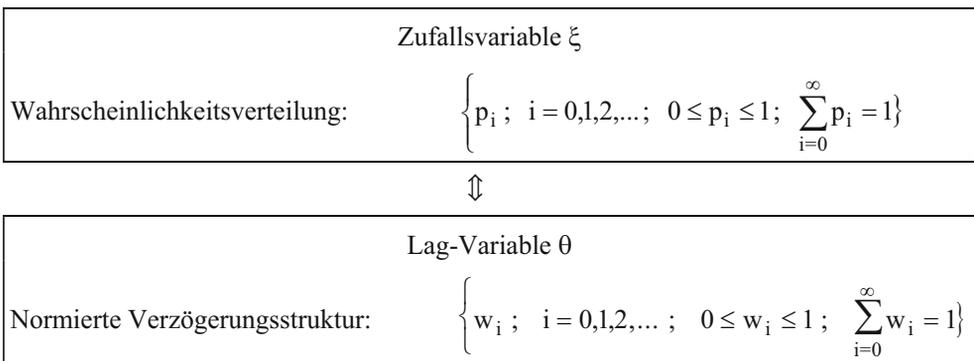
$$(2.5.6) \quad y_t = \alpha + \beta \sum_{i=0}^{\infty} w_i \cdot x_{t-i} + u_t.$$

Da die Lag-Gewichte w_i den Bedingungen

$$(2.5.7) \quad 0 \leq w_i \leq 1 \quad \text{und} \quad \sum_{i=0}^{\infty} w_i = 1$$

genügen, lässt sich die Verzögerungsstruktur ($w_i; i = 0, 1, 2, \dots$) formal wie eine Wahrscheinlichkeitsverteilung behandeln. An Stelle einer diskreten Zufallsvariablen ξ tritt die Lag-Variable θ , deren Ausprägungen $0, 1, 2, \dots$ die den normierten Reaktionskoeffizienten entsprechenden Lag-Gewichte w_0, w_1, w_2, \dots zugeordnet werden (Abbildung 2.5. 1). Mithin lässt sich bei der Analyse von Lag-Strukturen ein Teil des Instrumentariums der Wahrscheinlichkeitsrechnung übernehmen. Es sei hier jedoch ausdrücklich darauf hingewiesen, dass der Lag keine stochastische, sondern eine deterministische Größe ist; daher ist die Analogie zwischen einer Wahrscheinlichkeitsverteilung und einer normierten Verzögerungsstruktur auch nur rein formaler Natur.

Abbildung 2.5. 3: Formale Analogie zwischen den Zufallsvariablen ξ und der Lag-Variablen θ



Komplizierte Verzögerungsverteilungen lassen die Einführung des Lag-Operators L als zweckmäßig erscheinen. Seiner Definition

$$(2.5.8) \quad x_{t-1} = L \cdot x_t$$

entsprechend verschiebt er den Zeitindex einer Variablen um eine Einheit in die Vergangenheit. Als Operator kann L unabhängig von der ihm zugeordneten Variablen wie eine algebraische Größe behandelt werden. Unter Berücksichtigung von (2.5.8) können wir daher schreiben:

$$\begin{aligned}x_{t-2} &= L \cdot x_{t-1} = L(Lx_t) = L^2x_t \\x_{t-3} &= L \cdot x_{t-2} = L(L^2x_t) = L^3x_t\end{aligned}$$

Definieren wir schließlich noch

$$x_t = L^0 \cdot x_t,$$

wobei L^0 die Ausgangsreihe nicht verändert (neutraler Operator), so gilt allgemein

$$x_{t-i} = L^i \cdot x_t, \quad i = 0, 1, 2, \dots \quad ^{39}.$$

Durch Anwendung des Lag-Operators geht das Distributed-Lag-Modell (2.5.6) in die Form

$$\begin{aligned}y_t &= \alpha + \beta(w_0 + w_1L + w_2L^2 + \dots)x_t + u_t \\&= \alpha + \beta\left(\sum_{i=0}^{\infty} w_i L^i\right)x_t + u_t\end{aligned}$$

über. Unter Verwendung der Lag-Reihe

$$W(L) = \sum_{i=0}^{\infty} w_i L^i$$

erhalten wir schließlich die kompakte Form

$$(2.5.9) \quad y_t = \alpha + \beta \cdot W(L) \cdot x_t + u_t$$

des allgemeinen Modells verteilter Verzögerungen.

Obwohl die Reaktionskoeffizienten des allgemeinen Distributed-Lag-Modells bestimmten Restriktionen unterworfen sind, ist dieser Ansatz einer Schätzung noch nicht unmittelbar zugänglich. Im Wesentlichen treten folgende Probleme auf:

- Einer unendlichen Anzahl von zu schätzenden Parameter steht nur ein endlicher Stichprobenumfang gegenüber. Um überhaupt eine Schätzung vornehmen zu können, dürfte nur eine endliche Zahl zeitlich verzögerter exogener Variablen in die Regressionsgleichung einbezogen werden. Aufgrund unserer weitgehenden Unkenntnis über den zeitlichen Ablauf ökonomischer Prozesse ist es aber schwierig festzulegen, von welcher Verzögerungslänge an der Einfluss der x -Werte auf die endogene Variable vernachlässigbar gering ist.
- Selbst wenn obiges Problem befriedigend gelöst werden kann, so besteht die Gefahr, dass die bei ökonomischen Zeitreihen häufig recht starke Korrelation aufein-

³⁹ L^i verschiebt somit die Zeitreihe (x_t) um i Perioden in die Vergangenheit.

ander folgender Werte nicht nur zu einer Beeinträchtigung der Schätzergebnisse führt, sondern sie sogar unbrauchbar werden lässt. Starke Multikollinearität bedingt eine hohe Kovarianz zwischen den Parameterschätzern, was bedeutet, dass eine zufällige Überschätzung eines Parameters mit einer Unterschätzung eines anderen Parameters einhergeht. Außerdem kann die Situation eintreten, dass zwar der Gesamtzusammenhang statistisch gesichert ist, die Einzeleinflüsse jedoch aufgrund erhöhter Varianzen als nicht signifikant erscheinen. Derartig unsichere Schätzergebnisse eignen sich kaum zur Beurteilung ökonomischer Verhaltensweisen.

- Die Einbeziehung zeitverzögerter exogener Variablen in die Regressionsgleichung führt zu einem Verlust an Beobachtungswerten; der damit verbundene Informationsverlust fällt insbesondere bei kurzen Zeitreihen ins Gewicht.

Die Schwierigkeiten bei der Schätzung verteilter Verzögerungen lassen sich durch eine Verminderung der zu schätzenden Parameter reduzieren. Dies erreicht man, in dem man a priori bestimmte Annahmen über die Zeitstruktur des zu untersuchenden Prozesses setzt. Es sind bereits verschiedene Modelle entwickelt worden, die sich bezüglich ihrer Flexibilität teilweise jedoch erheblich unterscheiden. Hier sollen die beiden gebräuchlichsten Verzögerungsstrukturen und die damit zusammenhängenden Schätzmöglichkeiten aufgezeigt werden: das Koyck- und das Almon-Verfahren. Seltener werden dagegen die Pascalsche Lag-Verteilung (Solow, 1960) und die rationale Lag-Verteilung (Jorgenson, 1966) in ökonometrischen Studien verwendet.

2.5.3 Geometrische Lag-Modelle

2.5.3.1 Das Koyck-Modell

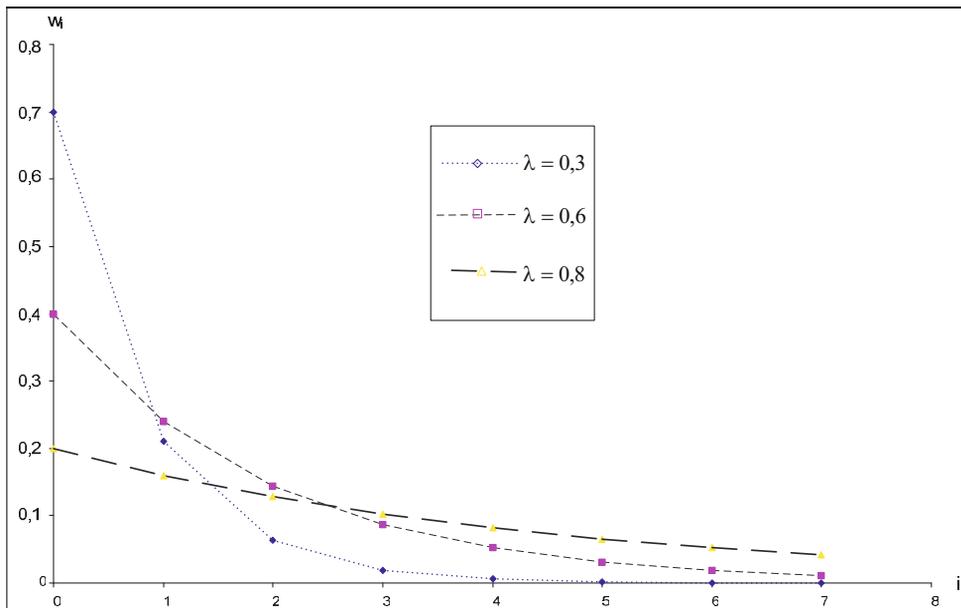
Um die dem allgemeinen Distributed-Lag-Modell inhärenten Schätzprobleme zu umgehen, legte Koyck (1954) seiner Investitionshypothese erstmals eine geometrisch fallende Lag-Verteilung zugrunde. Trotz ihrer recht restriktiven Struktur hat sie eine große praktische Bedeutung erlangt.

In Analogie zur geometrischen Wahrscheinlichkeitsverteilung errechnen sich die Lag-Gewichte nach der Formel

$$(2.5.10) \quad w_i = (1 - \lambda)\lambda^i, \quad i = 0, 1, 2, \dots$$

Der Parameter λ liegt dabei im Intervall $(0, 1)$. Abbildung 2.5. 4 zeigt, dass bei einem niedrigen Wert des Parameters λ die Lag-Verteilung sehr stark abfällt, während sie bei hohem λ relativ flach verläuft.

Abbildung 2.5. 4: Geometrische Lag-Verteilung für alternative Werte des Parameters λ



Unter Verwendung der Kenntnisse der geometrischen Wahrscheinlichkeitsverteilung ergibt sich als mittlerer Lag

$$\bar{\theta} = \frac{\lambda}{1 - \lambda}$$

und für die Varianz der Lag-Verteilung

$$s_{\theta}^2 = \frac{\lambda}{(1 - \lambda)^2}.$$

Wir ersetzen nun die nicht näher spezifizierten Gewichte des allgemeinen normierten Verzögerungsmodells durch (2.5.10) und erhalten die Reaktionsform der geometrischen Lag-Verteilung

$$\begin{aligned}
 (2.5.11) \quad y_t &= \alpha + \beta \sum_{i=0}^{\infty} (1-\lambda)\lambda^i \cdot x_{t-i} + u_t \\
 &= \alpha + \beta(1-\lambda) \sum_{i=0}^{\infty} \lambda^i \cdot x_{t-i} + u_t
 \end{aligned}$$

Unter Anwendung des Lag-Operators L ergibt sich

$$(2.5.12) \quad y_t = \alpha + \beta(1-\lambda) \left(\sum_{i=0}^{\infty} \lambda^i L^i \right) x_t + u_t .$$

Der Ausdruck

$$\sum_{i=0}^{\infty} \lambda^i L^i$$

stellt eine unendliche geometrische Reihe dar, die in $(1-\lambda L)^{-1}$ umgeformt werden kann ⁴⁰:

$$\sum_{i=0}^{\infty} \lambda^i L^i = \frac{1}{1-\lambda L} .$$

Damit geht (2.5.12) in die finale Form ⁴¹

$$(2.5.13) \quad y_t = \alpha + \frac{\beta(1-\lambda)}{1-\lambda L} \cdot x_t + u_t$$

des geometrischen Lag-Modells über, die sich jedoch noch nicht zur Parameterschätzung eignet. Um ein geeignetes Schätzmodell zu erhalten, multiplizieren wir (2.5.13) mit dem Polynom $(1-\lambda L)$ und bringen die zeitverzögerte endogene Variable y_{t-1} auf die rechte Seite. Die sich dadurch ergebende reduzierte Form, die im Gegensatz zur finalen Form auch zeitlich verzögerte endogene Größen als erklärende Variablen umfassen kann, entspricht dem Ergebnis der Koyck-Transformation:

$$\begin{aligned}
 (1-\lambda L)y_t &= \alpha(1-\lambda L) \cdot 1 + \beta(1-\lambda) \cdot x_t + (1-\lambda L) \cdot u_t \\
 y_t - \lambda y_{t-1} &= \alpha(1-\lambda) + \beta(1-\lambda) \cdot x_t + u_t - \lambda u_{t-1} \\
 y_t &= \alpha(1-\lambda) + \beta(1-\lambda) \cdot x_t + \lambda y_{t-1} + u_t - \lambda u_{t-1}
 \end{aligned}$$

$$(2.5.14) \quad y_t = \alpha^* + \beta^* x_t + \lambda y_{t-1} + v_t$$

mit

⁴⁰ Allgemein gilt für eine unendliche geometrische Reihe mit $0 < a < 1$: $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$.

⁴¹ In der finalen Form eines Modells ist die abhängige Variable eine Funktion der exogenen Variablen, des Störterms und etwaiger Anfangswerte der verzögerten endogenen Variablen.

$$(2.5.15) \quad \alpha^* = \alpha(1 - \lambda), \quad \beta^* = \beta(1 - \lambda) \quad \text{und} \quad v_t = u_t - \lambda u_{t-1} .$$

Der wesentliche Vorteil des Regressionsmodells (2.5.14) liegt zweifellos in der beträchtlichen Verminderung des Ausmaßes der Multikollinearität: x_t und y_{t-1} werden nicht so stark korreliert sein wie die zeitverzögerten Werte der exogenen Variablen. Hinzu kommt, dass die Schätzbasis aufgrund der erhöhten Anzahl der Freiheitsgrade sicherer ist. Da λ bereits als Koeffizient in der reduzierten Form geschätzt wird, bestimmen wir nun noch die beiden übrigen Koeffizienten α und β der strukturellen Form. Aus (2.5.15) erhält man

$$\alpha = \frac{\alpha^*}{1 - \lambda} \quad \text{und} \quad \beta = \frac{\beta^*}{1 - \lambda} .$$

Zwar ist im geometrischen Lag-Modell ein Teil der dem allgemeinen Distributed-Lag-Modell inhärenten Schätzschwierigkeiten ausgeräumt. Durch die Transformation der strukturellen in die reduzierte Form entstehen jedoch neue Komplikationen: erstens ist die Störvariable v_t unter sehr allgemeinen Bedingungen autokorreliert; und zweitens besteht zwischen der zeitlich verzögerten endogenen Variablen y_{t-1} und der Zufallsvariablen v_t eine stochastische Beziehung. Damit sind die Annahmen des klassischen Regressionsmodells in zwei Punkten verletzt.

Wir betrachten diese beiden Probleme nun genauer. Da die untransformierte Störvariable u_t mit dem klassischen Regressionsmodell in Verbindung gebracht werden kann, liegt es nahe, sie als unkorreliert zu betrachten. Für die Kovarianz zwischen v_t und v_{t-1} ergibt sich mit (2.5.15)

$$\begin{aligned} \text{Cov}(v_t, v_{t-1}) &= E(v_t, v_{t-1}) = E[(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2})] \\ &= E(u_t u_{t-1}) - \lambda E(u_t u_{t-2}) - \lambda E(u_{t-1}^2) + \lambda^2 E(u_{t-1} u_{t-2}) \\ &= -\lambda E(u_{t-1}^2) \end{aligned}$$

$$\text{wegen} \quad E(u_t u_{t-1}) = E(u_t u_{t-2}) = E(u_{t-1} u_{t-2}) = 0 .$$

Aufgrund der Homoskedastizität erhält man

$$\text{Cov}(v_t, v_{t-1}) = -\lambda \sigma^2 ,$$

was bedeutet, dass die Störgröße v_t des Koyck-Modells (2.5.14) autokorreliert ist. Dadurch würde zwar die Erwartungstreue der Parameterschätzer im Falle einer OLS-Schätzung unberührt bleiben, jedoch müsste ein Effizienzverlust in Kauf genommen werden.

Im klassischen Regressionsmodell sind die erklärenden Größen als exogene Variablen außerdem von der Störvariablen stochastisch unabhängig, zumindest jedoch unkorreliert. Das Auftreten der zeitlich verzögerten endogenen Variablen y_{t-1} als erklärende Größen führt demgegenüber zu einer Verletzung dieser Bedingung. Unter Vernachlässigung des konstanten Gliedes erhält man unter Verwendung der um eine Periode verzögerten Gleichung (2.5.11) für y_{t-1}

$$\begin{aligned} \text{Cov}(v_t, y_{t-1}) &= E \left[(u_t - \lambda u_{t-1}) \left(\beta(1-\lambda) \sum_{i=0}^{\infty} \lambda^i x_{t-i-1} + u_{t-1} \right) \right] \\ &= \beta(1-\lambda) E(u_t x_{t-1}) + \dots + E(u_t u_{t-1}) - \lambda(1-\lambda) \beta E(u_{t-1} x_{t-1}) - \dots - \lambda E(u_{t-1})^2 \end{aligned}$$

Da u und x unkorreliert sind und auch die Kovarianzen der Störvariablen u_t verschwinden, folgt

$$\text{Cov}(v_t, y_{t-1}) = -\lambda E(u_{t-1}^2)$$

und aufgrund der Homoskedastizität

$$\text{Cov}(v_t, y_{t-1}) = -\lambda \sigma^2 .$$

Die negative Korrelation zwischen v_t und y_{t-1} kommt dadurch zustande, dass sowohl v_t als auch y_{t-1} negativ mit u_{t-1} korreliert sind. Da y_{t-1} aufgrund des autoregressiven Schemas (2.5.14) alle zukünftigen Werte der endogenen Variablen y beeinflusst, gilt allgemein

$$\text{Cov}(v_t, y_{t-i}) \neq 0 \quad \text{für } i \geq 0 .$$

Hierdurch entstehen größere Schätzprobleme als bei einer Autokorrelation der Störvariablen. Tatsächlich lässt sich zeigen, dass aufgrund der stochastischen Beziehung zwischen der erklärenden Variablen y_{t-1} und der Störvariablen v_t die Eigenschaft der Erwartungstreue einer OLS-Schätzung verloren geht.

2.5.3.2 Anpassungs- und Erwartungshypothesen

Die Adäquanz der dem geometrischen Lag-Modell zugrunde liegenden Reaktionsform lässt sich nur in Abhängigkeit des jeweils zu untersuchenden Phänomens beurteilen. Es gilt daher, die hinter den Verhaltensweisen der Individuen stehenden ökonomischen Mechanismen aufzudecken. Obwohl Koyck die Annahme geometrisch verteilter Lags nicht näher begründet, kann sie aus den Hypothesen von Nerlove (1958) und Cagan (1956) über das Verhalten der Wirtschaftssubjekte abgeleitet werden. Nerlove legt dabei den Schwerpunkt auf den Anpassungsmechanismus, wohingegen Cagan die Erwartungsbildung in den Vordergrund stellt.

Die **Hypothese der partiellen Anpassung** (partial adjustment hypothesis) postuliert, dass die laufenden Werte einer unabhängigen Variablen x_t das geplante oder gewünschte Niveau y_t^e der abhängigen Variablen y_t bis auf eine Zufallsabweichung determinieren:

$$(2.5.16) \quad y_t^e = \alpha + \beta x_t + u_t .$$

Allerdings stehen einer sofortigen Anpassung an die gleichgewichtige Größe Y_t^e Hemmnisse verschiedenster Art entgegen: eine gewisse Trägheit im Verhalten der Wirtschaftseinheiten, Kosten der Anpassung, fehlende Informationen, Restriktionen technischer, finanzieller, rechtlicher und administrativer Art u.a. mehr. Diese Friktionen führen dazu, dass das gewünschte Niveau nicht vollständig in der Planungsperiode erreicht wird. Nur ein bestimmter Anteil jener zur Erreichung des Gleichgewichts erforderlichen Veränderungen wird auch tatsächlich realisiert, d.h. es gilt

$$(2.5.17) \quad y_t - y_{t-1} = \delta(y_t^e - y_{t-1}), \quad 0 < \delta < 1 ,$$

wobei δ der Anpassungskoeffizient ist. Je näher δ bei 1 liegt, umso kürzer ist der Anpassungszeitraum; ein niedrigerer Wert weist auf einen längeren Zeithorizont hin.

Nach Substitution von y_t^e durch (2.5.16) erhält man

$$(2.5.18) \quad y_t - y_{t-1} = \delta(\alpha + \beta x_t + u_t - y_{t-1}) ,$$

so dass sich nach Auflösen nach y_t und geeigneter Zusammenfassung

$$(2.5.19) \quad y_t = \delta\alpha + \delta\beta x_t + (1-\delta)y_{t-1} + \delta u_t .$$

ergibt. Die Ähnlichkeit zwischen Gleichung (2.5.19) und der reduzierten Form des Koyck-Modells ist unverkennbar. Beiden Modellen liegt eine geometrische Verzögerung

rungsstruktur zugrunde (vgl. Gaab, 1974, S. 13). Allerdings weist die aus dem Anpassungsmodell abgeleitete Endgleichung erheblich einfachere stochastische Eigenschaften auf als die reduzierte Form des Koyck-Modells. Unter der Annahme, dass die ursprüngliche Zufallsvariable u_t nicht autokorreliert ist, gilt diese Eigenschaft auch für die transformierte Störgröße δu_t . Ebenso besteht modellmäßig keine stochastische Abhängigkeit zwischen dem Störterm und der zeitlich verzögerten endogenen Variablen y_{t-1} . Sofern das Modell richtig spezifiziert ist, ergibt sich dadurch gegenüber dem Ausgangsmodell der geometrischen Lag-Verteilung eine schätztechnische Vereinfachung. In der Wirtschaftstheorie wird auf der Grundlage des Modells der partiellen Anpassung z.B. die Anpassung des Kapitalstocks auf eine Nachfrageänderung untersucht. Außerdem findet die Hypothese Anwendung in der dynamischen Geldnachfrageanalyse.

Vorstellungen über die zukünftige Entwicklung einer Variablen orientieren sich vielfach an bestimmten Zukunftserwartungen der sie determinierenden Größen. Eine derartige Verhaltensweise lässt sich durch die Funktion

$$(2.5.20) \quad y_t = \alpha + \beta x_t^e + u_t$$

beschreiben, in der y_t z.B. den Konsum und x_t^e das erwartete Einkommen sein können. Da die Erwartungsgröße x_t^e im Allgemeinen nicht beobachtbar ist,⁴² benötigt man eine Hypothese über die Erwartungsbildung der Wirtschaftssubjekte. Das **Modell der adaptiven Erwartungen** (adaptive expectations model) geht davon aus, dass die Individuen ihre bisherigen Zukunftserwartungen korrigieren, sofern sie eine Diskrepanz zwischen dem realisierten und von ihnen erwarteten Wert einer Größe feststellen. Allerdings werden Irrtümer in der Vorperiode nicht in vollem Umfang in die Erwartungsbildung eingehen, so dass die Erwartungshypothese die Form

$$(2.5.21) \quad x_t^e - x_{t-1}^e = \gamma(x_{t-1} - x_{t-1}^e), \quad 0 < \gamma < 1$$

annimmt. Der Erwartungskoeffizient γ gibt dabei an, in welchem Ausmaß Fehleinschätzung der Vergangenheit in die Antizipationen der Wirtschaftseinheiten eingehen. In dem einen Grenzfall, $\gamma=0$, schlägt sich eine Fehleinschätzung überhaupt nicht in den Zukunftserwartungen nieder; im anderen der hier ausgeschlossenen Grenzfälle, $\gamma=1$, erfolgt eine vollständige Erwartungskorrektur.

⁴² Es gibt allerdings auch Versuche, Plan- und Erwartungsgrößen direkt durch Befragung zu ermitteln. Derartige Erhebungen wurden aber bisher in ökonometrischen Modellen kaum berücksichtigt.

Um die Hypothese der adaptiven Erwartungen einer Schätzung zugänglich zu machen, müssen wir x_t^e in Gleichung (2.5.21) durch eine Funktion beobachtbarer Größen ersetzen. Durch Addition von γx_t^e auf beiden Seiten der Gleichung (2.5.21) ergibt sich

$$x_t^e - (1 - \gamma)x_{t-1}^e = \gamma x_t$$

und nach Auflösen nach der Erwartungsgröße x_t^e

$$x_t^e = (1 - \gamma)x_{t-1}^e + \gamma x_t .$$

Substituiert man nun hierin sukzessive die verzögerte Erwartungsgröße auf der rechten Seite, so erhält man schließlich die Beziehung

$$(2.5.22) \quad x_t^e = \gamma \sum_{i=0}^{\infty} (1 - \gamma)^i x_{t-i} ,$$

in der die aktuelle Erwartungsgröße x_t^e vollständig auf die realisierten Werte der exogenen Variablen x zurückgeführt worden ist. Mit (2.5.22) geht (2.5.20) in

$$(2.5.23) \quad y_t = \alpha + \beta \cdot \gamma \sum_{i=0}^{\infty} (1 - \gamma)^i \cdot x_{t-i} + u_t$$

über, die der Reaktionsform (2.5.11) der geometrischen Lag-Verteilung entspricht. Folglich lässt sich die Koyck-Transformation anwenden, die darin besteht, dass man die um eine Periode verzögerte und mit $(1 - \gamma)$ multiplizierte Gleichung (2.5.23),

$$(1 - \gamma)y_{t-1} = (1 - \gamma)\alpha + \beta\gamma \sum_{i=0}^{\infty} (1 - \gamma)^{i+1} x_{t-i-1} + (1 - \gamma)u_t ,$$

von dieser subtrahiert:

$$y_t - (1 - \gamma)y_{t-1} = \gamma\alpha + \beta\gamma x_t + u_t - (1 - \gamma)u_{t-1} .$$

Durch Umformung ergibt sich schließlich

$$(2.5.24) \quad y_t = \alpha\gamma + \beta\gamma x_t + (1 - \gamma)y_{t-1} + [u_t - (1 - \gamma)u_{t-1}] .$$

Ein Vergleich der reduzierten Formen beider Hypothesen zeigt, dass unterschiedliche Modellansätze zur gleichen dynamischen Regressionsgleichung führen können. Dadurch ergibt sich ein Identifikationsproblem. In der Regel gibt es keine Möglichkeit, zwischen beiden Hypothesen empirisch zu unterscheiden. Den Störtermen der Endgleichungen (2.5.19) und (2.5.24) liegen dagegen unterschiedliche stochastische Eigenschaften zugrunde. Ebenso wie im Koyck-Modell ist der Störterm in Gleichung (2.5.24) autokorreliert und nicht mehr von der erklärenden Variablen y_{t-1} unabhängig. Die dem

Koyck-Modell anhaftenden Schätzprobleme treten daher auch bei dieser Hypothese in unveränderter Form auf.

2.5.3.3 Die OLS-Methode und ihre Schätzeigenschaften

Aufgrund eines Stichprobenergebnisses gibt es keine Möglichkeit, zwischen den drei differierenden Modellansätzen zu differenzieren, die zu der dynamischen Regressionsgleichung

$$(2.5.25) \quad y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + v_t, \quad t = 1, 2, \dots, n$$

führen. Wenn man bei einer Schätzung unmittelbar das dynamische Regressionsmodell (2.5.25) zugrunde legt oder die Hypothese der partiellen Anpassung unterstellt, könnte man wie bei einer multiplen Regression allgemein a priori von einer unkorrelierten Störvariablen v_t ausgehen. Dann würde sich ein Schätzproblem allein durch das Auftreten der endogenen verzögerten Variablen y_{t-1} ergeben. Bei dem Koyck-Modell und der Hypothese der adaptiven Erwartungen ist dagegen auf jeden Fall zunächst einmal eine Autokorrelation der Störgröße zu unterstellen, was bei der Schätzung zusätzlich zu berücksichtigen ist.

Zum Zweck einer Diskussion der Eigenschaften von Schätzern der Parameter des dynamischen Regressionsmodells (2.5.25) gehen wir von der kompakten Darstellung

$$(2.5.26) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$$

mit

$$\mathbf{y} = (y_1, y_2, \dots, y_n)', \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)', \quad \mathbf{v} = (v_1, v_2, \dots, v_n)'$$

und

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & y_0 \\ 1 & x_2 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_{n-1} \end{bmatrix}$$

aus. \mathbf{y}_t ist darin ein $n \times 1$ -Vektor der endogenen Variablen, $\boldsymbol{\beta}$ ein 3×1 -Parameter Vektor und \mathbf{v} ein $n \times 1$ Vektor der Störvariablen. In der $n \times 3$ -Matrix der prädeterminierten Vari-

ablen gibt y_0 einen Anfangswert an, der gleich dem Beobachtungswert der endogenen Variablen in der Periode vor Beginn des Stützbereichs gesetzt werden kann.

Es geht nun um die Eigenschaften des OLS-Schätzers

$$(2.5.27) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y},$$

der allgemein für ein multiples Regressionsmodell bestimmt worden ist. Nach Einsetzen von (2.5.26) für \mathbf{y} lässt sich $\hat{\boldsymbol{\beta}}$ in der Form

$$(2.5.28) \quad \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v}$$

schreiben, so dass man nach Bildung des Erwartungswerts

$$(2.5.29) \quad E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v}].$$

erhält. Wie (2.5.29) zeigt, ist $\hat{\boldsymbol{\beta}}$ nur dann ein erwartungstreuer Schätzer für den Parametervektor $\boldsymbol{\beta}$, wenn der Term $E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v}]$ verschwindet. Wenn man den einfachsten Fall einer unkorrelierten Störvariablen v_t unterstellt, besteht zwischen den Komponenten von \mathbf{X}' und \mathbf{v} keine Korrelation, da die jeweils aktuelle Zufallsvariable v_t nicht nur von allen vergangenen und zukünftigen Werten der exogenen Variablen x , sondern auch von den Werten der zeitlich verzögerten Variablen y_{t-j} für alle $j \geq 1$ stochastisch unabhängig ist. Jedoch enthält die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ mit Ausnahme von y_n alle Stichprobenwerte der endogenen Variablen y , so dass $(\mathbf{X}'\mathbf{X})^{-1}$ und $\mathbf{X}'\mathbf{v}$ auf komplexe Weise voneinander abhängen.

Um diesen Zusammenhang zu verdeutlichen, unterdrücken wir die exogene Variable und die Scheinvariable. Der Erwartungswert des Vektors $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v}$ geht dann in den Ausdruck

$$E \left[\frac{\sum_{t=1}^n y_{t-1} v_t}{\sum_{t=1}^n y_{t-1}^2} \right]$$

über. Während v_t und y_{t-1} im Zähler paarweise unkorreliert sind, beeinflusst die Zufallsvariable v_t die einzelnen Summanden des Nennerterms. Und zwar existieren folgende Wirkungsketten:

$$\begin{array}{ccccccc}
 v_1 & \rightarrow & y_1^2 & \rightarrow & y_2^2 & \rightarrow & \cdots \rightarrow y_{n-1}^2 \\
 v_2 & & & \rightarrow & y_2^2 & \rightarrow & \cdots \rightarrow y_{n-1}^2 \\
 \vdots & & & & & & \vdots \\
 v_{n-1} & & & & & & \rightarrow y_{n-1}^2
 \end{array}$$

Nur der Anfangswert y_0 kann als von der Störgröße v_t unabhängig angesehen werden.

Aufgrund der stochastischen Abhängigkeiten darf $E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{v}]$ nicht in den Erwartungswert der Inversen $(\mathbf{X}'\mathbf{X})^{-1}$ und den Erwartungswert des Vektors $\mathbf{X}'\mathbf{v}$ aufgespalten werden, so dass im Allgemeinen

$$E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{v}] \neq \mathbf{0}$$

und damit auch

$$E(\hat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}$$

gilt.

Hurwicz (1950) untersuchte erstmals analytisch das Ausmaß der Verzerrung des OLS-Schätzers für ein dynamisches Regressionsmodell. Er fand heraus, dass die OLS-Methode die Parameter der Grundgesamtheit dem Betrag nach systematisch unterschätzt. In dem einfachen Fall

$$y_t = \beta y_{t-1} + v_t$$

beträgt der Bias für genügend großes n

$$(2.5.30) \quad E(\hat{\beta}) - \beta \approx -\frac{2\beta}{n}.$$

Allerdings verschwindet der Bias, wenn n über alle Grenzen hinaus wächst. $\hat{\beta}$ ist damit zumindest ein asymptotisch erwartungstreuer Schätzer für β .

Wie sich zeigen lässt, ist der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ nur dann ein konsistenter Schätzer für den Parametervektor $\boldsymbol{\beta}$, wenn die Störvariable v_t unkorreliert ist. Bei einer Autokorrelation der Störvariablen v_t geht die Konsistenzeigenschaft dagegen verloren. Um die Konsistenz des OLS-Schätzvektors $\hat{\boldsymbol{\beta}}$ bei unkorrelierten Störvariablen nachzuweisen, muss gezeigt werden, dass die Schätzer der Regressionskoeffizienten für $n \rightarrow \infty$ mit Wahrscheinlichkeit gegen die "wahren" Parameter konvergieren, d.h. dass

$$\lim_{n \rightarrow \infty} P(|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}| < \boldsymbol{\varepsilon}) = 1, \quad \boldsymbol{\varepsilon} > \mathbf{0}$$

oder kürzer

$$p \lim_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$$

gilt. Bei der Beweisführung kommt die Beziehung

$$(2.5.31) \quad p \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}' \mathbf{v} = \mathbf{0}$$

eine entscheidende Bedeutung zu. (2.5.31) ist genau dann erfüllt, wenn die Bedingungen

$$(2.5.32) \quad \lim_{n \rightarrow \infty} E \left(\frac{1}{n} \mathbf{X}' \mathbf{v} \right) = \mathbf{0}$$

und

$$(2.5.33) \quad \lim_{n \rightarrow \infty} \text{Cov} \left(\frac{1}{n} \mathbf{X}' \mathbf{v} \right) = \mathbf{0}$$

gelten, d.h. mit $n \rightarrow \infty$ müssen der Erwartungswert von $1/n \cdot \mathbf{X}' \mathbf{v}$ gegen den $k \times 1$ -Nullvektor $\mathbf{0}$ und die Kovarianzmatrix von $1/n \cdot \mathbf{X}' \mathbf{v}$ gegen die $k \times k$ -Nullmatrix streben. Wie bei der Untersuchung der Erwartungstreue des Schätzvektors $\hat{\boldsymbol{\beta}}$ bereits gezeigt wurde, gilt aufgrund der Unkorreliertheit zwischen v_t und y_{t+s} , $s \neq 0$, die Beziehung

$$(2.5.34) \quad E \left(\frac{1}{n} \mathbf{X}' \mathbf{v} \right) = \mathbf{0}$$

und damit erst recht (2.5.32). Um die Gültigkeit von (2.5.33) zu zeigen, bilden wir zunächst die Kovarianzmatrix von $1/\sqrt{n} \cdot \mathbf{X}' \mathbf{v}$:

$$\text{Cov} \left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v} \right) = E \left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v} \mathbf{v}' \mathbf{X} \frac{1}{\sqrt{n}} \right).$$

Wegen (2.5.34) können wir den Term

$$E \left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v} \mathbf{v}' \mathbf{X} \frac{1}{\sqrt{n}} \right)$$

aufspalten, so dass man

$$\text{Cov} \left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v} \right) = E(\mathbf{v} \mathbf{v}') \cdot E \left(\frac{1}{n} \mathbf{X}' \mathbf{X} \right)$$

erhält. Aufgrund der fehlenden Autokorrelation der Störterme gilt

$$E(\mathbf{v}\mathbf{v}') = \sigma_v^2 \mathbf{I} ,$$

womit

$$\text{Cov}\left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v}\right) = \sigma_v^2 E\left(\frac{1}{n} \mathbf{X}' \mathbf{X}\right)$$

folgt. Wir nehmen nun an, dass der Erwartungswert der Momentenmatrix $1/n \cdot \mathbf{X}' \mathbf{X}$ in der Grenze $n \rightarrow \infty$ existiert und mit dem Wahrscheinlichkeitslimes identisch ist:

$$(2.5.35) \quad \lim_{n \rightarrow \infty} E\left(\frac{1}{n} \mathbf{X}' \mathbf{X}\right) = \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}' \mathbf{X}\right) = \mathbf{Q} ,$$

wobei \mathbf{Q} eine reguläre $k \times k$ -Matrix ist. Dadurch ergibt sich

$$\lim_{n \rightarrow \infty} \text{Cov}\left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v}\right) = \sigma_v^2 \mathbf{Q} ,$$

so dass für die Kovarianzmatrix von $1/n \cdot \mathbf{X}' \mathbf{v}$

$$\lim_{n \rightarrow \infty} \text{Cov}\left(\frac{1}{n} \mathbf{X}' \mathbf{v}\right) = \lim_{n \rightarrow \infty} \text{Cov}\left(\frac{1}{\sqrt{n}} \mathbf{X}' \mathbf{v}\right) \cdot \lim_{n \rightarrow \infty} \left(\frac{1}{\sqrt{n}}\right) = \sigma_v^2 \mathbf{Q} \cdot \mathbf{0} = \mathbf{0}$$

folgt.

Damit lässt sich nun die Konsistenz des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ bei unkorrelierten Störvariablen beweisen. Durch Anwendung des Wahrscheinlichkeitslimes auf (2.5.29) erhält man

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} &= \text{plim}_{n \rightarrow \infty} \left[\boldsymbol{\beta} + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{v} \right] \\ &= \boldsymbol{\beta} + \text{plim}_{n \rightarrow \infty} \left[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{v} \right] . \end{aligned}$$

Aus dem Theorem von Slutsky folgt⁴³

$$\text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}' \mathbf{X}\right)^{-1} \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}' \mathbf{v}\right)$$

und wegen (2.5.31) und (2.5.35) schließlich

$$\text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \cdot \mathbf{0} = \boldsymbol{\beta} .$$

⁴³ Das Theorem von Slutsky behandelt Funktionszusammenhänge zwischen den Wahrscheinlichkeitslimes von Zufallsvariablen. Allgemein gilt $\text{plim}(\mathbf{X} \pm \mathbf{Y}) = \text{plim} \mathbf{X} \pm \text{plim} \mathbf{Y}$ und $\text{plim}(\mathbf{X}\mathbf{Y}) = \text{plim} \mathbf{X} \cdot \text{plim} \mathbf{Y}$, sofern die stochastischen Matrizen \mathbf{X} und \mathbf{Y} entsprechend dimensioniert sind. S. hierzu Schönfeld (1969), S. 179.

Weiterhin ist die OLS-Schätzung asymptotisch effizient (vgl. Johnston, 1972, S. 305), d.h., es gibt keinen anderen konsistenten Schätzvektor für β , dessen asymptotische Varianz geringer ist als die des OLS-Schätzers $\hat{\beta}$. Außerdem zeigt sich die OLS-Methode in Simulationsstudien auch bei kleinen Stichproben anderen Schätzverfahren als überlegen (Copas, 1966), so dass sie als eine praktikable Schätzmethode für dynamische Regressionsmodelle bei unkorrelierten Störvariablen angesehen werden kann.

Bei autokorrelierten Störvariablen ist dagegen nicht einmal die Mindestforderung der Konsistenz gesichert. Die fehlende Konsistenz des OLS-Schätzers resultiert aus der kontemporären Abhängigkeit der verzögerten endogenen Variablen und der Störvariablen. Einerseits beeinflusst die Störvariable u_{t-1} v_t , andererseits aber auch y_{t-1} , so dass die Kovarianz zwischen v_t und y_{t-1} im Allgemeinen einen von null verschiedenen Wert annimmt.

Um die Inkonsistenz der OLS-Schätzung bei autokorrelierten Störvariablen zu zeigen, ersetzen wir den Störvektor v in (2.5.28) durch $u - \lambda u_{-1}$, wobei der Zeitindex der Komponenten von u_{-1} im Vergleich zu u um eine Einheit in die Vergangenheit verschoben ist.

Es ergibt sich

$$\begin{aligned}\hat{\beta} &= \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{u} - \lambda \mathbf{u}_{-1}) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} - \lambda (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}_{-1}.\end{aligned}$$

Nach Anwendung des Wahrscheinlichkeitslimes erhalten wir

$$\begin{aligned}p \lim_{n \rightarrow \infty} \hat{\beta} &= p \lim_{n \rightarrow \infty} \left[\beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} - \lambda (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}_{-1} \right] \\ &= \beta + p \lim_{n \rightarrow \infty} \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} - \lambda (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}_{-1} \right]\end{aligned}$$

und aus dem Theorem von Slutsky folgt

$$p \lim_{n \rightarrow \infty} \hat{\beta} = \beta + p \lim_{n \rightarrow \infty} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} - \lambda p \lim_{n \rightarrow \infty} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}_{-1}.$$

Da die untransformierte Störvariable u unkorreliert ist, gilt gemäß (2.5.31)

$$p \lim_{n \rightarrow \infty} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} = \mathbf{0}$$

und daher

$$(2.5.36) \quad p \lim_{n \rightarrow \infty} \hat{\beta} = \beta - \lambda \cdot p \lim_{n \rightarrow \infty} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}_{-1}.$$

Da die Komponenten der Matrizen $(\mathbf{X}'\mathbf{X})^{-1}$ und \mathbf{X}' mit denen des verzögerten Störvektors \mathbf{u}_{-1} in stochastischer Beziehung stehen, strebt der Ausdruck $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}_{-1}$ nicht mit Wahrscheinlichkeit gegen den Nullvektor $\mathbf{0}$, was dazu führt, dass

$$p \lim_{n \rightarrow \infty} \hat{\boldsymbol{\beta}} \neq \boldsymbol{\beta}$$

ist.⁴⁴ Die Konsistenzeigenschaft geht bei der OLS-Methode somit bei einer Autokorrelation der Störvariablen verloren. In dieser Situation sind alternative Schätzverfahren wie z.B. die verallgemeinerte Methode der kleinsten Quadrate (GLS-Methode) oder die Methode der Instrumentvariablen (IV-Methode) in Betracht zu ziehen, die zumindest die Konsistenz der Schätzung sichern.⁴⁵

2.5.3.4 Der Durbin-h-Test

Bei ökonometrischen Modellen mit zeitlich verzögerten endogenen Variablen als Regressoren lassen sich mit der OLS-Methode keine erwartungstreuen Schätzer bestimmen. Der OLS-Schätzer für den Parametervektor $\boldsymbol{\beta}$ ist jedoch zumindest konsistent, wenn die Störgröße nicht autokorreliert ist. Bei einer Autokorrelation der Störgröße verschwindet die Verzerrung dagegen nicht einmal mit über alle Grenzen wachsendem Stichprobenumfang, so dass die OLS-Schätzung inkonsistent ist.

Unter diesem Aspekt kommt einer Prüfung der Residuen auf Autokorrelation eine besondere Bedeutung zu. Allerdings ist der Durbin-Watson-Test hier nicht anwendbar, da die Durbin-Watson-Statistik DW bei zeitlich verzögerten endogenen Variablen als Regressoren systematisch gegen den Wert 2 verzerrt ist. Der Durbin-Watson-Test tendiert also dahin, die Nullhypothese einer fehlenden Autokorrelation zu bevorzugen, selbst wenn bestimmte Indizien dagegen sprechen. Durbin (1970) hat jedoch einen Autokorrelationstest für das dynamische Regressionsmodell

⁴⁴ Nur in dem unwahrscheinlichen Fall, dass λ genau mit dem Autokorrelationskoeffizienten erster Ordnung der Residuen, ρ , übereinstimmt, würde die Konsistenz gesichert sein.

⁴⁵ Da die GLS-Methode bereits vorgestellt worden ist, können wir uns hier auf eine Präsentation der Methode der Instrumentvariablen beschränken. Bei der Anwendung der GLS-Methode auf dynamische Regressionsmodelle muss der autoregressive Parameter λ geeignet geschätzt werden, was am besten im Rahmen einer "Gittersuche" (grid search) erfolgt, da dann die Konsistenz des GLS-Schätzers gesichert ist. Danach wird derjenige Wert im Intervall $(0, 1)$ als Schätzwert für λ verwendet, der die Summe der gewogenen Abweichungsquadrate $\mathbf{v}\boldsymbol{\Omega}^{-1}\mathbf{v}$ minimiert. Vgl. Johnston (1972), S. 315.

$$(2.5.37) \quad y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + u_t$$

entwickelt, der als **Durbin-h-Test** bekannt geworden ist. Die Prüfgröße des Tests, die Durbin-h-Statistik genannt wird, ist durch

$$(2.5.38) \quad h = \hat{\rho} \sqrt{\frac{n}{1 - n \left[\widehat{\text{Var}}(\hat{\beta}_3) \right]}}$$

gegeben, in der $\hat{\rho}$ der geschätzte Autokorrelationskoeffizient erster Ordnung und $\widehat{\text{Var}}(\hat{\beta}_3)$ die geschätzte Varianz des Regressionskoeffizienten der zeitlich verzögerten endogenen Variablen y_{t-1} ist. Bei ökonometrischen Analysen lässt sich die Prüfgröße h durch Kenntnis der Durbin-Watson-Statistik DW unter Berücksichtigung der Beziehung

$$DW \approx 2(1 - \hat{\rho})$$

bestimmen:

$$(2.5.39) \quad h = \left(1 - \frac{DW}{2}\right) \cdot \sqrt{\frac{n}{1 - n \left[\widehat{\text{Var}}(\hat{\beta}_3) \right]}}$$

Unter der Nullhypothese

$$H_0 : \rho = 0$$

ist die Durbin-h-Statistik für große Stichproben standardnormalverteilt. Der Durbin-h-Test, mit dem die Autokorrelation erster Ordnung der Störgröße überprüft wird, lässt sich somit unter Verwendung der Standardtabelle der Normalverteilung durchführen. Die Testprozedur bleibt unverändert, wenn in der dynamischen Regressionsgleichung mehrere exogene Variablen oder zusätzliche zeitliche Verzögerungen der endogenen Variablen auftreten.

Vorausgesetzt werden muss bei der Anwendung des Durbin-h-Tests in dieser Form, dass die n -fache geschätzte Varianz des Regressionskoeffizienten von y_{t-1} kleiner als Eins ist. Ansonsten wäre die Prüfgröße h nicht definiert. In dieser Situation empfiehlt Durbin, die Regressionsgleichung

$$\hat{u}_t = \beta_1^* + \beta_2^* x_t + \beta_3^* y_{t-1} + \beta_4^* \hat{u}_{t-1} + v_t$$

mit der gewöhnlichen Methode der kleinsten Quadrate zu schätzen. Die Größen \hat{u}_t sind darin die OLS-Residuen des dynamischen Regressionsmodells (2.5.37). Der Autokorre-

lationstest ist in diesem Fall mit dem t-Test des Regressionskoeffizienten β_4^* der zeitlich verzögerten Residualgröße \hat{u}_{t-1} identisch. Bei einer Signifikanz des Regressionskoeffizienten β_4^* kann auf eine Autokorrelation erster Ordnung der Residuen geschlossen werden. Die Nullhypothese einer fehlenden Autokorrelation wird dagegen beibehalten, wenn sich der Regressionskoeffizient β_4^* als nichtsignifikant erweist.

2.5.3.5 Ein Beispiel

Eine Untersuchung der dynamischen Struktur der Geldnachfragefunktion ist erstmals von Goldfeld (1973) auf der Grundlage der Hypothese der partiellen Anpassung für die US-amerikanische Volkswirtschaft durchgeführt worden. Danach wird die Lücke für die zwischen der gewünschten Geldhaltung m_t^* in der aktuellen Periode und der realisierten Geldhaltung m_{t-1} der Vorperiode z.B. aufgrund von Anpassungskosten nicht sofort vollständig geschlossen. Bei Verwendung der ökonomischen Variablen in logarithmierter Form lautet die Hypothese der partiellen Anpassung der Geldhaltung

$$\ln m_t - \ln m_{t-1} = \delta(\ln m_t^* - \ln m_{t-1}),$$

in der δ einen Anpassungskoeffizienten bezeichnet, der zwischen Null und Eins liegt. Wenn die gewünschte Geldnachfrage nun von dem Aktivitätsniveau und den Opportunitätskosten der Geldhaltung abhängt,

$$\ln m_t^* = \beta_1 + \beta_2 \ln y_t + \beta_3 \ln r_t + u_t,$$

dann erhält man die Geldnachfragefunktion,

$$(2.5.40) \quad \ln m_t = \delta\beta_1 + \delta\beta_2 \ln y_t + \delta\beta_3 \ln r_t + (1 - \delta)\ln m_{t-1} + u_t,$$

in der die Geldmenge in verzögerter Form als erklärende Variable auftritt. Das Aktivitätsniveau wird hier z.B. durch das Bruttosozialprodukt y gemessen; die Opportunitätskosten werden durch einen repräsentativen Zinssatz r erfasst. Sofern die Wirtschaftseinheiten keiner Geldillusion unterliegen, sind das Bruttosozialprodukt und die Geldmenge als reale Größen in die Geldnachfragefunktion einzubeziehen.

Bei einer ökonometrischen Schätzung der dynamischen Geldnachfragefunktion (2.5.40) für die Bundesrepublik Deutschland verwenden wir Quartalsdaten für den Zeitraum von 1960/I bis 1990/II. Die Geldmenge ist in der Abgrenzung M1 verwendet worden, die

das Bargeld und die Sichtguthaben der Nichtbanken umfasst. Das Aktivitätsniveau wird durch das Bruttozialprodukt gemessen, und die Opportunitätskosten werden durch den Geldmarktzinssatz (Dreimonatsgeld) erfasst. Die beiden Niveaugrößen Geldmenge und Bruttozialprodukt werden mit dem Preisindex des Bruttozialproduktes deflationiert, d.h., sie gehen als reale Größen in die Geldnachfragefunktion ein. Alle ökonomischen Variablen werden wie üblich bei der Geldnachfrageanalyse logarithmiert. Die Daten sind den Monatsberichten der Deutschen Bundesbank entnommen.

Die ökonometrische Schätzung der Geldnachfragefunktion (2.5.40) erbrachte folgendes Ergebnis (t-Werte in Klammern):

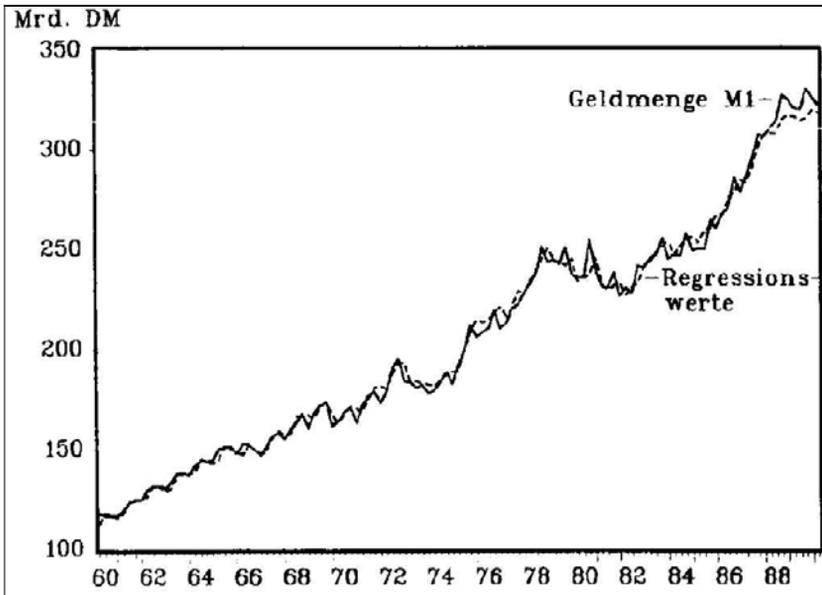
$$\ln \hat{m}_t = -0,304 + 0,305 \ln y_t - 0,044 \ln r_t + 0,743 \ln m_{t-1}$$

$$(-5,595) \quad (8,787) \quad (-7,413) \quad (24,695)$$

$$R^2 = 0,994, \bar{R}^2 = 0,994, F = 6729,284, DW = 2,254742 .$$

Die Einflussgrößen "erklären" insgesamt 99,4 % der Geldmengenstreuung. Der Gesamtzusammenhang ist hochsignifikant. Da die Geldnachfragefunktion eine zeitlich verzögerte endogene Variable als Einflussgröße enthält, ist der Durbin-Watson-Test zum Zwecke einer Überprüfung der Autokorrelation der Störvariablen nicht anwendbar.

Abbildung 2.5. 5: Geldnachfragefunktion mit Koyck-Lag



Jedoch lässt sich mit Hilfe der Durbin-Watson-Statistik die Durbin-h-Statistik bestimmen. Hierzu wird zusätzlich die Varianz des geschätzten Regressionskoeffizienten $\hat{\beta}_4$ der zeitlich verzögerten endogenen Variablen $\ln m_{t-1}$ benötigt:

$$\text{Var}(\hat{\beta}_4) = 0,0300784^2 = 0,00090471 .$$

Bei einem Stichprobenumfang von $n=121$ erhält man dann

$$h = \left(1 - \frac{2,254742}{2}\right) \sqrt{\frac{121}{1 - 121 \cdot 0,00090471}} = -1,4847 .$$

Bei einem Signifikanzniveau von 5 % lautet der kritische Wert $|z_{9,975}| = 1,96$, so dass die Nullhypothese einer fehlenden Autokorrelation nicht widerlegt werden kann.

Außerdem sind alle Regressionskoeffizienten auch auf dem 1 %-Signifikanzniveau statistisch gesichert. Sie haben bezüglich der ökonomischen Einflussgrößen durchweg das erwartete Vorzeichen. Die kurz- und langfristigen Einkommens- und Zinselastizitäten gehen aus folgender Tabelle hervor:

		Elastizität (ϵ)	
		Einkommenselastizität	Zinselastizität
Fristigkeit	kurzfristig	0,305	-0,044
	langfristig	1,187	-0,171

Die kurzfristigen Einkommens- und Zinselastizitäten von 0,305 und $-0,044$ entsprechen den geschätzten Regressionskoeffizienten der beiden ökonomischen Variablen. Dagegen ergeben sich die langfristigen Einkommens- und Zinselastizitäten unter Bezug auf den Anpassungskoeffizienten δ von 0,257, der sich unmittelbar aus dem Regressionskoeffizienten der verzögerten Geldmenge ergibt:

$$\epsilon_{y,M} = \frac{0,305}{1-0,743} = 1,187 \quad \text{und} \quad \epsilon_{r,M} = \frac{-0,044}{1-0,743} = -0,171 .$$

Die langfristigen Einkommens- und Zinselastizitäten lauten somit 1,187 und $-0,171$. Sie ergeben sich nach vollständiger Anpassung der Geldhaltung an das gewünschte Niveau. Wegen

$$\tilde{I}_{0,5} = \frac{\ln(1-0,5)}{\ln 0,743} = 2,33$$

und

$$\tilde{I}_{0,9} = \frac{\ln(1-0,9)}{\ln 0,743} = 7,75$$

hat sich die Anpassung allerdings zu 50 % bereits nach 2,3 Quartalen und zu 90 % nach 7,8 Quartalen vollzogen.⁴⁶ Der mittlere Anpassungslag ist mit

$$\mu = \frac{0,743}{1-0,743} = 2,89$$

um gut ein halbes Quartal länger als der Medianlag.

2.5.3.6 Die Methode der Instrumentvariablen (IV-Methode)

Als Alternative zur verallgemeinerten Methode der kleinsten Quadrate (GLS-Methode) kommt zur Schätzung dynamischer Regressionsmodelle bei Autokorrelation der Stör-

⁴⁶ Allgemein sind die Lag-Quantile \tilde{I}_p , $0 < p < 1$, durch $\tilde{I}_p = [\ln(1-p)]/[\ln(1-\delta)]$ gegeben. Vgl. Harvey (1981), S. 233 f.

variablen die Methode der Instrumentvariablen in Betracht. Ihr Prinzip besteht darin, die zwischen der zeitlich verzögerten endogenen Variablen und der Störvariablen auftretende Korrelation zumindest asymptotisch zu eliminieren, indem erstere durch geeignete Instrumente ersetzt werden. Dabei müssen als Instrumente Größen gewählt werden, die mit der zeitlich verzögerten endogenen Variablen in einem engen Zusammenhang stehen, ohne jedoch mit der Störvariablen korreliert zu sein.

Bei dem dynamischen Regressionsmodell

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t$$

ist die Matrix \mathbf{Z} der Instrumentvariablen wie die Beobachtungsmatrix \mathbf{X} von der Dimension $n \times 3$. Während die exogene Variable x als Instrument für sich selbst verwendet werden kann, ist für die zeitlich verzögerte endogene Variable eine geeignete Instrumentvariable zu finden. Allgemein fordern wir, dass die Matrix \mathbf{Z} der Instrumentvariablen so gewählt wird, dass sie die Eigenschaften

$$(2.5.41) \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}' \mathbf{v} = \mathbf{0}$$

und

$$(2.5.42) \quad \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}' \mathbf{X} = \mathbf{Q}_{ZX}, \mathbf{Q}_{ZX} \quad \text{regulär}$$

erfüllt. Damit wird sichergestellt, dass die Instrumentvariablen bei großen Stichproben mit der Störvariablen unkorreliert sind. Außerdem wird gefordert, dass die Momentenmatrix \mathbf{Q}_{ZX} zwischen den Instrumentvariablen und Regressoren des dynamischen Regressionsmodells asymptotisch existiert und nichtsingulär ist. Wenn man nun das dynamische Regressionsmodell

$$(2.5.43) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \mathbf{y}$$

nicht mit der Transponierten der Beobachtungsmatrix \mathbf{X} wie bei der OLS-Methode, sondern mit der Transponierten der Matrix \mathbf{Z} der Instrumentvariablen multipliziert, dann erhält man das System

$$\mathbf{Z}' \mathbf{y} = \mathbf{Z}' \mathbf{X} \boldsymbol{\beta} + \mathbf{Z}' \mathbf{v} \ ,$$

das mit dem Normalgleichungssystem (2.1.14) vergleichbar ist. Der **Instrumentvariablen-Schätzer** (IV-Schätzer) $\hat{\boldsymbol{\beta}}_{IV}$ für $\boldsymbol{\beta}$ ergibt sich dann durch Invertierung der

Produktmatrix $\mathbf{Z}'\mathbf{X}$ unter Vernachlässigung desjenigen Terms, der von der Störvariablen abhängig ist:

$$(2.5.44) \quad \hat{\boldsymbol{\beta}}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y}.$$

Wie sich zeigen lässt, ist $\hat{\boldsymbol{\beta}}_{IV}$ unter den Bedingungen (2.5.41) und (2.5.42) ein konsistenter Schätzer für den Parametervektor $\boldsymbol{\beta}$. Substituiert man das dynamische Regressionsmodell (2.5.43) in (2.5.44), so erhält man

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{v}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{v}.$$

Die Anwendung des Wahrscheinlichkeitslimes führt zu

$$\begin{aligned} \text{p lim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}}_{IV} &= \text{p lim}_{n \rightarrow \infty} \left[\boldsymbol{\beta} + \left(\frac{1}{n} \mathbf{Z}'\mathbf{X} \right)^{-1} \cdot \frac{1}{n} \mathbf{Z}'\mathbf{v} \right] \\ &= \boldsymbol{\beta} + \text{p lim}_{n \rightarrow \infty} \left[\frac{1}{n} (\mathbf{Z}'\mathbf{X})^{-1} \cdot \frac{1}{n} \mathbf{Z}'\mathbf{v} \right], \end{aligned}$$

was aufgrund des Slutsky-Theorems in

$$\text{p lim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} + \text{p lim}_{n \rightarrow \infty} \left[\frac{1}{n} (\mathbf{Z}'\mathbf{X})^{-1} \right] \cdot \text{p lim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{Z}'\mathbf{v} \right)$$

übergeht. Unter Verwendung der Bedingungen (2.5.41) und (2.5.42) erhält man schließlich

$$\text{p lim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} + \mathbf{Q}_{ZX}^{-1} \cdot \mathbf{0} = \boldsymbol{\beta},$$

womit die Konsistenz von $\hat{\boldsymbol{\beta}}_{IV}$ bewiesen ist.

Es ist jedoch schwierig, Variablen zu finden, die die Beziehung zwischen den zu ersetzenden zeitlich verzögerten endogenen Variablen adäquat wiedergeben. Bei unserem Schätzmodell liegt es nahe, x_{t-1} als Instrument für y_{t-1} zu verwenden. Insbesondere bei mehreren Variablen besteht außerdem die Möglichkeit, an Stelle von x_{t-1} die zeitlich verzögerten Regressionswerte \hat{y}_{t-1} aus einer OLS-Schätzung als Instrumentvariable für y_{t-1} zu benutzen. \hat{y}_{t-1} hat die Eigenschaft, stark mit y_{t-1} korreliert, aber symptotisch mit der Störvariablen unkorreliert zu sein. Als Schätzwert involviert \hat{y}_{t-1} jedoch Zufallsschwankungen, die verhindern, dass der Small-Sample-Bias verschwindet. Bei k exogenen Variablen x_1, x_2, \dots, x_k ist das Instrument \hat{y}_{t-1} durch

$$\hat{y}_{t-1} = \hat{\beta}_0^* + \hat{\beta}_1^* x_{1,t-1} + \hat{\beta}_2^* x_{2,t-1} + \dots + \hat{\beta}_k^* x_{k,t-1}$$

gegeben. Mit den in der 1. Stufe ermittelten Regressionswerten \hat{y}_{t-1} erfolgt eine Parameterschätzung der dynamischen Regressionsgleichung

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 \hat{y}_{t-1} + v_t$$

in der 2. Stufe. Diese Vorgehensweise entspricht dem Prinzip der zweistufigen Methode der kleinsten Quadrate. Für $k=1$ sind die Schätzer der Instrumente x_{t-1} und \hat{y}_{t-1} identisch. Allerdings lässt sich zeigen, dass letztere Schätzfunktion für $k>1$ eine höhere Effizienz aufweist, da die Stärke des Zusammenhangs zwischen \hat{y}_{t-1} und y_{t-1} zunimmt. In der einfachen Beziehung

$$y_t = \beta y_{t-1} + v_t$$

beträgt die Varianz von $\hat{\beta}$

$$\text{Var}(\hat{\beta}) = \sigma^2 \frac{1}{\sum \hat{y}_{t-1} y_{t-1}},$$

woran deutlich wird, dass die IV-Schätzung umso unsicherer wird, je schwächer die Instrumentvariable mit der zu ersetzenden zeitlich verzögerten endogenen Variable korreliert ist. Im Extremfall, wenn $\sum \hat{y}_{t-1} y_{t-1} = 0$ ist, strebt $\text{Var}(\hat{\beta})$ gegen unendlich. Wie Simulationsstudien jedoch ergeben haben, nimmt der Erklärungswert für $k>2$ meist nur noch geringfügig zu (vgl. Gaab, 1974, S. 31).

Der Vorteil der IV-Methode liegt in der Einfachheit und anschaulichen Interpretationsmöglichkeit. Außerdem sind keine Annahmen über den zugrunde liegenden Störprozess erforderlich. Simulationsstudien haben jedoch gezeigt, dass die Varianzen der IV-Schätzungen bei kleinem und mittlerem Stichprobenumfang häufig recht groß sind, so dass mit Effizienzverlusten gerechnet werden muss (vgl. Schönfeld, 1971, S. 92).

2.5.4 Das Almon-Verfahren

In einer Untersuchung der Verzögerungsstruktur zwischen Kapitalzuweisungen und Investitionsausgaben hat Almon (1965) ein flexibles Verfahren zur Schätzung von Distributed-Lag-Modellen vorgestellt, das in der angewandten Ökonometrie eine große Bedeutung erlangt hat. Bei diesem Verfahren wird vorausgesetzt, dass der Einfluss der

unabhängigen Variablen x auf die abhängige Variable y nicht unbegrenzt ist, sondern nach s Perioden verschwindet:

$$(2.5.45) \quad y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_s x_{t-s} + u_t.$$

Da der Gesamtlag von s Perioden vor allem bei unterjährigen Daten im Allgemeinen recht groß sein wird, ist eine direkte Schätzung von (2.5.45) aufgrund des Multikollinearitätsproblems zu umgehen. Im Almon-Verfahren werden die Reaktionskoeffizienten β_i daher durch ein Lag-Polynom r -ten Grades, $r < s$, approximiert⁴⁷:

$$(2.5.46) \quad \beta_i = a_0 + a_1 \cdot i + a_2 \cdot i^2 + \dots + a_r \cdot i^r, \quad i = 0, 1, 2, \dots, s.$$

Aufgrund dieser Darstellung spricht man beim Almon-Verfahren auch von einer polynomialen Lag-Verteilung. Vorteile gegenüber dem direkten Ansatz ergeben sich nur dann, wenn der Polynomgrad r wesentlich kleiner als der Gesamtlag s ist.

Direkt geschätzt werden nicht die Reaktionskoeffizienten β_i , $i = 0, 1, 2, \dots, s$, sondern die Lag-Koeffizienten a_k , $k = 0, 1, 2, \dots, r$. Substituiert man (2.5.46) aus diesem Grund in das Distributed-Lag-Modell (2.5.45), so erhält man

$$y_t = \alpha + a_0 x_t + \left(a_0 + \sum_{k=1}^r a_k \right) x_{t-1} + \left(a_0 + \sum_{k=1}^r 2^k a_k \right) x_{t-2} + \dots + \left(a_0 + \sum_{k=1}^r s^k a_k \right) x_{t-s} + u_t$$

und nach geeigneter Umstellung

$$(2.5.47) \quad \begin{aligned} y_t &= \alpha + a_0 \sum_{i=0}^s x_{t-i} + \sum_{i=1}^s \sum_{k=1}^r i^k a_k x_{t-i} + u_t \\ &= \alpha + a_0 \sum_{i=0}^s x_{t-i} + \sum_{k=1}^r a_k \sum_{i=1}^s i^k x_{t-i} + u_t \\ &= \alpha + a_0 \sum_{i=0}^s x_{t-i} + a_1 \sum_{i=1}^s i x_{t-i} + a_2 \sum_{i=1}^s i^2 x_{t-i} + \dots + a_r \sum_{i=1}^s i^r x_{t-i} + u_t \end{aligned}$$

Die $r+1$ Größen

$$z_{0t} = \sum_{i=0}^s x_{t-i}, z_{1t} = \sum_{i=0}^s i x_{t-i}, z_{2t} = \sum_{i=0}^s i^2 x_{t-i}, \dots, z_{rt} = \sum_{i=0}^s i^r x_{t-i}$$

sind nun die unabhängigen Variablen im polynomialen Lag-Modell,

$$(2.5.48) \quad y_t = \alpha + a_0 z_{0t} + a_1 z_{1t} + a_2 z_{2t} + \dots + a_r z_{rt} + u_t,$$

⁴⁷ Grundlage für diese Formel ist ein Theorem von Weierstrass, nachdem jede stetige Funktion innerhalb eines geschlossenen Intervalls mit beliebiger Genauigkeit approximiert werden kann.

das mit der OLS-Methode geschätzt werden kann. Wenn OLS-Schätzwerte für die Lag-Koeffizienten a_0, a_1, \dots, a_r vorliegen, können die Reaktionskoeffizienten β_i über die Beziehung (2.5.46) berechnet werden.

Um die Reaktionskoeffizienten β_i auf Signifikanz testen zu können, wird die Kovarianzmatrix der Lag-Koeffizienten a_k benötigt. Aus diesem Grund bietet sich die Verwendung der Matrizennotation für das Almon-Verfahren an. Das Distributed-Lag-Modell mit einem Gesamtlag von s lautet dann

$$(2.5.49) \quad \mathbf{y} = \boldsymbol{\alpha}\mathbf{e} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

mit

$$\begin{aligned} \mathbf{e} &= (1 \quad 1 \quad \cdots \quad 1)', \\ \mathbf{y} &= (y_1 \quad y_2 \quad \cdots \quad y_n)', \\ \boldsymbol{\beta} &= (\alpha \quad \beta_0 \quad \beta_1 \quad \cdots \quad \beta_s)', \\ \mathbf{u} &= (u_1 \quad u_2 \quad \cdots \quad u_n)', \end{aligned}$$

und

$$\mathbf{X} = \begin{bmatrix} x_1 & x_0 & x_{-1} & \cdots & x_{1-s} \\ x_2 & x_1 & x_0 & \cdots & x_{2-s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n-1} & x_{n-2} & \cdots & x_{n-s} \end{bmatrix}.$$

Der Vektor der Reaktionskoeffizienten $\boldsymbol{\beta}$ lässt sich unter Beachtung von (2.5.46) in der Form

$$(2.5.50) \quad \boldsymbol{\beta} = \mathbf{H} \cdot \mathbf{a}$$

mit

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 2^1 & 2^2 & \cdots & 2^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s^1 & s^2 & \cdots & s^r \end{bmatrix}$$

und $\mathbf{a} = (a_0 \ a_1 \ \dots \ a_r)'$

darstellen. Substituiert man nun $\boldsymbol{\beta}$ in (2.5.49) durch (2.5.50), so ergibt sich

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\alpha}\mathbf{e} + \mathbf{X}(\mathbf{H}\mathbf{a}) + \mathbf{u} \\ &= \boldsymbol{\alpha}\mathbf{e} + (\mathbf{X}\mathbf{H})\mathbf{a} + \mathbf{u} \end{aligned}$$

und mit

$$\mathbf{Z} = \mathbf{X}\mathbf{H}$$

schließlich die Schätzgleichung

$$(2.5.51) \quad \mathbf{y} = \boldsymbol{\alpha}\mathbf{e} + \mathbf{Z}\mathbf{a} + \mathbf{u}$$

des Almon-Verfahrens.

Der OLS-Schätzer für den Vektor \mathbf{a} der Lag-Koeffizienten lautet

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} ,$$

so dass der OLS-Schätzer für $\boldsymbol{\beta}$ von der Form

$$(2.5.52) \quad \hat{\boldsymbol{\beta}} = \mathbf{H}\hat{\mathbf{a}} = \mathbf{H}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

ist. Die Kovarianzmatrix von $\hat{\mathbf{a}}$ ist durch

$$(2.5.53) \quad \text{Cov}(\hat{\mathbf{a}}) = \sigma_u^2 (\mathbf{Z}'\mathbf{Z})^{-1}$$

gegeben. Zur Bestimmung der Varianzen der Reaktionskoeffizienten ist zu beachten, dass die β_i lineare Funktionen der Lag-Koeffizienten a_k sind, die sich in der Form

$$\beta_i = \mathbf{h}'_i \mathbf{a} , \quad i = 0, 1, 2, \dots, s$$

darstellen lassen mit \mathbf{h}_i als $(i+1)$ -ter Zeile von \mathbf{H} . Mithin erhält man

$$\text{Var}(\hat{\beta}_i) = \text{Var}(\mathbf{h}'_i \hat{\mathbf{a}}) = \mathbf{h}'_i \text{Var}(\hat{\mathbf{a}}) \mathbf{h}_i$$

und unter Berücksichtigung von (2.5.53) schließlich

$$(2.5.54) \quad \text{Var}(\hat{\beta}_i) = \sigma_u^2 \mathbf{h}'_i (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{h}_i .$$

Unter Verwendung dieser Beziehung lassen sich die Reaktionskoeffizienten mit dem bekannten t-Test auf Signifikanz überprüfen.

Bei der Anwendung der Almon-Schätztechnik ist es erforderlich, sowohl die Länge der Verzögerung als auch den Grad des Lag-Polynoms festzulegen. Da die ökonomische Theorie hierüber kaum eine Aussage machen kann, sind in der Regel hierzu statistische Kriterien heranzuziehen. Wegen der schnell zunehmenden Multikollinearität kommt als Näherungsfunktion ein Polynom höheren als dritten oder vierten Grades wohl kaum in Betracht. Man kann daher unterschiedliche Verzögerungslängen für alternative Polynomgrade vorgeben und die Wahl an Hand des korrigierten Bestimmtheitsmaßes unter Berücksichtigung der Signifikanz der Reaktionskoeffizienten treffen. Vom Unter-

suchungsgegenstand aus betrachtet ist es entscheidend, ob die ermittelten Reaktionskoeffizienten die theoretisch erwartete Verzögerungsstruktur widerspiegeln. So müssen insbesondere alle β_1 -Koeffizienten das gleiche Vorzeichen besitzen und der Summenkoeffizient darf häufig nicht den Wert Eins überschreiben. Gerade die Flexibilität des Almon-Modells ermöglicht es dem Forscher, ohne zu viel restriktive Vorgaben dynamische Hypothesen einem empirischen Test zu unterziehen.

Aufgaben

- (2.5.1) Erläutern Sie den Begriff der verteilten Verzögerung anhand der Investitionsstätigkeit der Bauwirtschaft!
- (2.5.2) Was versteht man unter einem exogen-dynamischen Regressionsmodell und welche Schätzprobleme birgt es?
- (2.5.3) Skizzieren Sie anhand des in Aufgabe (2.5.1) genannten Beispiels den Unterschied zwischen dem „impact multiplier“ und „total multiplier“!
- (2.5.4) Geben Sie die Lag-Polynom-Darstellung des Modells
$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + u_t$$
 an!
- (2.5.5) Wie lässt sich die Koyck-Transformation eines exogen-dynamischen Regressionsmodells in ein endogen-dynamisches Regressionsmodell der Form $y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + v_t$ inhaltlich klären?
- (2.5.6) Zeigen Sie, wie sich die Habit-Persistence-Hypothese
$$C_t = \beta_0 + \beta_1 y_t + \beta_2 C_{t-1} + v_t$$
aus der Hypothese der partiellen Anpassung ableiten lässt!
- (2.5.7) Schätzen Sie die in Aufgabe (2.5.6) wiedergegebene Habit-Persistence-Hypothese (Daten s. Beispiel 2.1.1) unter Verwendung der OLS-Methode! Welche Folgerungen hinsichtlich der Schätzeigenschaften lassen sich aus dem Durbin-h-Test ziehen?
- (2.5.8) Wie groß sind die kurz- und langfristigen Einkommenselastizitäten des privaten Konsums auf der Basis der partiellen Anpassungshypothese unter Verwendung der in Aufgabe (2.5.7) ermittelten Schätzergebnisse?
- (2.5.9) Bestimmen Sie für die in Aufgabe (2.5.7) ökonometrisch geschätzte Konsumfunktion der Medianlag und mittleren Anpassungslag (Interpretation)!
- (2.5.10) Welche Grundidee einer ökonometrischen Schätzung von Distributed-Lag-Modellen steht hinter dem Almon-Verfahren?

2.6 Modelle mit qualitativen Variablen

2.6.1 Vorbemerkungen

Ökonomische Variablen sind in ihrer überwiegenden Mehrzahl quantitativ, so dass bei den ökonometrischen Modellen zunächst einmal auf diese Variablenart abgestellt worden ist. Aus verschiedenen Gründen haben jedoch **Regressionsmodelle mit qualitativen Variablen** in der Ökonometrie einen festen Stellenwert. Als exogene Variablen kommen qualitative Größen einerseits dann in Betracht, wenn in ökonomischer Hinsicht Verhaltensunterschiede zwischen verschiedenen Personengruppen zu erwarten sind. So ist z.B. ein unterschiedliches Konsum- und Sparverhalten in den verschiedenen sozialen Schichten der Bevölkerung bekannt. Bei bestimmten Produkten ist z.B. mit einem unterschiedlichen Kaufverhalten von Frauen und Männern zu rechnen. **Qualitative Regressoren**, die bei einer ökonometrischen Analyse berücksichtigt werden müssten, sind hier die soziale Schicht und das Geschlecht. Diese Art von qualitativen Variablen, die man als Dummy-Variablen bezeichnet, kommt in der Regel nur bei Querschnittsanalysen vor. Bei unterjährigen Zeitreihendaten sind dagegen gelegentlich saisonale Einflüsse durch Dummy-Variablen zu modellieren.

Mit einem **Strukturbruchtest** kann überprüft werden, ob das ökonometrische Modell im gesamten Beobachtungszeitraum dieselbe Struktur besitzt. Nach bestimmten wirtschaftlichen Ereignissen wie z.B. einem Ölpreisschock oder dem Aufkommen von Finanzinnovationen kann sich das Verhalten der Wirtschaftssubjekte grundlegend ändern, so dass sich eine Unterteilung des Beobachtungszeitraumes als erforderlich erweist. Die Einordnung von Strukturbruchtests unter dieser Rubrik ergibt sich daraus, dass er prinzipiell in Form einer Dummy-Variablen in einem ökonometrischen Modell berücksichtigt werden kann.

Während qualitative Regressoren durch ein entsprechendes "Design" der Beobachtungsmatrix X in ein multiples Regressionsmodell einbezogen werden können, sieht die Situation bei einer **qualitativen endogenen Variablen** völlig anders aus. Wenn z.B. die Kaufwahrscheinlichkeit für ein Produkt in Abhängigkeit bestimmter Einflussgrößen erklärt werden soll, liegt eine qualitative abhängige Variable vor, die bei einem Kauf des Produkts durch eine Person oder einen Haushalt den Wert Eins und bei einem Nichtkauf den Wert Null annimmt. Eine ökonometrische Analyse basiert hier auf sog. Logit-

oder Probitmodellen, da die Schätzung von Anteilswerten Probleme aufwirft. Wie sich zeigen wird, können die klassischen Annahmen des multiplen Regressionsmodells bei ökonometrischen Modellen mit einer qualitativen abhängigen Variablen nicht mehr vorausgesetzt werden.

2.6.2 Qualitative Regressoren

Obwohl ökonomische Variablen, die zur Erklärung eines Tatbestandes in Betracht kommen, überwiegend quantitative Größen sind, liegt bei bestimmten Phänomenen ein qualitativer Einfluss vor. Wir wollen die Behandlung derartiger qualitativer Einflüsse anhand einer einfachen Nachfragefunktion aufzeigen, bei der die Nachfrage q_t nach einer bestimmten Güterart in Abhängigkeit von den Gesamtausgaben x_t modelliert wird:

$$(2.6.1) \quad q_t = \beta_0 + \beta_1 x_t + u_t .$$

Die Störvariable u_t soll darin den klassischen Annahmen des Regressionsmodells genügen. Eine Vernachlässigung des Preises der betrachteten Güterart und der Preise konkurrierender Güter ist z.B. bei einer Querschnittsanalyse immer dann zulässig, wenn die Preise innerhalb der Erhebungsperiode im Wesentlichen konstant bleiben. Je nach der Güterart kann jedoch ein Einfluss des Geschlechts auf die Nachfrage vorliegen. So ist z.B. bei Kosmetikartikeln zwischen weiblichen und männlichen Personen ein unterschiedliches Nachfrageverhalten zu erwarten.

Wie kann nun der Einfluss der qualitativen Variablen Geschlecht in der Nachfrageanalyse berücksichtigt werden? Wir definieren hierzu eine Dummy-Variable g_t , die den Wert 1 bei einer weiblichen und den Wert 0 bei einer männlichen Erhebungseinheit t annimmt:

$$g_t = \begin{cases} 1, & \text{falls Frau} \\ 0, & \text{falls Mann.} \end{cases}$$

Die Nachfragefunktion (2.6.1) lautet dann bei einer Erweiterung um den qualitativen Regressor Geschlecht

$$(2.6.2) \quad q_t = \beta_0 + \beta_1 x_t + \beta_2 g_t + u_t .$$

Bei acht Erhebungseinheiten mit der Reihenfolge 2x weiblich, 1x männlich, 1x weiblich, 3x männlich, 1x weiblich beim Merkmal Geschlecht hätte die Beobachtungsmatrix \mathbf{X} folgende Gestalt:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & 1 \\ 1 & x_2 & 1 \\ 1 & x_3 & 0 \\ 1 & x_4 & 1 \\ 1 & x_5 & 0 \\ 1 & x_6 & 0 \\ 1 & x_7 & 0 \\ 1 & x_8 & 1 \end{bmatrix}$$

Es ergibt sich somit ein ganz bestimmter Aufbau oder Design der Beobachtungsmatrix \mathbf{X} , so dass man sie bei der Einbeziehung qualitativer Variablen gelegentlich auch als Designmatrix bezeichnet.⁴⁸

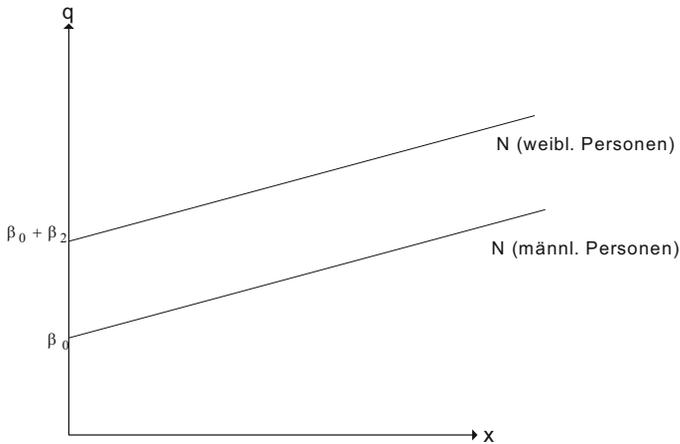
Eine Interpretation des Einflusses des qualitativen Regressors Geschlecht erhält man, indem man die Nachfragefunktion (2.6.2) getrennt für männliche und weibliche Personen betrachtet:

$$\text{für männliche Personen: } q_t = \beta_0 + \beta_1 x_t + u_t$$

$$\text{für weibliche Personen: } q_t = (\beta_0 + \beta_2) + \beta_1 x_t + u_t .$$

Die Einbeziehung der qualitativen Variablen Geschlecht bewirkt eine Parallelverschiebung der Nachfragefunktion (N), was in Abbildung 2.6. 1 für den Fall $\beta_2 > 0$ dargestellt worden ist. Die Dummy-Variable misst in diesem Fall die bei gegebenen Gesamtausgaben größere Nachfrage der Frauen im Vergleich zu Männern nach der betrachteten Güterart.

⁴⁸ Der Begriff der Designmatrix, der aus der kategorialen Datenanalyse (Varianzanalyse, kategoriale Regressionsanalyse) stammt, wird jedoch in der Ökonometrie weniger häufig verwendet, da in ökonometrischen Modellen in der Regel nicht ausschließlich qualitative Regressoren verwendet werden.

Abbildung 2.6. 1: Verschiebung der Nachfragefunktion

Die Schätzung eines Regressionsmodells mit qualitativen Regressoren kann mit der gewöhnlichen Methode der kleinsten Quadrate erfolgen. Der OLS-Schätzer für β_2 kann mit dem t-Test in der üblichen Form auf Signifikanz geprüft werden. Ein signifikantes Testergebnis bedeutet z.B. im Falle der Nachfragefunktion, dass ein unterschiedliches Nachfrageverhalten zwischen Frauen und Männern in Bezug auf die betrachtete Güterart statistisch gesichert ist. Zu beachten ist allerdings, dass die Nachfragefunktion (2.6.2) implizit die Hypothese einer gleichen marginalen Konsumneigung enthält, da der Regressionskoeffizient β_1 in beiden Personengruppen als gleich angesehen wird. Mit dem ökonometrischen Modell (2.6.2) lässt sich allein die Relevanz einer Parallelverschiebung der Nachfragefunktion überprüfen.

Gegenüber einer separaten Schätzung der Nachfragefunktion für beide Personengruppen hat der Dummy-Variablenansatz den Vorteil einer größeren Anzahl von Freiheitsgraden, die die Schätzer zuverlässiger werden lassen. Die Möglichkeit einer unterschiedlichen marginalen Konsumneigung kann durch das ökonometrische Modell

$$(2.6.3) \quad q_t = \beta_0 + \beta_1 x_t + \beta_2 x_t g_t + u_t$$

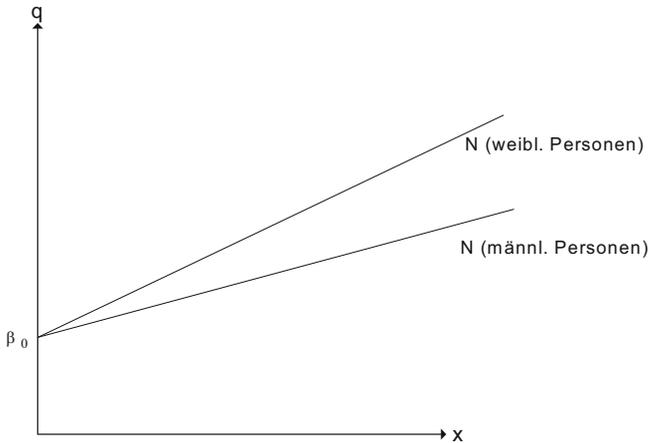
erfasst werden. In diesem Fall, der in Abbildung 2.6. 2 für $\beta_2 > 0$ dargestellt worden ist, gibt β_2 die höhere oder geringere marginale Konsumneigung der weiblichen gegenüber den männlichen Nachfragern wieder:

$$\text{für männliche Personen: } q_t = \beta_0 + \beta_1 x_t + u_t$$

$$\text{für weibliche Personen: } q_t = \beta_0 + (\beta_1 + \beta_2)x_t + u_t .$$

Natürlich ist es ebenfalls möglich, beide Verhaltensunterschiede zwischen den Personengruppen in einem ökonometrischen Modell zu erfassen.

Abbildung 2.6. 2: Veränderung der marginalen Konsumneigung



Anhand einer qualitativen Variablen mit drei Ausprägungen wollen wir allgemein die Einbeziehung von qualitativen Regressoren in ökonometrischen Modellen erörtern. Hierzu wird die Nachfrage nach einer Güterart neben den Gesamtausgaben in Abhängigkeit von der sozialen Stellung eines Haushalts betrachtet. Wenn man nun für jede der drei Kategorien der sozialen Stellung (s) eine Dummy-Variable einführen würde, dann lautete die Nachfragefunktion

$$q_t = \beta_0 + \beta_1 x_t + \beta_2 s_{1t} + \beta_3 s_{2t} + \beta_4 s_{3t} + u_t$$

mit

$$s_{ht} = \begin{cases} 1, & \text{falls soziale Stellung } h \\ 0, & \text{sonst} \end{cases}, \quad h = 1, 2, 3 .$$

Damit würde sich z.B. bei 8 Haushalten folgendes Design der Beobachtungsmatrix \mathbf{X} ergeben:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & 0 & 1 & 0 \\ 1 & x_2 & 0 & 0 & 1 \\ 1 & x_3 & 0 & 0 & 1 \\ 1 & x_4 & 1 & 0 & 0 \\ 1 & x_5 & 0 & 1 & 0 \\ 1 & x_6 & 1 & 0 & 0 \\ 1 & x_7 & 1 & 0 & 0 \\ 1 & x_8 & 0 & 1 & 0 \end{bmatrix}$$

Der erste Haushalt liegt hinsichtlich der sozialen Stellung in der zweiten Kategorie, der zweite und dritte Haushalt in der dritten Kategorie, usw. Bei diesem Design ergibt die Summe der drei letzten Spalten genau die erste Spalte:

$$\text{Spalte 1} = \text{Spalte 3} + \text{Spalte 4} + \text{Spalte 5} .$$

Die Spalten der Beobachtungsmatrix, die die Variablenwerte enthalten, sind daher nicht linear unabhängig. Der Rang der Beobachtungsmatrix \mathbf{X} ist um Eins niedriger als die Anzahl ihrer Spalten, was zur Folge hat, dass die Produktmatrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar ist. Der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ ließe sich bei einem derartigen Design folglich nicht ermitteln.

Aus diesem Grund muss bei der Einbeziehung eines qualitativen Regressors in ein ökonomisches Modell wie folgt verfahren werden. Bei einem qualitativen Regressor mit h Kategorien dürfen nur $h-1$ Dummy-Variablen gebildet werden. In unserem Beispiel könnte man z.B. für die Kategorien 2 und 3 der sozialen Stellung jeweils eine Dummy-Variable bilden,

$$s_{2t} = \begin{cases} 1, & \text{falls soziale Stellung 2} \\ 0, & \text{sonst} \end{cases}$$

und

$$s_{3t} = \begin{cases} 1, & \text{falls soziale Stellung 3} \\ 0, & \text{sonst,} \end{cases}$$

womit die Nachfragefunktion

$$q_t = \beta_0 + \beta_1 x_t + \beta_2 s_{2t} + \beta_3 s_{3t} + u_t$$

lautet. Daraus ergibt sich ein Design der Beobachtungsmatrix \mathbf{X} , das eine Bestimmung des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ der unbekanntenen Regressionskoeffizienten ermöglicht. Für die betrachteten Haushalte nimmt die Beobachtungsmatrix die Gestalt

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & 1 & 0 \\ 1 & x_2 & 0 & 1 \\ 1 & x_3 & 0 & 1 \\ 1 & x_4 & 0 & 0 \\ 1 & x_5 & 1 & 0 \\ 1 & x_6 & 0 & 0 \\ 1 & x_7 & 0 & 0 \\ 1 & x_8 & 1 & 0 \end{bmatrix}$$

an, die keine lineare Abhängigkeit zwischen den Spalten der Dummy-Variablen und der Spalte der Scheinvariablen mehr enthält.

Die Interpretation der Koeffizienten β_2 und β_3 geht aus der separaten Darstellung der Nachfragefunktion über die drei Kategorien der sozialen Stellung hervor:

$$\text{für soziale Stellung 1: } q_t = \beta_0 + \beta_1 x_t + u_t$$

$$\text{für soziale Stellung 2: } q_t = (\beta_0 + \beta_2) + \beta_1 x_t + u_t$$

$$\text{für soziale Stellung 3: } q_t = (\beta_0 + \beta_3) + \beta_1 x_t + u_t .$$

Die Regressionskoeffizienten β_2 und β_3 messen also die Mehr- oder Mindernachfrage eines Haushalts in der zweiten bzw. dritten Kategorie der sozialen Stellung gegenüber einem Haushalt in der ersten Kategorie. Die soziale Stellung 1 dient hier somit als Referenzgröße. Alternativ hätte natürlich auch die soziale Stellung 2 oder 3 als Referenzgröße verwendet werden können, was bei der Interpretation entsprechend zu beachten ist.

Dummy-Variablen werden in ökonometrischen Modellen häufig bei unterjährigen Daten verwendet, um saisonale Einflüsse zu berücksichtigen. Prinzipiell erfolgt die Modellierung genauso wie bei den hier betrachteten qualitativen Variablen. Bei Quartalsdaten sind z.B. 3 und bei Monatsdaten 11 Dummy-Variablen zu bilden. Die Regressionskoeffizienten der Dummy-Variablen geben hierbei die saisonalen Ausschläge in den Einheitsperioden gegenüber einem Referenzquartal oder -monat wieder.

2.6.3 Strukturbruchtest

In einer Volkswirtschaft treten von Zeit zu Zeit Ereignisse auf, die ihre Struktur in bestimmter Hinsicht dauerhaft verändern. Nach einem Krieg, einem Ölpreisschock, Basisinnovationen oder einer ordnungspolitischen Grundsatzentscheidung können die wirtschaftlichen Rahmenbedingungen eine neue Orientierung erfahren haben, die nicht ohne Einfluss auf die Strukturparameter eines ökonometrischen Modells zu bleiben braucht. Wenn dies der Fall ist, spricht der Ökonometriker von einem Strukturbruch. Vor dem Strukturbruch in der Periode n_1+1 hatte das ökonometrische Modell

$$(2.6.4) \quad y_t = \sum_{j=1}^k \beta_j^1 \cdot x_{jt} + u_t, \quad t = 1, 2, \dots, n_1$$

Gültigkeit; danach hat das Modell die Form

$$(2.6.5) \quad y_t = \sum_{j=1}^k \beta_j^2 \cdot x_{jt} + u_t, \quad t = n_1 + 1, n_1 + 2, \dots, n$$

Für den gesamten Beobachtungszeitraum kann das ökonometrische Modell unter Verwendung zweier Dummy-Variablen d_1 und d_2 , die dem Strukturbruch Rechnung tragen sollen, wie folgt formuliert werden:

$$(2.6.6) \quad y_t = \sum_{j=1}^k \beta_j^1 \cdot x_{jt} \cdot d_{1t} + \sum_{j=1}^k \beta_j^2 \cdot x_{jt} \cdot d_{2t} + u_t, \quad t = 1, 2, \dots, n$$

mit

$$d_{1t} = \begin{cases} 1 & \text{für } t=1, 2, \dots, n_1 \\ 0 & \text{für } t=n_1+1, n_1+2, \dots, n \end{cases}$$

und

$$d_{2t} = \begin{cases} 0 & \text{für } t=1, 2, \dots, n_1 \\ 1 & \text{für } t=n_1+1, n_1+2, \dots, n. \end{cases}$$

Auf diese Weise wird der Tatsache Rechnung getragen, dass sich bei einem Strukturbruch alle Regressionskoeffizienten verändern können.

Der Ökonometriker hat nun zu beurteilen, ob ein bestimmtes Ereignis tatsächlich zu einem Strukturbruch geführt hat oder nicht. Es ist also zu testen, ob sich die ökonomischen Randbedingungen in einem Ausmaß geändert haben, das es nicht mehr erlaubt,

von einer Konstanz der Strukturparameter des ökonometrischen Modells über den gesamten Beobachtungszeitraum auszugehen. Ein statistischer Test, der eine Aussage über diesen Tatbestand ermöglicht, ist von Chow (1960) entwickelt worden.

Bei dem **Chow-Test** sind drei Regressionen durchzuführen. Zum einen werden die Regressionsgleichungen (2.6.4) und (2.6.5) für die Teilperiode vor und nach dem Strukturbruch geschätzt. Hieraus erhält man die beiden Quadratsummen der Residuen in den beiden Teilperioden:

$$\text{vor dem Strukturbruch: } Q_1 = \sum_{t=1}^{n_1} \hat{u}_t^2 \quad \text{mit } n_1 - k \text{ Freiheitsgraden}$$

$$\text{nach dem Strukturbruch: } Q_2 = \sum_{t=n_1+1}^n \hat{u}_t^2 \quad \text{mit } n_2 - k \text{ Freiheitsgraden .}$$

Außerdem wird die Regressionsgleichung

$$(2.6.7) \quad y_t = \sum_{j=1}^k \beta_j \cdot x_{jt} + u_t, \quad t = 1, 2, \dots, n$$

für die gesamte Beobachtungsperiode geschätzt, die zu folgender Quadratsumme der Residuen führt:

$$Q = \sum_{t=1}^n \hat{u}_t^2 \quad \text{mit } n-k \text{ Freiheitsgraden .}$$

Bei einer Stabilität der Strukturparameter im Beobachtungszeitraum ist zu erwarten, dass sich die Modellanpassung nicht entscheidend verbessert, wenn man die beiden Teilperioden vor und nach einem vermuteten Strukturbruch separat betrachtet. Die Quadratsumme Q der Residuen wird dann nicht wesentlich größer sein als die beiden Quadratsummen Q_1 und Q_2 zusammen genommen. Falls sich die Strukturparameter dagegen bedeutsam verändert haben, wird die Residualquadratsumme Q dagegen die beiden Quadratsummen Q_1 und Q_2 merklich übersteigen.

Darin liegt die Ratio des Chow-Tests der Nullhypothese

$$(2.6.8) \quad H_0: \beta_j^1 = \beta_j^2 \quad \text{für alle } j = 1, 2, \dots, k .$$

Die Prüfgröße des Tests ist durch das Verhältnis der Differenz der Abweichungsquadratsummen $Q - (Q_1 + Q_2)$ und der Summe der Abweichungsquadratsummen $Q_1 + Q_2$ nach Division durch die Anzahl der Freiheitsgrade gegeben:

$$(2.6.9) \quad F = \frac{[Q - (Q_1 + Q_2)] / k}{(Q_1 + Q_2) / (n - 2k)} .$$

Die Anzahl der Freiheitsgrade der Zähler- und Nennergröße ergibt sich darin aus denen der beteiligten Quadratsummen:

$$(n - k) - [(n_1 - k) + (n_2 - k)] = k \quad (\text{für Zählergröße})$$

und

$$(n_1 - k) + (n_2 - k) = n - 2k \quad (\text{für Nennergröße}) .$$

Auf einem Signifikanzniveau α ist die Nullhypothese abzulehnen, wenn die Prüfgröße F das $(1-\alpha)$ -Quantil $F_{1-\alpha;k;n-2k}$ der F-Verteilung mit k und $n-2k$ Freiheitsgraden überschreitet.

$$F > F_{k;n-2k;1-\alpha} \rightarrow H_0 \text{ ablehnen} .$$

Die Differenz zwischen den Residualquadratsummen für den gesamten Beobachtungszeitraum und den beiden Teilzeiträumen vor und nach dem vermuteten Strukturbruch lässt sich dann nicht mehr als zufällig betrachten. Vielmehr wird sie auf unterschiedliche Strukturparameter in den beiden Teilzeiträumen zurückgeführt. Aufgrund des Testergebnisses ist dann davon auszugehen, dass im Beobachtungszeitraum tatsächlich ein Strukturbruch vorliegt.

Beispiel 2.6. 1: Für 1975 gab die Deutsche Bundesbank erstmals eine Zielvorgabe in Bezug auf die Zentralbankgeldmenge bekannt. Die Zentralbankgeldmenge und später die Geldmenge M_3 lösten seitdem die freien Liquiditätsreserven und das Zinsniveau als Indikatoren und Zwischenzeile der Geldpolitik ab. Es stellt sich die Frage, ob die Umorientierung in der Geldpolitik Auswirkungen auf die Geldnachfrage im Sinne eines Strukturbruchs gezeitigt hat, was hier mit Hilfe des Chow-Tests beispielhaft geprüft werden soll.

Ausgangspunkt ist hierbei die in Beispiel 2.1. 2 OLS-geschätzte Geldnachfragefunktion für den gesamten Beobachtungszeitraum von 1970 bis 1989⁴⁹,

⁴⁹ Das hier präsentierte Schätzergebnis, das mit dem Programm EViews ermittelt worden ist, weicht rundungsbedingt mit den in den Beispielen 2.1.2 und 2.2.3 manuell berechneten Werten geringfügig ab.

$$\ln m_t = -4,741 + 1,424 \cdot \ln y_t - 0,077 \cdot \ln r_t,$$

(-10,900)
(24,903)
(-4,076)

$$R^2 = 0,978, \quad DW = 1,192,$$

für die die Residuenquadratsumme Q den Wert 0,014855 annimmt. Für den Zeitraum von 1970 bis 1974 vor dem zu überprüfenden Strukturbruch führt eine OLS-Schätzung zu der numerisch spezifizierten Geldnachfragefunktion

$$\ln m_t = -2,267 + 1,090 \cdot \ln y_t - 0,098 \cdot \ln r_t,$$

(-2,497)
(8,552)
(-4,327)

$$R^2 = 0,974, \quad DW = 2,932,$$

während sich für den Zeitraum von 1975 bis 1989 die Geldnachfragefunktion

$$\ln m_t = -5,637 + 1,544 \cdot \ln y_t - 0,080 \cdot \ln r_t,$$

(-8,385)
(17,079)
(-3,504)

$$R^2 = 0,961, \quad DW = 1,323,$$

ergibt. Die hiermit verbundenen Residualquadratsummen lauten $Q_1 = 0,000321$ und $Q_2 = 0,010708$. Mit $k=3$ und $n=20$ nimmt die Prüfgröße des Chow-Tests den Wert

$$F = \frac{[Q - (Q_1 + Q_2)]/K}{(Q_1 + Q_2)/(n - 2k)} = \frac{[0,014855 - (0,000321 + 0,010708)]/3}{(0,000321 + 0,010708)/(20 - 2 \cdot 3)}$$

$$= \frac{0,001275}{0,000788} = 1,631$$

an. Für ein Signifikanzniveau von 5% ist dieser Wert mit dem kritischen Wert

$$F_{3;14;0,95} = 3,34$$

einer F-Verteilung mit 3 Freiheitsgraden des Zählers und 14 Freiheitsgraden des Nenners zu vergleichen. Wegen

$$(F = 1,631) < (F_{3;14;0,95} = 3,34)$$

kann die Nullhypothese $H_0 : \beta_j^1 = \beta_j^2$ für $j=1,2,3$ nicht abgelehnt werden, was bedeutet, dass mit dem Wechsel der geldpolitischen Strategie nicht zugleich auch ein Strukturbruch in Bezug auf das Geldnachfrageverhalten des Publikums nachgewiesen werden kann. ♦

2.6.4 Qualitative abhängige Variablen

2.6.4.1 Qualitative Wahlhandlungsprobleme

Die ökonomische Theorie zielt hauptsächlich darauf ab, das Niveau einer abhängigen Variablen durch eine oder mehrere Einflussgrößen zu erklären. An dieser Zielsetzung hat sich schwerpunktmäßig die Entwicklung ökonometrischer Methoden orientiert. Bei einem Vorliegen qualitativer Regressoren kann das ökonometrische Instrumentarium ohne weiteres übernommen werden, sofern ihre Effekte korrekt in der Beobachtungs- oder Designmatrix \mathbf{X} berücksichtigt werden. Eine andere Situation ergibt sich, wenn ökonomische Wahlhandlungen zu erklären sind. Wie lässt sich z.B. die Entscheidung eines Individuums, seine Arbeitskraft anzubieten, ökonomisch deuten? Welche Einflussgrößen sind maßgeblich für den Entschluss eines Konsumenten für den Kauf oder Nichtkauf eines Produkts? Wodurch kann die Wahl eines bestimmten Verkehrs- oder Transportmittels ökonomisch erklärt werden? In all diesen Problemstellungen hat die zu erklärende Variable kein metrisches Messniveau, sondern diskrete Ausprägungen. Man spricht dann von **qualitativen Wahlhandlungsmodellen** (qualitative choice models oder quantal choice models), auf die die klassischen ökonometrischen Methoden nicht von vornherein anwendbar sind.

Qualitative Wahlhandlungsmodelle tragen zu einer Mikrofundierung der ökonomischen Theorie bei. Sie setzen in der Regel Mikrodaten voraus, die z.B. aus Unternehmens- und Verbraucherbefragungen stammen können. Die typischen Anwendungen zielen auf eine Querschnittsanalyse der Mikrodaten ab, wenn auch speziellen Fragestellungen durchaus im Rahmen einer Längsschnittanalyse nachgegangen werden kann. Wir stellen hier die binären Wahlhandlungsmodelle vor, bei denen es um die ökonomische Erklärung von zwei Alternativen wie z.B. Kauf oder Nichtkauf geht. Eine Verallgemeinerung bilden die multinomialen Wahlhandlungsmodelle, bei denen die zu erklärende Variable mehr als zwei diskrete Ausprägungen hat. Methodisch treten dadurch jedoch keine zusätzlichen Probleme auf, weshalb auf ihre Behandlung hier verzichtet wird.

2.6.4.2 Lineares Wahrscheinlichkeitsmodell und Logit-Modell

Zum Zwecke einer ökonometrischen Modellierung von Wahlhandlungen betrachten wir eine dichotome abhängige Variable Y , die z.B. bei dem Kauf eines Produkts oder dem Angebot seiner Arbeitskraft den Wert 1 und bei einem Nichtkauf oder fehlender Bereitschaft zu einer Arbeitspartizipation den Wert 0 annimmt. Während die Kaufentscheidung z.B. vom Einkommen, dem Konsumentenzinssatz, dem Geschlecht und der sozialen Stellung beeinflusst werden kann, sind z.B. der Lohnsatz und das Geschlecht maßgebliche Determinanten für eine Entscheidung über die Erwerbstätigkeit einer Person. Allgemein gibt der $k \times 1$ -Vektor \mathbf{x}_t die Werte der erklärenden Größen, unter denen sich auch Dummy-Variablen befinden können, für die t -te statistische Einheit wieder. Da z.B. die Kaufentscheidung nicht als deterministisch angesehen werden kann, muss der Tatsache Rechnung getragen werden, dass ein Konsument mit einem hohen Einkommen ein bestimmtes Produkt kauft, wohingegen sich ein anderer Konsument mit einem vergleichbar hohen Einkommen für einen Nichtkauf entscheidet. Wenn das Einkommen jedoch eine substantielle Einflussgröße ist, so muss bei einem nicht inferioren Gut die Kaufwahrscheinlichkeit *ceteris paribus* mit zunehmendem Einkommen steigen. Bei einem Wahlhandlungsmodell wird die Wahrscheinlichkeit dafür, dass sich ein Wirtschaftssubjekt z.B. für den Kauf eines Produkts oder eine Partizipation am Erwerbsleben entscheidet, durch die Einflussgrößen x_1, x_2, \dots, x_k erklärt:

$$P(y_t = 1) = p_t = F(\boldsymbol{\beta}' \mathbf{x}_t), t = 1, 2, \dots, n$$

mit

$$\mathbf{x}_t = (x_{1t} \quad x_{2t} \quad \dots \quad x_{kt})'$$

und

$$\boldsymbol{\beta} = (\beta_1 \quad \beta_2 \quad \dots \quad \beta_k)'$$

Der Vektor $\boldsymbol{\beta}$ enthält die unbekannt Parameter des Wahlhandlungsmodells, denen in Abhängigkeit von der Spezifikation der Verteilungsfunktion F eine unterschiedliche Interpretation zukommt. In der Regel ist x_1 eine Scheinvariable ($x_{1t}=1$ für alle t), so dass der Parameter β_1 das absolute Glied bezeichnet.

Im einfachsten Wahlhandlungsmodell verändert sich z.B. die Kaufwahrscheinlichkeit linear mit einer Veränderung der Einflussgrößen. Man erhält das **lineare Wahrscheinlichkeitsmodell**, in dem als Basisverteilung eine Gleichverteilung unterstellt wird:

$$p_t = F(\beta_1 \mathbf{x}_t) = \beta' \mathbf{x}_t .$$

Die Größe p_t gibt die Wahrscheinlichkeit dafür an, dass ein betrachtetes Ereignis wie z.B. der Kauf eines Produkts realisiert wird. Dann lässt sich p_t jedoch gleichzeitig als Erwartungswert der Zufallsvariablen y_t interpretieren:

$$(2.6.10) \quad E(y_t) = p_t = \beta' \mathbf{x}_t .$$

Die Zufallsvariable y_t setzt sich demnach aus der systematischen Komponente p_t und einer Störgröße u_t zusammen, die den möglichen Abweichungen zwischen den beobachteten und erwarteten Wahlhandlungsentscheidungen der Individuen Rechnung trägt:

$$(2.6.11) \quad y_t = p_t + u_t = \beta' \mathbf{x}_t + u_t .$$

Da y_t nur die Werte 1 und 0 annehmen kann, folgt u_t einer Zweipunktverteilung, die aus der folgenden Tabelle ersichtlich ist:

y_t	u_t	$P(u_t)$
1	$1 - \beta' \mathbf{x}_t$	$\beta' \mathbf{x}_t$
0	$-\beta' \mathbf{x}_t$	$1 - \beta' \mathbf{x}_t$

Man erkennt unmittelbar, dass der Erwartungswert von u_t gleich Null ist:

$$E(u_t) = (1 - \beta' \mathbf{x}_t) \beta' \mathbf{x}_t - \beta' \mathbf{x}_t (1 - \beta' \mathbf{x}_t) = 0 .$$

Die Varianz der Störgröße ist damit durch

$$\begin{aligned} \text{Var}(u_t) &= E(u_t^2) = (1 - \beta' \mathbf{x}_t)^2 \beta' \mathbf{x}_t + (\beta' \mathbf{x}_t)^2 (1 - \beta' \mathbf{x}_t) \\ &= (1 - \beta' \mathbf{x}_t) \beta' \mathbf{x}_t [(1 - \beta' \mathbf{x}_t) \beta' \mathbf{x}_t] \\ &= (1 - \beta' \mathbf{x}_t) \beta' \mathbf{x}_t \end{aligned}$$

gegeben. Aufgrund der Abhängigkeit der Varianz von den Beobachtungen \mathbf{x}_t sind die Störterme u_t heteroskedastisch.

Dass das lineare Wahrscheinlichkeitsmodell Interpretationsprobleme aufwerfen kann, wird bereits daran deutlich, dass z.B. bei einer OLS-Schätzung der Regressionsgleichung (2.6.11) im allgemeinen nicht gewährleistet ist, dass die systematische Komponente im $(0, 1)$ -Intervall liegt. Eine Interpretation als Wahrscheinlichkeit für das betrachtete Ereignis wäre dann nicht mehr ohne weiteres möglich. Eine OLS-Schätzung des linearen Wahrscheinlichkeitsmodells ist allerdings aufgrund der Heteroskedastizität der Störgröße gar nicht adäquat. Im Gegensatz zum multiplen Regressionsmodell bei

quantitativen zu erklärenden Variablen weiß man beim linearen Wahrscheinlichkeitsmodell von vornherein, dass die Varianz der Störgröße von den konkreten Werten der exogenen Variablen abhängt. Die Annahme der Homoskedastizität ist beim linearen Wahrscheinlichkeitsmodell damit stets verletzt. Da nicht gesichert ist, dass die systematische Komponente im $(0, 1)$ -Intervall liegt, können sich unter Umständen außerdem negative Schätzwerte für die Varianz der Störgröße ergeben, womit z.B. ein Signifikanztest nicht mehr sinnvoll durchgeführt werden könnte.

Nicht unproblematisch ist außerdem die beim linearen Wahrscheinlichkeitsmodell unterstellte Annahme einer identischen Zunahme der Wahrscheinlichkeit eines Ereignisses auf einen Impuls der exogenen Variablen unabhängig von ihrem realisierten Niveau. Denn man kann sich vorstellen, dass sich z.B. Einkommensänderungen auf sehr niedrigem oder sehr hohem Niveau anders auf die Kaufwahrscheinlichkeit auswirken als gleichwertige Veränderungen in den mittleren Einkommensbereichen. Eine derartige Situation liegt z.B. oft bei Entscheidungen über den Kauf dauerhafter Gebrauchs- oder Investitionsgüter wie z.B. bei einem PKW-Kauf oder dem Kauf von Wohnungseigentum vor.

In der empirischen Wirtschaftsforschung werden aus diesen Gründen anstelle des linearen Wahrscheinlichkeitsmodells **Probit- und Logit-Modelle** zur Erklärung ökonomischer Wahlhandlungen herangezogen, die durch die Spezifikationen

$$(2.6.12) \quad p_t = F(\boldsymbol{\beta}'\mathbf{x}_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_t} e^{-u_t^2/2} du_t \quad (\text{Probit-Modell})$$

und

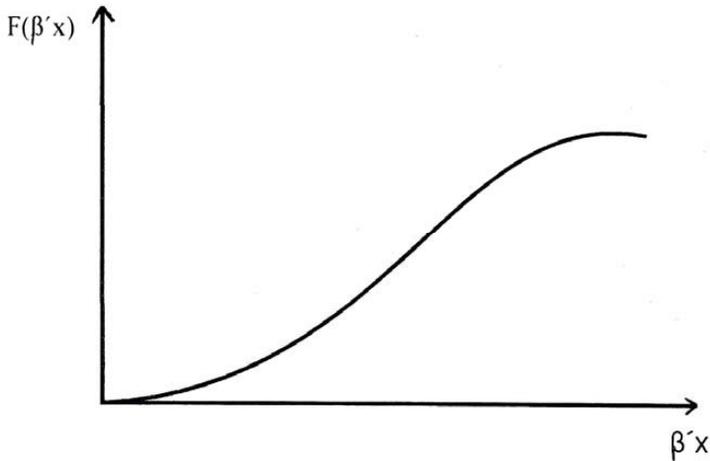
$$(2.6.13) \quad p_t = F(\boldsymbol{\beta}'\mathbf{x}_t) = \frac{1}{1 + e^{-\boldsymbol{\beta}'\mathbf{x}_t}} = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_t}}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_t}} \quad (\text{Logit-Modell})$$

der Verteilungsfunktion F gegeben sind. Während im Probit-Modell z.B. die Kaufwahrscheinlichkeit durch eine Standard-Normalverteilung gegeben ist, liegt dem Logit-Modell die Verteilungsfunktion einer logistischen Verteilung zugrunde. Die Normalverteilung lässt sich begründen, wenn man für ein Individuum t die unbeobachtbare Indexvariable

$$y_t^* = \boldsymbol{\beta}'\mathbf{x}_t + u_t$$

eingeführt, die den Grad der Neigung des Individuums z.B. für eine Kaufentscheidung oder Arbeitspartizipation erfasst. Das Individuum wird sich für den Kauf eines Gutes entscheiden,

Abbildung 2.6. 3: Wahrscheinlichkeitsverteilung des Logit-Modells



wenn die Indexvariable y_t^* eine bestimmte Schranke überschreitet, bei der der Besitz des Gutes als vorteilhaft angesehen wird⁵⁰. Wenn wir die Schranke ohne Einschränkung der Allgemeinheit gleich Null setzen, dann lässt sich die Beziehung zwischen den unbeobachtbaren Indexvariablen y_t^* und der beobachtbaren abhängigen Variablen y_t durch

$$\begin{aligned} y_t &= 1 && \text{falls } y_t^* > 0 \\ y_t &= 0 && \text{falls } y_t^* \leq 0 \end{aligned}$$

wiedergeben. Die Wahrscheinlichkeit, dass sich ein Individuum z.B. für den Kauf eines Produkts oder die Aufnahme einer Erwerbstätigkeit entscheidet, ist dann durch

$$P(y_t^* > 0) = P(\beta'x_t + u_t > 0) = P(u_t > -\beta'x_t)$$

und bei einer Symmetrie der Wahrscheinlichkeitsverteilung durch

⁵⁰ Wenn man anstelle der unbeobachtbaren Indexvariablen y_t^* einen Nutzenindex u_t einführt, lässt sich das Wahlhandlungsmodell aus einem nutzentheoretischen Ansatz entwickeln. Da der Nutzen in diesem Kontext als Zufallsvariable aufzufassen ist, ergibt sich das qualitative Wahlhandlungsmodell durch Maximierung des erwarteten Nutzens eines Individuums. Vgl. Judge et al. (1985), S. 753ff.

$$P(y_t^* > 0) = P(u_t < \beta' x_t) = F(\beta' x_t)$$

gegeben. Wenn man sich vorstellt, dass sich die Indexvariable y_t^* aus einer Vielzahl unabhängiger Einflussfaktoren zusammensetzt, kann aufgrund des Zentralen Grenzwertsatzes für F die Verteilungsfunktion einer Normalverteilung zugrunde gelegt werden. Ökonomische Wahlhandlungen lassen sich dann auf der Grundlage eines Probit-Modells untersuchen. Allerdings ist die Schätzung eines Probit-Modells recht aufwendig. Da sich die Wahrscheinlichkeitsfunktionen des Probit- und Logit-Modells in ihrer Form entsprechen, wird daher letzterem Modell bei empirischen Untersuchungen häufig der Vorzug geben. Die logistische Wahrscheinlichkeitsverteilung besitzt an den Rändern eine größere Wahrscheinlichkeitsmasse als die Normalverteilung, was vom Standpunkt der robusten Statistik durchaus als vorteilhaft zu bewerten ist. Wenn bei Anwendungen ein größerer Anteil von "outside values" als bei einer Normalverteilung erwartet werden kann, bietet das Logit-Modell eine adäquatere Grundlage zur Analyse von ökonomischen Wahlhandlungen. Aus diesem Grund werden wir das Logit-Modell in den Mittelpunkt unserer weiteren Erörterungen stellen.

Im Gegensatz zum linearen Wahrscheinlichkeitsmodell geben die Koeffizienten β_j beim **Logit-Modell** nicht mehr unmittelbar die Effekte einer Veränderung einer erklärenden Variablen auf die Wahrscheinlichkeit an. Aus der Darstellung

$$\frac{p_t}{1-p_t} = \frac{e^{\beta' x_t}}{1 + e^{\beta' x_t}} \bigg/ \frac{1}{1 + e^{\beta' x_t}} = e^{\beta' x_t}$$

folgt nach Logarithmierung die Beziehung

$$(2.6.14) \quad \ln \frac{p_t}{1-p_t} = \beta' x_t,$$

aus der ersichtlich ist, dass die Parameter β_j beim Logit-Modell die Impulse auf die sog. "log-odds ratio" $\ln(p_t/1-p_t)$ messen, die aus einer Veränderung der exogenen Größen um eine Einheit resultieren. Die "log-odds ratio" gibt die Chance für eine Realisierung des betrachteten Ereignisses im Vergleich zum Komplementärereignis wieder. Für sie hat sich die Bezeichnung "Logit" eingebürgert.⁵¹

⁵¹ Bei ausschließlich qualitativen erklärenden Variablen in Form von Dummy-Variablen wird es im Allgemeinen für einen speziellen Designvektor \mathbf{x}_t wiederholte Beobachtungen geben. Wenn man dann die Wahrscheinlichkeit p_t durch den beobachteten Anteilswert \bar{p}_t schätzt, dann lassen sich die Koeffizienten β_j aus der Logitform einer Regressionsgleichung,

Die Einflüsse einer Veränderung der erklärenden Variablen auf die Wahrscheinlichkeit des betrachteten Ereignisses lassen sich dagegen durch die partiellen Ableitungen

$$\frac{\partial p_t}{\partial x_{tj}} = \frac{\partial F(\boldsymbol{\beta}' \mathbf{x}_t)}{\partial x_{tj}} = f(\boldsymbol{\beta}' \mathbf{x}_t) \cdot \beta_j$$

angeben, in der f die Dichtefunktion der Logitverteilung bezeichnet:

$$f(\boldsymbol{\beta}' \mathbf{x}_t) = \frac{e^{\boldsymbol{\beta}' \mathbf{x}_t}}{(1 + e^{\boldsymbol{\beta}' \mathbf{x}_t})^2}.$$

Durch die Koeffizienten β_j allein ist nur noch die Richtung determiniert, in die sich die Wahrscheinlichkeiten p_t bei bestimmten Impulsen der exogenen Variablen verändern. Das Ausmaß der Veränderung hängt jedoch zusätzlich von der Steilheit der Verteilungsfunktion der Logitverteilung im Punkt $\boldsymbol{\beta}' \mathbf{x}_t$ ab. Je steiler die Verteilungsfunktion beim aktuellen Niveau verläuft, umso größer ist der Effekt einer Veränderung der exogenen Variablen auf die Wahrscheinlichkeit des betrachteten Ereignisses.

2.6.4.3 Maximum-Likelihood-Schätzung des Logit-Modells

Es ist p_t die Wahrscheinlichkeit, dass die Zufallsvariable y_t den Wert 1 annimmt, d.h., dass sich z.B. ein Konsument für den Kauf eines Produkts oder eine Erwerbsperson für die Erwerbstätigkeit entscheidet. Bei Gültigkeit des Logit-Modells ist die Wahrscheinlichkeit p_t durch (2.6.13) gegeben. Da der Vektor \mathbf{x}_t der erklärenden Variablen bei ökonomischen Anwendungen der Wahlhandlungstheorie häufig aus gemischten Variablen, d.h. qualitativen und quantitativen Variablen, besteht, sei davon ausgegangen, dass sich die Beobachtungsvektoren alle voneinander unterscheiden. Unter dieser Bedingung ist die **Maximum-Likelihood-Methode** das adäquate Schätzverfahren.

$$\ln \frac{\bar{p}_t}{1 - \bar{p}_t} = \boldsymbol{\beta}' \mathbf{x}_t + u_t,$$

ökonometrisch schätzen. Aufgrund der Heteroskedastizität der Störvariablen wäre in einer solchen Situation die verallgemeinerte Methode der kleinsten Quadrate (GLS-Methode) ein adäquates Schätzverfahren. S. hierzu z.B. Fomby, Hill und Johnson (1984), S. 344ff.

Da jede Beobachtung y_t als Realisation einer Binomialverteilung mit der Grundwahrscheinlichkeit p_t und einer Wiederholung aufzufassen ist, kann die Likelihoodfunktion L bei Unabhängigkeit der Beobachtungen in der Form

$$L = \prod_{t=1}^n p_t^{y_t} (1-p_t)^{1-y_t} = \prod_{t=1}^n \left(\frac{e^{\beta'x_t}}{1+e^{\beta'x_t}} \right)^{y_t} \left(1 - \frac{e^{\beta'x_t}}{1+e^{\beta'x_t}} \right)^{1-y_t}$$

geschrieben werden. Zum Zwecke einer Schätzung des Parametervektors β empfiehlt es sich, von der logarithmierten Likelihoodfunktion L^* auszugehen, die durch

$$\begin{aligned} L^* &= \sum_{t=1}^n y_t \ln p_t + (1-y_t) \ln(1-p_t) \\ &= \sum_{t=1}^n y_t \ln \frac{e^{\beta'x_t}}{1+e^{\beta'x_t}} + (1-y_t) \ln \frac{1}{1+e^{\beta'x_t}} \end{aligned}$$

gegeben ist. Man erhält die Bedingungen erster Ordnung für ein Maximum der Likelihoodfunktion, indem man L^* nach dem Parametervektor β differenziert:

$$\frac{\partial L^*}{\partial \beta} = \sum_{t=1}^n y_t \frac{e^{\beta'x_t}}{(1+e^{\beta'x_t})^2} x_t + (1-y_t) \frac{1}{1-p_t} \frac{-e^{\beta'x_t}}{(1+e^{\beta'x_t})^2} x_t = \mathbf{0} .$$

Nach einigen Umformungen erhält man hieraus

$$\frac{\partial L^*}{\partial \beta} = \sum_{t=1}^n (y_t - p_t) x_t = \mathbf{0}$$

so dass der Parametervektor β aus dem Gleichungssystem

$$(2.6.15) \quad \sum_{t=1}^n \cdot [y_t - p_t(\beta)] \cdot x_t = \mathbf{0} \quad \text{mit} \quad p_t(\beta) = \frac{e^{\beta'x_t}}{1+e^{\beta'x_t}}$$

bestimmt werden kann. Da (2.6.15) ein nicht-lineares Gleichungssystem in β ist, kann der unbekannte Parametervektor allerdings nicht analytisch ermittelt werden. Vielmehr ist ein iteratives Verfahren anzuwenden, das numerisch zuverlässig mit einer hohen Konvergenzgeschwindigkeit arbeitet, d.h. die ML-Lösung nach wenigen Schritten erzeugt. Ein solches Iterationsverfahren ist mit dem Newton-Raphson-Verfahren verfügbar.

Ausgehend von einem Anfangsschätzer $\boldsymbol{\beta}^{(0)}$ wird mit dem Newton-Raphson-Verfahren im m -ten Schritt ein verbesserter Parameterschätzer aus einem iterativen Schema der Form

$$(2.6.16) \quad \boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \left[\frac{\partial^2 \cdot L^*}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}} \right]^{-1} \left(\frac{\partial L^*}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}} \right)$$

ermittelt ⁵². Die Iterationsvorschrift (2.6.16) basiert auf einer Taylor-Reihenzerlegung zweiter Ordnung. Für das Logit-Modell sind der Vektor der ersten Ableitungen (= Gradient) und die Matrix der zweiten Ableitungen (= Hesse-Matrix) durch

$$\frac{\partial L^*}{\partial \boldsymbol{\beta}} = \sum_{t=1}^n \left[y_t \frac{1}{1 + e^{\boldsymbol{\beta}' \mathbf{x}_t}} - (1 - y_t) \frac{1}{1 + e^{-\boldsymbol{\beta}' \mathbf{x}_t}} \right] \cdot \mathbf{x}_t$$

und

$$\frac{\partial L^*}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{t=1}^n \frac{e^{-\boldsymbol{\beta}' \mathbf{x}_t}}{(1 + e^{-\boldsymbol{\beta}' \mathbf{x}_t})^2} \cdot \mathbf{x}_t \mathbf{x}_t'$$

gegeben (vgl. Fomby, Hill und Johnson, 1984, S. 350).

Die Hesse'sche Inverse kann zugleich als Schätzer für die Kovarianzmatrix des ML-Schätzers $\hat{\boldsymbol{\beta}}_{ML}$ verwendet werden. Ihre Diagonalelemente, die die Varianzen der Komponenten von $\hat{\boldsymbol{\beta}}_{ML}$ enthalten, lassen sich für einen Signifikanztest der geschätzten Koeffizienten des Logit-Modells verwenden.

2.6.4.4 Likelihood-Verhältnis-Test und Pseudo-R²

Die Gültigkeit des Logit-Modells überhaupt lässt sich unter Verwendung eines Likelihood-Verhältnisses überprüfen. Wenn die Nullhypothese

$$(2.6.17) \quad \beta_2 = \beta_3 = \dots = \beta_p = 0$$

gültig ist, kann die Wahrscheinlichkeit dafür, dass die Zufallsvariable y den Wert 1 annimmt, dadurch geschätzt werden, dass man die Anzahl n_1 der statistischen Einheiten,

⁵² Das Iterationsverfahren wird so lange wiederholt, bis ein Stoppkriterium erfüllt ist. So kann z.B. eine Konvergenz als praktisch gegeben angesehen werden, wenn sich die Veränderungen der Parameterschätzer $\boldsymbol{\beta}^{(m)}$ eine bestimmte Schranke unterschreiten oder sich die Werte der logarithmierten Likelihoodfunktion nicht mehr merklich verändern.

die die betrachtete Merkmalsausprägung haben, auf die Anzahl n der Beobachtungen bezieht:

$$\hat{p} = \frac{n_1}{n} \rightarrow p(y=1) = p .$$

Eine Bedingtheit durch exogene Variablen entfällt dann, da sie keinerlei Erklärungsgehalt besitzen. Unter der Nullhypothese ist die logarithmierte Likelihoodfunktion durch

$$L_0^* = \ln L_0 = n_1 \ln\left(\frac{n_1}{n}\right) + (n - n_1) \ln\left(\frac{n - n_1}{n}\right)$$

gegeben. Während L_0^* als restringierte log-Likelihoodfunktion bezeichnet wird, heißt die unter dem Logit-Modell resultierende logarithmische Likelihoodfunktion L^* unrestringiert. Unter der Nullhypothese (2.6.17) folgt die Likelihood-Ratio-Statistik

$$l = -2(L_0^* - L^*)$$

asymptotisch einer χ^2 -Verteilung mit $k-1$ Freiheitsgraden. Eine Annahme der Nullhypothese impliziert, dass exogenen Variablen x_2, x_3, \dots, x_k keinen Einfluss auf die Wahrscheinlichkeit des betrachteten Ereignisses haben, so dass ihre Schätzung allein durch den Anteilswert \hat{p} gerechtfertigt ist.

Als Maß für die Güte der Anpassung ist ein dem Determinationskoeffizienten vergleichbares Maß entwickelt worden, das aus diesem Grund als Pseudo- R^2 bezeichnet wird:

$$PR^2 = 1 - \frac{L^*}{L_0^*} .$$

Das Pseudo- R^2 kann als Gütemaß verstanden werden, das den Grad der "Unsicherheit" in den Daten misst, der durch das Logit-Modell "erklärt" wird (vgl. Judge et al., 1980, S. 602ff.). Sofern p_i stets 1 ist, wenn $y_i=1$ ist und 0, wenn $y_i=0$ ist, dann gilt für die unrestringierte log-Likelihoodfunktion $L^* = \ln L = 0$, d.h. PR^2 erreicht dann seinen maximalen Wert. In diesem Extremfall leisten die erklärenden Variablen eine perfekte Modellanpassung. Offenbar kann diese Situation jedoch nur dann eintreten, wenn $\beta' x_i$ gegen $-\infty$ oder $+\infty$ geht. Der minimale Wert des Pseudo- R^2 wird genau dann erreicht, wenn die unrestringierte log-Likelihoodfunktion mit der restringierten übereinstimmt, d.h. wenn die exogenen Variablen keinerlei Erklärungsgehalt besitzen. Zwar kann

grundsätzlich mit wachsendem PR^2 von einer verbesserten Modellanpassung gesprochen werden, doch besitzen die Werte des Pseudo- R^2 nicht die anschauliche Interpretation, die dem Determinantenkoeffizienten beim multiplen Regressionsmodell zukommt.

Beispiel 2.6. 2: Der Kauf eines Produkts wird in Abhängigkeit vom Einkommen und Geschlecht eines Konsumenten untersucht. Aus einer Querschnittserhebung hat sich folgende Datenbasis ergeben:

Konsument t	Produktkauf y_t	Einkommen x_{2t}	Geschlecht x_{3t}
1	nein (0)	2500	weibl. (1)
2	ja (1)	3000	männl. (0)
3	ja (1)	3200	männl. (0)
4	ja (1)	3500	männl. (1)
5	nein (0)	2800	weibl. (1)
6	nein (0)	3000	weibl. (1)
7	ja (1)	4000	männl. (0)
8	nein (0)	2200	weibl. (1)
9	ja (1)	3800	männl. (0)
10	nein (0)	2900	weibl. (1)
11	ja (1)	2600	weibl. (1)
12	nein (0)	2400	männl. (0)
13	ja (1)	4200	männl. (0)
14	ja (1)	3700	männl. (0)
15	nein (0)	3300	weibl. (1)
16	ja (1)	4100	weibl. (1)
17	nein (0)	2500	weibl. (1)
18	ja (1)	3900	männl. (0)
19	ja (1)	4400	männl. (0)
20	nein (0)	2000	weibl. (1)
21	nein (0)	3100	weibl. (1)
22	ja (1)	3400	männl. (0)
23	ja (1)	3500	weibl. (1)
24	ja (1)	3800	männl. (0)
25	nein (0)	2200	weibl. (1)

Die qualitativen Variablen Produktkauf und Geschlecht sind Dummy-Variablen, so dass ihre Ausprägungen mit 0 und 1 kodiert werden. Konsument 1, der das Produkt nicht gekauft hat ($y_1=0$), wird dann durch den Beobachtungsvektor $\mathbf{x}'_1 = (0 \ 2500 \ 1)$ be-

schrieben, Konsument 2, der sich für den Kauf des Produkts entschieden hat ($y_2=1$), hat den Beobachtungsvektor $\mathbf{x}'_2 = (0 \ 3000 \ 0)$. Die erste Komponente steht dabei stets für die Scheinvariable, die in allen Beobachtungsvektoren den Wert 1 annimmt.

Das Logit-Modell für den Produktkauf (y) in Abhängigkeit vom Einkommen (x_2) und Geschlecht (x_3) lautet

$$p_t = \frac{e^{\beta'x_t}}{1 + e^{\beta'x_t}} = \frac{1}{1 + e^{-\beta'x_t}}$$

mit

$$\beta' = (\beta_1 \ \beta_2 \ \beta_3) \quad \text{und} \quad \mathbf{x}'_t = (1 \ x_{2t} \ x_{3t}).$$

Da hier keine Wiederholungen vorliegen, erfolgt die Schätzung des Parametervektors β mit der Maximum-Likelihood-Methode.

Zum Zwecke der Modellschätzung kann auf ökonometrische Programmsysteme zurückgegriffen werden. Bei dem Programm EViews konvergiert das Verfahren nach 5 Iterationen zu folgender Lösung:

Variable	Regressionskoeffizient	Standardabweichung	t-Wert	Überschreitungswahrscheinlichkeit
Konstante	-11,72644	5,96009	-1,967	0,0619
Einkommen	0,00448	0,00204	2,196	0,0388
Geschlecht	-2,87446	1,83644	-1,565	0,1318

Die Überschreitungswahrscheinlichkeiten geben dabei das tatsächliche Signifikanzniveau an. Man erkennt unmittelbar, dass allein das Einkommen bei einer Irrtumswahrscheinlichkeit von 5 % einen signifikanten Einfluss auf die Kaufentscheidung ausübt. Aus dem negativen Vorzeichen des Regressionskoeffizienten der Variablen Geschlecht geht hervor, dass das Produkt tendenziell eher von Männern als von Frauen gekauft wird. Allerdings ist bei der Interpretation zu beachten, dass der Einfluss des Geschlechts auf den Produktkauf bei einem Signifikanzniveau von 5 % nicht statistisch gesichert ist.

Mit dem Maximum-Likelihood-Schätzer

$$\hat{\beta}'_{ML} = (-11,72644 \ 0,00448 \ -2,87446)$$

lässt sich auf der Grundlage von (2.6.13) für jeden Beobachtungsvektor \mathbf{x}_t die Kaufwahrscheinlichkeit p_t ermitteln. So beträgt die Kaufwahrscheinlichkeit für den Konsu-

menten 1, der sich für den Nichtkauf entschieden hat, aufgrund seiner Merkmalsausprägungen

$$p_1 = \frac{1}{1 + \exp(11,72644 - 0,00448 \cdot 2500 + 2,87446)} = 0,0323 .$$

Dagegen ordnet das Modell dem Konsumenten 2, der das Produkt gekauft hat, tatsächlich von vornherein eine hohe Kaufwahrscheinlichkeit zu:

$$p_2 = \frac{1}{1 + \exp(11,72644 - 0,00448 \cdot 3000)} = 0,8473 .$$

Eine Einkommenserhöhung von z.B. 100 DM würde dann beim Konsumenten 1 zu einer Erhöhung der Kaufwahrscheinlichkeit von

$$\begin{aligned} & \frac{\exp(-11,72644 + 0,00448 \cdot 2500 - 2,87446)}{[1 + \exp(-11,72644 + 0,00448 \cdot 2500 - 2,87446)]^2} \cdot 0,00448 \cdot 100 \\ & = 0,03560 \cdot 0,00448 \cdot 100 = 0,0159 \end{aligned}$$

führen, während sich die Kaufwahrscheinlichkeit beim Konsumenten 2 um

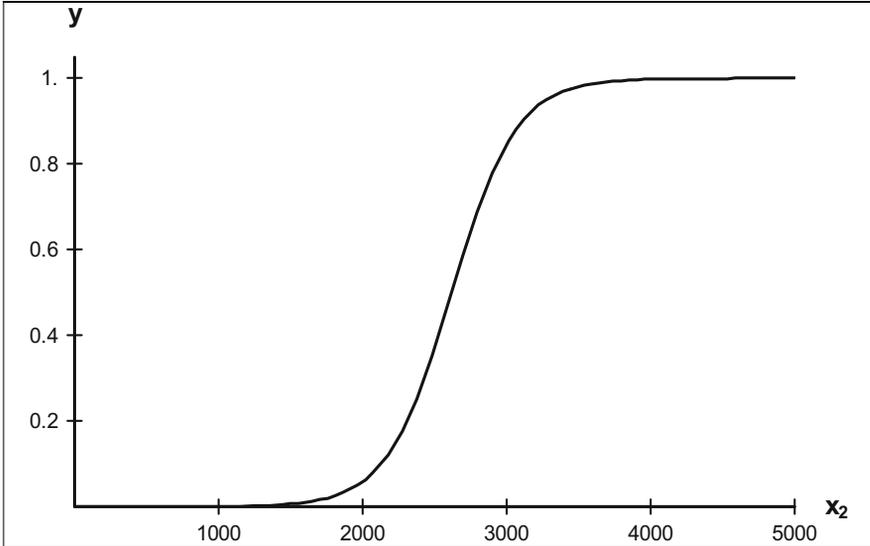
$$\begin{aligned} & \frac{\exp(-11,72644 + 0,00448 \cdot 3000)}{[1 + \exp(-11,72644 + 0,00448 \cdot 3000)]^2} \cdot 0,00448 \cdot 100 \\ & = 0,12938 \cdot 0,00448 \cdot 100 = 0,0580 \end{aligned}$$

erhöhen würde.⁵³ Die verschiedenen Zuwächse ergeben sich aus der unterschiedlichen Steilheit der Verteilungsfunktion in den betrachteten Punkten, was aus Abbildung 2.6. 4 hervorgeht.

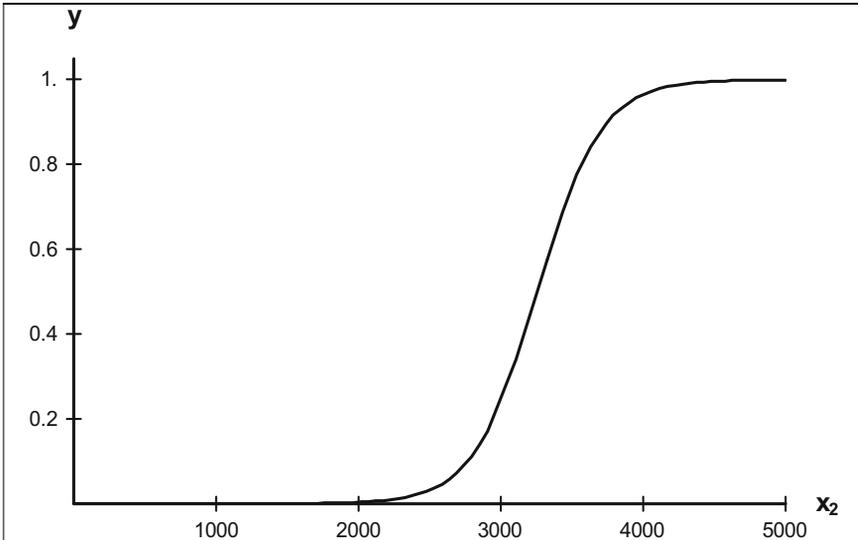
⁵³ Da die Steigung der Verteilungsfunktion des Logit-Modells streng genommen nur in einer infinitesimal kleinen Umgebung um das Ausgangseinkommen gilt, können die Veränderungen der Kaufwahrscheinlichkeiten hiermit nur approximativ bestimmt werden.

Abbildung 2.6. 4: Abhängigkeit der Kaufwahrscheinlichkeit vom Einkommen

a) bei Männern



b) bei Frauen



Schließlich wird noch der Wert der unter dem Logit-Modell gültigen logarithmierten Likelihoodfunktion L^* ausgewiesen: ⁵⁴

$$L^* = -6,072 .$$

Diese Größe wird für einen Teil des Gesamtzusammenhangs und der Berechnung des Anpassungsmaßes PR^2 benötigt. Hierzu ist außerdem eine Kenntnis des Wertes der logarithmierten Likelihoodfunktion L^*_0 erforderlich. Bei einem Stichprobenumfang von $n=25$ und einer Käuferzahl von $n_1=14$ erhält man für L^*_0 einen Wert von

$$L^*_0 = 14 \cdot \ln\left(\frac{14}{25}\right) + 11 \cdot \ln\left(\frac{11}{25}\right) = -17,148 .$$

Die Likelihood-Ratio-Statistik l nimmt damit den Wert

$$l = -2[-17,148 - (-6,072)] = 22,152$$

an. Ein Vergleich mit dem kritischen Wert

$$\chi^2_{0,95;2} = 5,99$$

zeigt, dass die Nullhypothese (2.6.17) auf einem Signifikanzniveau von 5 % eindeutig abzulehnen ist, so dass die durch das Logit-Modell hergestellte Beziehung zwischen den Kaufentscheidungen der Konsumenten und dem Einkommen und Geschlecht statistisch gesichert ist.

Für das Anpassungsmaß PR^2 ergibt sich der Wert

$$PR^2 = 1 - \frac{-6,072}{-17,148} = 0,646 ,$$

was bedeutet, dass die "Unsicherheit", die in den Daten vorhanden ist, in einem hohen Ausmaß durch das Logit-Modell vermindert werden kann. ♦

⁵⁴ L^* entspricht dem unter der Bezeichnung "Log likelihood" von EViews ausgewiesenen Wert.

Aufgaben

- (2.6.1)** Die Produktion y einer Industriebranche lässt sich in einem Quartal auf die um eine Periode verzögerten Auftragseingänge x sowie additiver saisonaler Einflüsse zurückführen. Geben Sie das zugehörige Design der Beobachtungsmatrix X für einen Zeitraum von drei Jahren wieder!
- (2.6.2)** Überprüfen Sie mit Hilfe des Chow-Tests, ob der Ölpreisschock 1980 einen Strukturbruch in der Keynes'schen Konsumfunktion (Daten s. Beispiel 2.1.1) verursacht haben könnte!
- (2.6.3)** Interpretieren Sie das Ergebnis einer OLS-Schätzung der Keynes'schen Konsumfunktion (Daten s. Beispiel 2.1.1) unter Verwendung der Dummy-Variablen d_1 (1 für 1974 bis 1979, 0 sonst) und d_2 (1 für 1980 bis 1992, 0 sonst)!
- (2.6.4)** Welche Einwände lassen sich gegen das lineare Wahrscheinlichkeitsmodell bei qualitativen abhängigen Variablen aus ökonometrischer Sicht erheben?
- (2.6.5)** Die Wahl eines Verkehrsmittels V (ÖPNV oder Pkw) für den Weg zur Arbeitsstätte soll auf der Basis eines Logit-Modells in Abhängigkeit von der Entfernung E (in km) und der beruflichen Stellung S (S_1 oder S_2) untersucht werden (Daten s. Anhang B, Tab. B2).
- Geben Sie das Logit-Modell der Verkehrsmittelwahl an und schätzen Sie es mit der Maximum-Likelihood-Methode!
 - Interpretieren Sie die Relevanz der beiden betrachteten Einflussgrößen für die Verkehrsmittelwahl!
 - Um wie viel Prozent würde sich die Wahrscheinlichkeit für die Wahl eines Pkws als Verkehrsmittel für den Weg zur Arbeitsstätte beim Berufstätigen 1 erhöhen, wenn die Entfernung um 2 km ansteigen würde?
 - Wie lässt sich die Güte der Anpassung des Logit-Modells zur Verkehrsmittelwahl quantifizieren?

2.7 Ökonometrische Prognose

2.7.1 Punktprognose

Wirtschaftspolitische Entscheidungen setzen im Allgemeinen eine Vorstellung über die zukünftige Entwicklung voraus. Ökonometrische Modelle können hierüber eine Orientierung vermitteln, wenn in ihnen die Strukturzusammenhänge adäquat abgebildet sind. Voraussetzung für eine "gute" Prognose auf der Grundlage eines ökonometrischen Modells ist jedoch eine Stabilität der Strukturparameter im Prognosezeitraum. Bei ökonometrischen Eingleichungsmodellen ist stets eine endogene Variable zu prognostizieren, während bei ökonometrischen Mehrgleichungsmodellen eine Vorhersage der zukünftigen Werte mehrerer endogenen Variablen vorzunehmen ist. Die Werte der exogenen Variablen müssen dabei vorgegeben werden. Am günstigsten ist es natürlich, wenn zur Erklärung zeitverzögerte exogene Variablen herangezogen werden können, deren Beobachtungswerte bereits vorliegen und für die Prognoseerstellung verwendet werden können. Ansonsten müssen sie selbst prognostiziert werden, z.B. unter Anwendung von Verfahren der Zeitreihenanalyse, sofern nicht auf externe Prognosen zurückgegriffen werden kann. Auf jeden Fall wird unterstellt, dass die exogenen Variablen einfacher zu prognostizieren sind als die endogenen Variablen, da der praktische Nutzen einer ökonometrischen Prognose ansonsten zweifelhaft wäre.

Eine ökonometrische Prognose ist in diesem Sinne immer eine bedingte Prognose, da sie von gegebenen Werten der exogenen Variablen ausgeht. Prognosefehler können sich daher nicht allein dadurch ergeben, dass das ökonometrische Modell die Strukturzusammenhänge unzureichend abbildet oder keine hinreichende Stabilität der Strukturparameter im Prognosezeitraum vorliegt. Vielmehr ergeben sie sich zusätzlich aufgrund einer Fehleinschätzung der Entwicklung der exogenen Variablen. Diese Problematik ist stets bei Ex ante-Prognosen zu berücksichtigen. Bei Ex post-Prognosen ist es dagegen möglich, die Modellstruktur selbst hinsichtlich ihrer Prognoseeignung zu überprüfen, indem man für einen in der Vergangenheit liegenden Prognosezeitraum auf die Beobachtungswerte der exogenen Variablen zurückgreift.

Hier soll die Prognoseerstellung für ein ökonometrisches Eingleichungsmodell erörtert werden. Bei einer Punktprognose wird der voraussichtliche Wert $\hat{y}_{n,h}$ $h = 1, 2, \dots, L$

einer endogenen Variablen y in der Periode $n+h$ unter Berücksichtigung der zurzeit n gegebenen Informationen ermittelt. Sofern $L=1$ ist, spricht man von einer Ein-Schritt-Prognose; falls $L>1$ ist, liegt eine Mehr-Schritt-Prognose vor. Allgemein gibt L die Anzahl der Prognoseperioden an und $n+L$ den Prognosehorizont.

Es seien $\hat{x}_{1,n+h}, \hat{x}_{2,n+h}, \dots, \hat{x}_{k,n+h}$ Prognosewerte für die exogenen Variablen eines ökonometrischen Eingleichungsmodells, die als gegeben betrachtet werden. Dann ist eine BLU-Prognose von y_{t+h} durch

$$(2.7.1) \quad \hat{y}_{n,h} = \sum_{j=1}^k \hat{\beta}_j \cdot \hat{x}_{j,n+h}$$

gegeben, wobei die Größen $\hat{\beta}_j$ OLS-Schätzer der unbekanntenen Regressionskoeffizienten β_j sind⁵⁵. Vereinbarungsgemäß steht der Regressor x_1 wiederum für die Scheinvariable, so dass $\hat{x}_{1,n+h}$ für alle h gleich Eins ist.

2.7.2 Intervallprognose

Man kann allerdings nicht erwarten, dass der Prognosewert $\hat{y}_{n,h}$ vollständig mit dem Zeitreihenwert y_{n+h} übereinstimmen wird. Das Prognoseprinzip stellt allein darauf ab, dass die Vorhersagen im Mittel "gut" sind, sofern das ökonometrische Modell keine Spezifikationsfehler enthält. Aus diesem Grund ist es wünschenswert, eine Kenntnis über den Unschärfebereich einer Punktprognose zu erhalten. Die Punktprognose ist dann durch eine Intervallprognose zu ergänzen.

Um das Prognoseintervall für den Wert y_{n+h} der endogenen Variablen zu entwickeln, gehen wir von der Prognosegleichung

$$(2.7.2) \quad \hat{y}_{n,h} = \mathbf{x}'_{n+h} \cdot \hat{\beta}$$

mit

⁵⁵ Der Prognosewert $\hat{y}_{n,h}$ ist eine beste lineare unverzerrte Prognose von y_{n+h} . Diese Eigenschaft lässt sich im Prinzip genau so wie beim OLS-Schätzer $\hat{\beta}$ zeigen, so dass hier auf einen Beweis verzichtet wird. Man beachte allerdings, dass durch $\hat{y}_{n,h}$ die zukünftige Realisation einer Zufallsvariablen eingeschätzt wird, während $\hat{\beta}$ ein Schätzer für den unbekanntenen Parametervektor ist.

$$\mathbf{x}_{n+h} = (x_{1,n+h} \ x_{2,n+h} \ \dots \ x_{k,n+h})'$$

bei der vorausgesetzt worden ist, dass die exogenen Variablen als gegeben betrachtet werden können. Der h-Schritt-Prognosefehler

$$(2.7.3) \quad \hat{u}_{n,h} = y_{n+h} - \hat{y}_{n,h}$$

lässt sich dann wegen

$$y_{n+h} = \mathbf{x}'_{n+h} \cdot \boldsymbol{\beta} + u_{n+h}$$

in der Form

$$(2.7.4) \quad \hat{u}_{n,h} = u_{n+h} - \mathbf{x}'_{n+h} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

schreiben. Aufgrund von

$$E(u_{n+h}) = 0 \quad \text{und} \quad E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

erhält man einen Erwartungswert des Prognosefehlers $\hat{u}_{n,h}$ von Null,

$$(2.7.5) \quad E(\hat{u}_{n,h}) = 0,$$

was bedeutet, dass die Prognose $\hat{y}_{n,h}$ unverzerrt ist ⁵⁶.

Für die Intervallprognose wird außerdem die Varianz $\hat{u}_{n,h}$ benötigt. Aufgrund der Unabhängigkeit von $\hat{u}_{n,h}$ und $\mathbf{x}'_{n+h} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ erhält man unter Verwendung von (2.7.4)

$$\begin{aligned} \text{Var}(\hat{u}_{n,h}) &= \text{Var}[u_{n+h} - \mathbf{x}'_{n+h} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] \\ \text{Var}(\hat{u}_{n,h}) &= \text{Var}[u_{n+h} - \mathbf{x}'_{n+h} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] \\ &= \text{Var}(u_{n+h}) + \text{Var}[\mathbf{x}'_{n+h} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] \\ &= \text{Var}(u_{n+h}) + \text{Var}(\mathbf{x}'_{n+h} \hat{\boldsymbol{\beta}}) \end{aligned}$$

Diese Umformung ist zulässig, weil

$$\text{Var}[\mathbf{x}'_{n+h} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] = \text{Var}(\mathbf{x}'_{n+h} \boldsymbol{\beta} - \mathbf{x}'_{n+h} \hat{\boldsymbol{\beta}}) = \text{Var}(\mathbf{x}'_{n+h} \hat{\boldsymbol{\beta}})$$

ist, da die Elemente von \mathbf{x}_{n+h} und $\boldsymbol{\beta}$ konstante Größen sind. Wenn man voraussetzt, dass die Varianz der Störvariablen in der Prognoseperiode gegenüber dem Beobachtungszeitraum unverändert bleibt, gilt

$$\text{Var}(u_{n,h}) = \sigma^2,$$

⁵⁶ Allgemeiner reicht für die Gültigkeit dieser Eigenschaft aus, dass die Werte der exogenen Variablen in der Prognoseperiode im Mittel korrekt prognostiziert werden.

so dass mit $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ sich für die Varianz des Prognosefehlers $\hat{u}_{n,h}$

$$\begin{aligned}\text{Var}(\hat{u}_{n,h}) &= \text{Var}(u_{n,h}) + \mathbf{x}'_{n,h} \text{Var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_{n,h} \\ &= \sigma^2 + \sigma^2 \cdot \mathbf{x}'_{n,h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_{n,h}\end{aligned}$$

und schließlich

$$(2.7.6) \quad \text{Var}(\hat{u}_{n,h}) = \sigma^2 \left(1 + \mathbf{x}'_{n,h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_{n,h} \right)$$

ergibt ⁵⁷.

In Anwendungen kann an Stelle der unbekanntes Störvarianz σ^2 die erwartungstreue Schätzfunktion $\hat{\sigma}^2$ verwendet werden. Bei einer Normalverteilung der Störvariablen $u_{n,h}$ ist der standardisierte Prognosefehler dann t-verteilt mit n-k Freiheitsgraden. Das Prognoseintervall ist dann durch

$$(2.7.7) \quad \mathbb{P} \left(\hat{y}_{n,h} - t_{1-\alpha/2; n-k} \cdot \sigma \sqrt{1 + \mathbf{x}'_{n,h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_{n,h}} \leq y_{n+h} \leq \hat{y}_{n,h} + t_{1-\alpha/2; n-k} \cdot \hat{\sigma} \sqrt{1 + \mathbf{x}'_{n,h} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_{n,h}} \right) = 1 - \alpha$$

gegeben. $t_{1-\alpha/2; n-k}$ ist darin das $(1-\alpha/2)$ -Quantil einer t-Verteilung mit n-k Freiheitsgraden. Das Prognoseintervall (2.7.7) besitzt eine Sicherheitswahrscheinlichkeit von $1-\alpha$.

Im Spezialfall einer einfachen Regression hat man

$$\mathbf{x}'_{n+h} = (1 \quad x_{n+h})$$

und

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{n}{n \sum x_t^2 - (\sum x_t)^2} & - \frac{\sum x_t}{n \sum x_t^2 - (\sum x_t)^2} \\ - \frac{\sum x_t}{n \sum x_t^2 - (\sum x_t)^2} & + \frac{\sum x_t^2}{n \sum x_t^2 - (\sum x_t)^2} \end{bmatrix},$$

so dass die quadratische Form in (2.7.6)

⁵⁷ Hierbei ist von der Regel $\text{Var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}' \text{Var}(\mathbf{y}) \mathbf{a}$ Gebrauch gemacht worden, wobei \mathbf{a} ein Koeffizientenvektor und \mathbf{y} ein Zufallsvektor ist.

$$\begin{aligned} \mathbf{x}'_n (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_{n,h} &= \frac{n - 2 \cdot x_{n+h} \sum x_t + x_{n+h}^2 \sum x_t^2}{n \sum x_t^2 - (\sum x_t)^2} \\ &= \frac{1}{n} + \frac{(x_{n+h} - \bar{x})^2}{\sum (x_t - \bar{x})^2} \end{aligned}$$

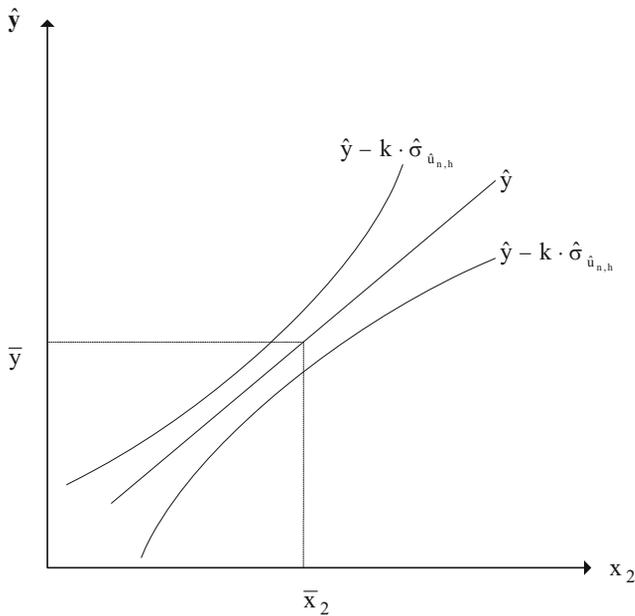
lautet. Für $k=2$ ergibt sich somit für die Varianz des Prognosefehlers der Ausdruck

$$\text{Var}(\hat{u}_{n,h}) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+h} - \bar{x})^2}{\sum (x_t - \bar{x})^2} \right),$$

womit man das Prognoseintervall

$$\begin{aligned} &P \left(\hat{y}_{n,h} - t_{1-\alpha/2; n-k} \cdot \hat{\sigma} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+h} - \bar{x})^2}{\sum (x_t - \bar{x})^2}} \right. \\ &\leq y_{n+h} \leq \hat{y}_{n,h} + t_{1-\alpha/2; n-k} \cdot \hat{\sigma} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+h} - \bar{x})^2}{\sum (x_t - \bar{x})^2}} \left. \right) \\ &= 1 - \alpha \end{aligned}$$

erhält. Daraus ist ersichtlich, dass das Prognoseintervall umso breiter wird, je mehr der Wert x_{n+h} der exogenen Variablen in der Prognoseperiode von seinem Mittelwert abweicht. Wenn die exogene Variable x trendbehaftet ist, ergibt sich somit mit zunehmendem Prognosehorizont h eine größere Ungenauigkeit der Vorhersage. Die typische Gestalt des Prognoseintervalls ist für den Fall der einfachen Regression in Abbildung 2.7. 1 wiedergegeben.

Abbildung 2.7. 1: Prognoseintervall

Beispiel 2.7. 1: Um den Privaten Verbrauch auf der Grundlage der makroökonomischen Konsumfunktion

$$\hat{C}_t = C_0 + c_1 \cdot Y_t^V$$

für das Jahr 1993 prognostizieren zu können, ist ein Kenntnis der voraussichtlichen Höhe des realen verfügbaren Einkommens erforderlich. Man könnte hierzu z.B. Prognosen der Wirtschaftsforschungsinstitute über das verfügbare Einkommen in jeweiligen Preisen und den Anstieg der Verbraucherpreise heranziehen. Hier soll das reale verfügbare Einkommen beispielhaft für das Jahr 1993 zeitreihenanalytisch mittels eines Exponentialtrends vorausgeschätzt werden. Man erhält für 1993 damit den Prognosewert 1420,8 Mrd. DM. Mit diesem Prognosewert lässt sich das reale Niveau des Privaten Verbrauchs für 1993 auf der Grundlage der im Stützzeitraum von 1974 bis 1992 geschätzten Konsumfunktion prognostizieren:

$$\hat{C}_{92,1} = 37,3 + 0,864 \cdot 1420,8 = 1264,9 \text{ .}$$

Nach dieser Prognose ist im Rezessionsjahr 1993 mit einem Rückgang des Privaten Verbrauchs in konstanten Preisen um etwa 1,7 % zu rechnen.

Hierbei handelt es sich um eine Punktprognose, die aus zeitreihenanalytischen Annahmen über die Entwicklung des verfügbaren Einkommens unter Berücksichtigung der Veränderung der Verbraucherpreise abgeleitet wurde. Die Punktprognose soll hier beispielhaft durch eine Intervallprognose ergänzt werden. Da die Stichprobe 19 Beobachtungen umfasst, ist zur Bestimmung der Intervallgrenzen bei einer Sicherheitswahrscheinlichkeit von 95 % das 0,975-Quantil der t-Verteilung mit 17 Freiheitsgraden zu verwenden:

$$t_{0,975;17} = 2,110.$$

Mit einer Residualquadratsumme von 9274,849 erhält man den erwartungstreuen Schätzer

$$\hat{\sigma}^2 = \frac{9274,849}{17} = 545,579$$

für die Störvarianz, woraus sich für den Standardfehler der Schätzung der Wert

$$\hat{\sigma} = \sqrt{545,579} = 23,358$$

ergibt. Außerdem wird noch die Summe der quadrierten Abweichungen des als Regressor fungierenden verfügbaren Einkommens von seinem Mittelwert

$$\bar{Y}^v = \frac{19866,93}{19} = 1045,63$$

benötigt:

$$\sum (Y_t^v - \bar{Y}^v)^2 = 297727,02.$$

Unter Kenntnis dieser Größen lässt sich das Prognoseintervall für den Privaten Verbrauch bestimmen:

$$\begin{aligned} & 1264,9 \pm 2,110 \cdot 23,358 \cdot \sqrt{1 + \frac{1}{19} + \frac{(1420,8 - 1045,63)^2}{297727,02}} \\ & = 1264,9 \pm 2,110 \cdot 23,358 \cdot \sqrt{1 + \frac{1}{19} + \frac{140752,53}{297727,02}} \\ & = 1264,9 \pm 60,9. \end{aligned}$$

Der praktische Nutzen des Prognoseintervalls wird häufig in Zweifel gezogen, da es im Allgemeinen zu breit ist und damit zu viel Spielraum für alternative Entwicklungstendenzen zulässt. Wirtschaftspolitische und unternehmerische Entscheidungsträger

wünschen sich außerdem häufig vom Prognostiker die Angabe eines konkreten Prognosewertes, um eine eindeutigere Orientierung zu erhalten. Obwohl das Prognoseintervall gewisse Unsicherheitsgrade widerspiegelt, wird es aus diesen Gründen bei der Veröffentlichung einer Wirtschaftsprognose nur selten ausgewiesen. Hierdurch kann leicht eine größere Präzision einer ökonomischen Prognose suggeriert werden, als sie tatsächlich aufgrund ihres stochastischen Charakters gegeben ist. ♦

2.7.3 Güte der Prognose

Vor der Anwendung eines ökonometrischen Modells zum Zwecke einer Ex ante-Prognose sollte die Prognosequalität für einen Zeitraum in der Vergangenheit überprüft werden. Auf der Grundlage von Ex post-Prognosen erhält man außerdem Informationen über die Güte konkurrierender Prognoseverfahren. Gleichmaßen bieten Ex post-Prognosen eine Möglichkeit zur Beurteilung der Prognoseeignung alternativer ökonometrischer Modelle.

Der Stützzeitraum, für den das ökonometrische Modell geschätzt wird, erstreckt sich über die Perioden 1 bis n . Der Ex post-Prognosezeitraum umfasst die Perioden $n+1$ bis maximal $n+L$. Für den Zeitraum der Ex post-Prognose sind die Werte der exogenen Variablen bekannt. Wenn sie bei der Prognose vorgegeben werden, erhält man eine Einschätzung der Prognoseleistung, die die strukturelle Form des Modells hervorzubringen in der Lage ist. Es kann jedoch zusätzlich sinnvoll sein, eine Ex post-Prognose unter Verwendung von Schätzwerten für die exogenen Variablen zu erstellen. Hieraus wird man im Allgemeinen eine realistischere Vorstellung über Leistungsfähigkeit des ökonometrischen Modells bei Ex ante-Prognosen erhalten, da die Unsicherheit der exogenen Größen berücksichtigt wird.

Man erhält ein Bild über die Prognosegüte, indem man die tatsächlichen und prognostizierten Zeitreihenwerte der endogenen Variablen y im Ex post-Prognosezeitraum gemeinsam in einem Zeitreihendiagramm einzeichnet. Darüber hinaus sind Gütekriterien verfügbar, mit denen verschiedene Aspekte der prognostischen Qualität eines ökonometrischen Modells quantitativ erfasst werden können. Als Gütekriterien kommen vor allem Prognosefehlermaße, Vergleichskriterien und Wendepunkt-Fehlermaße in Be-

tracht. Insbesondere gibt es bei den Prognosefehlermaßen eine Vielzahl von Alternativen. Wir beschränken uns hier auf einige wichtige Gütekriterien, die bei ökonometrischen Prognosen Anwendung finden.

Prognosefehlermaße geben in der Regel einen Durchschnitt der Abweichungen zwischen dem tatsächlichen Zeitreihenwert y_{n+h} und dem Prognosewert $\hat{y}_{n+h-1,1}$ in der Ex post-Prognoseperiode wieder.⁵⁸ In Anlehnung an die Form der Streuungsmaße in der deskriptiven Statistik liegt es nahe, die absoluten oder die quadrierten Abweichungen zwischen y_{n+h} und $\hat{y}_{n+h-1,1}$ als Grundelemente bei der Bildung eines Prognosefehlermaßes zu verwenden. Der **mittlere absolute Prognosefehler** (mean absolute error) wird aus den absoluten Abweichungen $|y_{n+h} - \hat{y}_{n+h-1,1}|$ gebildet,

$$(2.7.8) \quad \text{MAE} = \frac{1}{L} \sum_{h=1}^L |y_{n+h} - \hat{y}_{n+h-1,1}|,$$

während die Wurzel des **mittleren quadratischen Fehlers** (root mean square error) aus den quadrierten Abweichungen $(y_{n+h} - \hat{y}_{n+h-1,1})^2$ zusammensetzt:

$$(2.7.9) \quad \text{RMSE} = \sqrt{\frac{1}{L} \sum_{h=1}^L (y_{n+h} - \hat{y}_{n+h-1,1})^2}.$$

Beide Maße sind in der Größenordnung vergleichbar, da beim RMSE die Quadrierung der Einzelabweichungen durch das Ziehen der Wurzel aus der Summe der Abweichungsquadrate größtmäßig wieder rückgängig gemacht wird. Allerdings ist zu beachten, dass sich größere Prognosefehler erheblich stärker im RMSE als im MAE niederschlagen. Allgemein ist der mittlere absolute Fehler stets kleiner oder gleich der Wurzel des mittleren quadratischen Fehlers.

Beide Fehlermaße eignen sich für einen Vergleich alternativer Prognoseverfahren für die betrachtete endogene Variable. Häufig kommt es jedoch zusätzlich darauf an, Informationen darüber zu erhalten, ob bestimmte ökonomische Variablen besser prognostizierbar sind als andere. Um hierüber eine Aussage zu erhalten, ist anstelle eines absoluten ein relatives Prognosefehlermaß zu verwenden. Mit den relativen absoluten Fehlern $|y_{n+h} - \hat{y}_{n+h-1,1}|/y_{n+h}$ ergibt sich der **mittlere absolute prozentuale Prognosefehler** (mean absolute percentage error) in der Form

⁵⁸ Der Übersichtlichkeit halber beziehen wir uns bei der Darstellung der Prognosefehlermaße auf Ein-Schritt-Prognosen.

$$(2.7.10) \quad (2.7.10) \quad \text{MAPE} = \frac{1}{L} \sum_{h=1}^L \frac{|y_{n+h} - \hat{y}_{n+h-1,l}|}{y_{n+h}}.$$

Er gibt an, um wie viel Prozent die Prognosen im Ex post-Prognosezeitraum von den tatsächlichen Zeitreihenwerten der endogenen Variablen abgewichen sind. Ein entsprechender prozentualer Durchschnitt auf der Grundlage der Wurzel des mittleren quadratischen Fehlers ist durch

$$(2.7.11) \quad \text{RMSPE} = \sqrt{\frac{1}{L} \sum_{h=1}^L \left(\frac{y_{n+h} - \hat{y}_{n+h-1,l}}{y_{n+h}} \right)^2}$$

gegeben.

Als **Vergleichskriterium** hat der **Theil'sche Ungleichheitskoeffizient** U, der durch

$$(2.7.12) \quad U = \sqrt{\frac{\frac{1}{L} \sum (\Delta \hat{y}_{n+h-1,l} - \Delta y_{n+h})^2}{\frac{1}{L} \sum \Delta y_{n+h}^2}}$$

definiert ist, eine große Bedeutung bei der Beurteilung der Prognosegüte eines ökonomischen Modells erlangt⁵⁹. Die Größen $\Delta \hat{y}_{n+h-1,l}$ und Δy_{n+h} geben darin die prognostizierten und tatsächlichen Veränderungen der endogenen Variablen y in der Prognoseperiode h gegenüber der jeweiligen Ausgangsperiode n+h-1 an:

$$\Delta \hat{y}_{n+h-1,l} = \hat{y}_{n+h-1,l} - y_{n+h}$$

und

$$\Delta y_{n+h} = y_{n+h} - y_{n+h-1}.$$

Der Ungleichheitskoeffizient U ist stets positiv und erreicht die untere Schranke 0, wenn die tatsächliche Veränderung der endogenen Variablen y exakt prognostiziert wird. Je nach Ausmaß einer Fehlprognose kann U beliebig groß werden. Als kritisch ist dabei der Wert 1 des Theil'schen Ungleichheitskoeffizienten anzusehen. Dieser Wert wird nämlich genau dann erreicht, wenn für die Zukunft der Status quo unterstellt wird. Man bezeichnet eine solche Vorhersage, die die Vergangenheit unverändert in die Zukunft projiziert, als "naive Prognose". Wenn eine ökonometrische Prognose die zukünftige Entwicklung der endogenen Variablen besser als eine naive Prognose vorhersagt,

⁵⁹ Theil (1966), S. 26ff. Zu beachten ist, dass Theil selbst seinen Ungleichheitskoeffizienten unterschiedlich definiert (vgl. Theil, 1961, S. 30ff.). In der von uns verwendeten Form (2.7.12) hat der Theil'sche Ungleichheitskoeffizient u.E. die anschaulichste und klarste Interpretierbarkeit.

muss der Theil'sche Ungleichheitskoeffizient kleiner als 1 sein. Bei einem Ungleichheitskoeffizienten, der größer als 1 ist, kommt dem ökonometrischen Modell unter prognostischen Aspekten keinerlei Informationsgehalt zu.

Das Quadrat des Zählers des Theil'schen Ungleichheitskoeffizienten lässt sich in drei Komponenten zerlegen, aus denen die Art des Prognosefehlers hervorgeht:

$$(2.7.13) \quad F = \frac{1}{L} \sum (\Delta \hat{y}_{n+h-1,1} - \Delta y_{n+h})^2 = B + V + K$$

mit

$$B = (\overline{\Delta \hat{y}} - \overline{\Delta y})^2, \\ V = (S_{\Delta \hat{y}}^2 - S_{\Delta y}^2)$$

und

$$K = 2 \cdot (1 - r_{\Delta \hat{y}, \Delta y}) \cdot S_{\Delta \hat{y}} \cdot S_{\Delta y}.$$

Darin bezeichnen $\overline{\Delta \hat{y}}$ und $\overline{\Delta y}$ die Mittelwerte der prognostizierten und tatsächlichen Veränderungen der endogenen Variablen y innerhalb des Prognosezeitraums, $S_{\Delta \hat{y}}^2$ und $S_{\Delta y}^2$ entsprechen den Varianzen und $r_{\Delta \hat{y}, \Delta y}$ den Korrelationskoeffizienten. B gibt somit das Ausmaß der Verzerrung (Bias) an, wohingegen V und K den Varianz- und Kovarianzfehler enthalten. Die Anteile der Fehlerkomponenten am gesamten Prognosefehler F sind durch B/F , V/F und K/F gegeben.

Der Biasfehler ist gleich 0, wenn die endogene Variable im Mittel korrekt prognostiziert worden ist. Das Ausmaß der Schwankungen um den Mittelwert spiegelt sich in den beiden anderen Fehlerkomponenten wieder. Wenn zudem die Schwankungen korrekt erfasst werden, nimmt der Varianzfehler den Wert 0 an. Die Prognose enthielte dann keine systematischen Fehler, sondern allein unvermeidbare zufallsbedingte Fehler, die den Kovarianzfehler kennzeichnen. In diesem Sinne könnte man dann von einer optimalen Prognose sprechen.

Ein **Wendepunktfehler** liegt vor, wenn die prognostizierten und tatsächlichen Änderungen $\Delta \hat{y}_{n+h-1,1}$ und Δy_{n+h} ein unterschiedliches Vorzeichen besitzen. Man spricht von einem Wendepunktfehler 1. Art, wenn ein Wendepunkt vorausgesagt worden ist, der jedoch nicht eintritt. Ein Wendepunktfehler 2. Art liegt dagegen vor, wenn ein Wendepunkt eintritt, ohne vorhergesagt zu sein. Wendepunktfehler sind vor allem bei Konjunkturprognosen als kritisch zu bewerten, da sie Entscheidungsträgern falsche Sig-

nale für den Einsatz konjunktureller Maßnahmen vermitteln. Unter diesem Aspekt steht weniger das Ausmaß als vielmehr die Richtung der zukünftigen Entwicklung im Blickpunkt. Da bei Konjunkturprognosen eine Vielzahl von ökonomischen Variablen betrachtet wird, geben Wendepunkt-Fehlermaße aus diesem Grund im allgemeinen den Anteil der Wendepunktfehler einer bestimmten Art wieder. Bei der Wendepunkt-Fehlerquote 1. Art, WQ_1 , wird die Anzahl F_1 der irrtümlich vorausgesagten Wendepunkte auf die Anzahl \hat{W} der prognostizierten Wendepunkte bezogen:

$$(2.7.14) \quad WQ_1 = \frac{F_1}{\hat{W}} .$$

Dagegen wird bei der Wendepunkt-Fehlerquote 2. Art, WQ_2 , die Anzahl F_2 der nicht erkannten Wendepunkte auf die Anzahl W der tatsächlich eingetretenen Wendepunkte bezogen:

$$(2.7.15) \quad WQ_2 = \frac{F_2}{W} .$$

Eine allgemeine Wendepunkt-Fehlerquote WQ ist durch

$$(2.7.16) \quad WQ = \frac{F_1 + F_2}{\hat{W} + F_2}$$

definiert.

Alle drei Wendepunkt-Fehlermaße liegen zwischen 0 und 1. Sie sind gleich null, wenn keine Wendepunktfehler vorliegt. Die Wendepunkt-Fehlerquote 1. Art nimmt den Wert eins an, wenn keine vorausgesagten Wendepunkte tatsächlich eingetreten sind. Demgegenüber nimmt die Wendepunkt-Fehlerquote 2. Art den Wert eins an, wenn keiner der tatsächlich eingetretenen Wendepunkte prognostiziert worden ist. Schließlich wird WQ gleich eins, wenn die prognostizierten und eingetretenen Wendepunkte durchweg nicht übereinstimmen.

In der Praxis der empirischen Wirtschaftsforschung treten Wendepunktfehler 2. Art häufiger auf als Wendepunktfehler 1. Art. Prognostiker neigen offenbar dazu, länger von einer gleichgerichteten Tendenz auszugehen, als dies in der wirtschaftlichen Wirklichkeit gegeben ist.

Aufgaben

- (2.7.1) Inwiefern ist eine ökonometrische Prognose stets eine bedingte Prognose?
- (2.7.2) Worin bestehen die Unterschiede zwischen einer Ex post- und einer Ex ante-Prognose?
- (2.7.3) Wie lautet die Punkt- und Intervallprognose der Geldnachfrage für das Jahr 1990, wenn man die Werte der exogenen Variablen für das Prognosejahr als bekannt ($y_{90}=2112.3$, $r_{90}=8,48$) voraussetzt?
- (2.7.4) Mittels welcher Prognosefehlermaße lässt sich die Güte ökonometrischer Prognosen beurteilen?
- (2.7.5) Erläutern Sie die Ratio des Theil'schen Ungleichheitskoeffizienten!

2.8 Tests auf Parameterinstabilität

2.8.1 Vorbemerkungen

Die Durchführung einer Regressionsanalyse setzt implizit voraus, dass das unterstellte Modell korrekt spezifiziert ist. Dies bedeutet zum einen, dass die Störterme den Eigenschaften eines reinen Zufallsprozesses folgen, also weder autokorreliert noch heteroskedastisch sind. Des Weiteren wird eine Normalverteilung der Störgrößen vorausgesetzt. Zum anderen ist eine stabile Beziehung zwischen der endogenen Variablen und den Erklärungsgrößen gefordert. Insbesondere sollten die ermittelten Einflüsse der exogenen Variablen im Zeitablauf im Kern konstant bleiben. Wenn sich dagegen die Regressionsparameter im Zeitverlauf verändern, wären Prognosen aufgrund eines Modells mit festen Koeffizienten verzerrt. Schließlich wird vorausgesetzt, dass alle relevanten Erklärungsvariablen berücksichtigt sind und dass der funktionale Zusammenhang der Realität entspricht.

Die Tests auf Parameterinstabilität sollen überprüfen, ob diese Annahmen tatsächlich erfüllt sind und fördern auf diese Weise eine kritischere Modellspezifikation. Eine festgestellte Instabilität kann einerseits substantiell begründet sein und mag dann entsprechende Modelle notwendig machen. Zum anderen kann die Diagnose aber auch mit Fehlspezifikationen des Modells begründet sein. Werden z.B. wichtige Erklärungsvariablen vergessen, kann sich dies unter Umständen in zeitvariablen Einflüssen der restlichen Variablen niederschlagen. Diese beruhen aber nicht auf einer tatsächlichen, sondern vielmehr auf einer vermeintlichen Instabilität, die sich allein aufgrund der Fehlspezifikation ergibt. Die hier dargestellten Tests können zwischen diesen beiden Alternativen prinzipiell nicht diskriminieren. Bei einer Ablehnung der Stabilitätshypothese wird also lediglich angezeigt, dass das unterstellte Regressionsmodell auf irgendeine Weise fehlerhaft ist.

Die vorher dargestellten Tests auf autokorrelierte und heteroskedastische Störterme lassen sich bereits als Tests auf Parameterinstabilität interpretieren. Denn die Modell-

defekte können unter anderem auftreten, wenn die Regressionsparameter im Vektor β im Zeitablauf variieren oder wenn wichtige Erklärungsgrößen im Regressionsmodell fehlen. Insofern sind Autokorrelation und Heteroskedastizität stets Indizien für Fehler im systematischen Modellteil.

In diesem Kapitel testen wir auf instabile Regressionskoeffizienten im Parametervektor β , die ergänzend zum bereits diskutierten Chow-Test auf Strukturbruch anwendbar sind. Außerdem werden Tests auf die generelle Modellspezifikation dargestellt. Dabei geht es unter anderem um das Problem vergessener Variablen und die adäquate funktionale Form des Regressionszusammenhangs.

Hier werden als erstes der CUSUM- und CUSUMSQ-Test auf instabile Regressionskoeffizienten im Parametervektor β vorgestellt, die ergänzend zum bereits diskutierten Chow-Test auf Strukturbruch anwendbar sind. Außerdem werden der RESET- und der Harvey-Collier-Test als Tests auf die generelle Modellspezifikation dargestellt. Schließlich skizzieren wir den Jarque-Bera-Test auf Normalverteilung, der inzwischen Standard bei der Überprüfung der Verteilungsannahme ist.

2.8.2 CUSUM- und CUSUMSQ-Tests

Hier werden der CUSUM- und CUSUMSQ-Test erörtert, mit denen sich eine vorhandene Instabilität des Vektors der Regressionskoeffizienten β aufdecken lässt. Bei einer Ablehnung der Stabilitätshypothese kann jedoch nicht ohne weiteres gefolgert werden, dass β variiert und das Modell im übrigen korrekt ist. Denn die fehlende Konstanz der Regressionsparameter mag mit einer Fehlspezifikation des Modells begründet sein.

Wir haben bereits an früherer Stelle den Chow-Test kennen gelernt. Mit dem Verfahren wird ein möglicher Strukturbruch in der Regressionsbeziehung beurteilt, der mit einem Wechsel des Parameterregimes verbunden ist. Dabei ist der Zeitpunkt des Strukturbruchs konstant. Dieser Test lässt sich für den gesamten Parametervektor oder völlig

entsprechend nur für bestimmte Regressionskoeffizienten durchführen. Darüber hinaus ist ein analoger Test anwendbar, wenn ein Unterzeitraum eine zu geringe Anzahl an Beobachtungen aufweist. Dies könnte z.B. der Fall sein, wenn sich der vermeintliche Strukturbruch in der Nähe des Randes der Untersuchungsperiode befindet.

Als Beispiel soll der zweite Teilbereich der Untersuchungsperiode nur $n_2=1$ Beobachtung enthalten. Es lässt sich dann testen, ob einperiodige Prognosefehler signifikant von 0 verschieden sind oder äquivalent, ob die letzte Beobachtung von der gleichen Parameterstruktur wie die vorangegangenen $n_1=n-1$ Beobachtungen erzeugt worden ist. Unter der Nullhypothese ist der Strukturbruch vernachlässigbar, so dass man eine Regression für die gesamten n Beobachtungen schätzen kann. Bei Geltung der Alternative wird die letzte Beobachtung von einer anderen Parameterstruktur erzeugt, so dass ein Modell nur für die ersten $n-1$ Beobachtungen geschätzt werden kann. Die gefundene Parameterstruktur lässt sich dann für die Prognose der n -ten Beobachtung verwenden. Damit entsteht ein Prognosefehler, der durch

$$(2.8.1) \quad \tilde{u}_n = y_n - \mathbf{x}'_n \hat{\boldsymbol{\beta}}_{n-1}$$

gegeben ist. Der Parametervektor $\boldsymbol{\beta}_{n-1}$ wird mit der OLS-Methode auf der Basis der ersten $n_1 = n - 1$ Beobachtungen ermittelt, so dass

$$\hat{\boldsymbol{\beta}}_{n-1} = (\mathbf{X}'_{n-1} \mathbf{X}_{n-1})^{-1} \mathbf{X}'_{n-1} \mathbf{y}_{n-1}$$

gilt. Ferner gilt

$$\mathbf{y}_{n-1} = \mathbf{X}'_{n-1} \boldsymbol{\beta}_{n-1} + \mathbf{u}_{n-1}$$

und

$$y_n = \mathbf{x}'_n \boldsymbol{\beta}_n + u_n,$$

so dass man nach Einsetzen in (2.8.1) für den Prognosefehler

$$(2.8.2) \quad \tilde{u}_n = \mathbf{x}'_n \boldsymbol{\beta}_n + u_n - \mathbf{x}'_n \boldsymbol{\beta}_{n-1} - \mathbf{x}'_n (\mathbf{X}'_{n-1} \mathbf{X}_{n-1})^{-1} \mathbf{X}'_{n-1} \mathbf{u}_{n-1}$$

erhält. Da der Störprozess u_t die klassischen Eigenschaften erfüllt, ergibt sich für den Erwartungswert und die Varianz des Prognosefehlers

$$(2.8.3) \quad E(\tilde{u}_n) = \mathbf{x}'_n (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n-1})$$

und

$$(2.8.4) \quad \text{Var}(\tilde{\mathbf{u}}_n) = \sigma^2 (1 + \mathbf{x}'_n (\mathbf{X}'_{n-1} \mathbf{X}_{n-1})^{-1} \mathbf{x}_n),$$

wobei die unbekannte Störtermvarianz im reduzierten Beobachtungszeitraum durch

$$(2.8.5) \quad \hat{\sigma}^2 = \hat{\mathbf{u}}'_{n-1} \hat{\mathbf{u}}_{n-1} / (n_1 - k)$$

geschätzt werden kann. Unter der Nullhypothese eines unveränderten Parameterregimes ist der Erwartungswert des Prognosefehlers gleich 0. Dann lässt sich die Prüfgröße

$$(2.8.6) \quad t = \frac{y_n - \mathbf{x}'_n \hat{\boldsymbol{\beta}}_{n-1}}{\hat{\sigma} \sqrt{(1 + \mathbf{x}'_n (\mathbf{X}'_{n-1} \mathbf{X}_{n-1})^{-1} \mathbf{x}_n)}}$$

durch Division des einperiodigen Prognosefehlers durch seine geschätzte Varianz bilden, die asymptotisch t-verteilt ist mit $n_1 - k$ Freiheitsgraden. Durch diesen Test lässt sich rasch überprüfen, ob neu hinzukommende Beobachtungen mit dem bereits ermittelten Modell noch kompatibel sind.

Ein weiteres Verfahren, mit dem sich die Stabilität des Vektors der Regressionskoeffizienten überprüfen lässt, ist durch den CUSUM (cumulative sum)-Test gegeben (Brown, Durbin und Evans, 1975). Dabei werden rekursive Residuen, mithin einperiodige Prognosefehler mit konstanter Varianz verwendet. In einem ersten Schritt wird das Regressionsmodell nur für die ersten $r-1$ Beobachtungen geschätzt. Für den Zeitpunkt r ergibt sich der Vorhersagefehler

$$(2.8.7) \quad \tilde{u}_r = (y_r - \mathbf{x}'_r \hat{\boldsymbol{\beta}}_{r-1}) f_r^{-1},$$

der mit dem Faktor

$$(2.8.8) \quad f_r = \left(1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r \right)^{1/2}$$

normiert wird. Wegen (2.8.4) bewirkt die Normierung, dass u_r eine konstante Varianz von σ^2 hat. Die Prognosefehler sind daher homoskedastisch und unkorreliert, sofern die Störterme des Regressionsmodells die klassischen Eigenschaften erfüllen. Ferner haben sie unter der Nullhypothese eines konstanten Parameterregimes einen Erwartungswert von 0. Dagegen wäre ihr Erwartungswert aufgrund von (2.8.3) ungleich 0, wenn der Vektor der Regressionskoeffizienten variiert.

Bei n Beobachtungen und k Regressoren lassen sich insgesamt $n-k$ rekursive Residuen, also homoskedastische einperiodige Prognosefehler berechnen. Der erste Prognosefehler ergibt sich in der Periode $k+1$ auf der Basis eines Modells, das für die ersten k Beobachtungen geschätzt ist. Nach Ermittlung der Ein-Schritt-Prognose wird der Stütz-

zeitraum der Regression um eine weitere Beobachtung verlängert. Das Verfahren wird wiederholt, bis das Ende des Beobachtungszeitraums erreicht ist.

Der CUSUM-Test unterstellt unter der Nullhypothese einen konstanten Parametervektor β . Falls der Vektor der Regressionskoeffizienten bis zu einem bestimmten Zeitpunkt $r=k+1, \dots, n$ konstant ist und danach variiert, haben die rekursiven Residuen wegen (2.8.3) bis zum Zeitpunkt des Strukturbruchs einen Erwartungswert von 0. Ab der $(r+1)$ ten Beobachtung ist der Erwartungswert dagegen ungleich 0. Daher kann man aus der Entwicklung der kumulierten rekursiven Residuen

$$(2.8.9) \quad W_r = \frac{1}{\hat{\sigma}} \sum_{j=k+1}^r \tilde{u}_j$$

im (r, W_r) -Diagramm Informationen über einen möglichen Strukturbruch erhalten. Unter der Nullhypothese eines unveränderten Parametervektors ist $E(W_r)=0$, da die Cusums W_r Linearkombinationen der rekursiven Residuen sind. Die Nullhypothese wird daher verworfen, wenn sich die Cusums signifikant von 0 unterscheiden. Da das Regressionsmodell von konstanten Parametern ausgeht, wird die endogene Variable nach einem Strukturbruch über- oder unterschätzt. Entsprechend sind die rekursiven Residuen ab diesem Zeitpunkt systematisch positiv oder negativ. Die Cusums nehmen betragsmäßig zu und werden nach Ablauf einiger Perioden, etwa in der Periode r^* ihre Signifikanzgrenzen überschreiten. Dann wäre der Bruch auf eine oder mehrere Perioden vor r^* zu datieren. Je stärker der Strukturbruch ausgeprägt ist, desto näher liegt er am Zeitpunkt r^* , ab dem die Cusums nicht mehr vernachlässigbar sind.

Darüber hinaus werden in der Partialsumme (2.8.9) standardisierte rekursive Residuen verwendet, wobei sich σ zunächst durch die Standardabweichung der OLS-Residuen abschätzen lässt. Diese Strategie kann jedoch zu einer Überschätzung der wahren Standardabweichung führen. In der Folge liegen die Cusums näher bei 0, so dass die Güte des Tests nachlässt und höhere Fehler 2. Art zu erwarten sind. Daher kann alternativ der Schätzer nach Harvey (1975)

$$(2.8.10) \quad \hat{\sigma} = \left(\frac{1}{n-k-1} \sum_{r=k+1}^n (\tilde{u}_r - \bar{\tilde{u}})^2 \right)^{1/2} \quad \text{mit} \quad \bar{\tilde{u}} = \frac{1}{n-k} \sum_{r=k+1}^n \tilde{u}_r$$

verwendet werden, der die Standardabweichung auf der Basis der rekursiven Residuen berechnet. Im Gegensatz zum Mittelwert der OLS-Residuen kann das Mittel der rekursiven Residuen von 0 verschieden sein. Der Mittelwert der rekursiven Residuen kann bei einem Strukturbruch wachsen, wodurch die geschätzte Standardabweichung kleiner

wird. Damit wird es eventuell eher möglich, dass die Cusums bei einem Strukturbruch tatsächlich signifikant von 0 verschieden sind.

Die Prüfgröße (2.8.9) folgt keiner Standardverteilung. Der Annahmehereich der Nullhypothese lässt sich durch ein Paar linearer Funktionen im (r, W_r) -Diagramm angeben. Da die Varianz der Cusums approximativ durch $E(W_v^2) = r - k$ gegeben ist und daher mit steigendem r wächst, werden die linearen Funktionen so festgelegt, dass ihr vertikaler Abstand mit steigendem r zunimmt. Sofern die Cusums die Grenzen des Annahmehereichs durchlaufen, wird die Nullhypothese abgelehnt.

Es lässt sich zeigen, dass der Annahmehereich der Nullhypothese durch das Paar symmetrisch verlaufender Geraden

$$(2.8.11) \quad -a\sqrt{n-k} - 2a\frac{r-k}{\sqrt{n-k}} \quad \text{und} \quad a\sqrt{n-k} + 2a\frac{r-k}{\sqrt{n-k}}$$

gegeben ist. Der Parameter a wird in Abhängigkeit von der gewählten Irrtumswahrscheinlichkeit festgelegt (Brown, Durban und Evans, 1975). Die Größe a ist bei einem Signifikanzniveau von 5% gleich 0,948; bei einem Signifikanzniveau von 10% nimmt a den Wert 0,850 an.

Der CUSUM-Test ist nur eingeschränkt praktikabel, wenn die Variationen im Vektor der Regressionskoeffizienten nicht systematisch sind. In diesem Fall können sich positive und negative Prognosefehler in den Cusums saldieren, so dass das Verfahren eventuell fälschlicherweise Stabilität suggerieren kann. Ein ähnlicher Effekt ist zu erwarten, wenn im Beobachtungszeitraum mehrere Strukturbrüche vorliegen und sowohl positive als auch negative Prognosefehler verursachen. In solchen Situationen hat die Methode nur eine geringe Güte. Als Alternative zum CUSUM-Test bietet sich hier der CUSUMSQ (**cumulative sum of squares**)-Test an, der auch auf den rekursiven Residuen basiert. Die Cusums of Squares sind durch

$$(2.8.12) \quad Z_r = \frac{\sum_{j=k+1}^r \tilde{u}_j^2}{\sum_{j=k+1}^n \tilde{u}_j^2} = W_r / W_n$$

definiert. Die Cusum of Squares geben die Anteile der Summe der bis zum Zeitpunkt r quadrierten rekursiven Residuen an der gesamten Residuenquadratsumme an. Unter der Nullhypothese eines konstanten Parameterregimes folgen die Cusum of Squares einer Beta-Verteilung. Ihr Erwartungswert ist durch $E(Z_r) = (r - k)/(n - k)$ gegeben und steigt linear von 0 ($r=k$) bis auf 1 ($r=n$). Sofern die Cusums of Squares im (r, Z_r) -Dia-

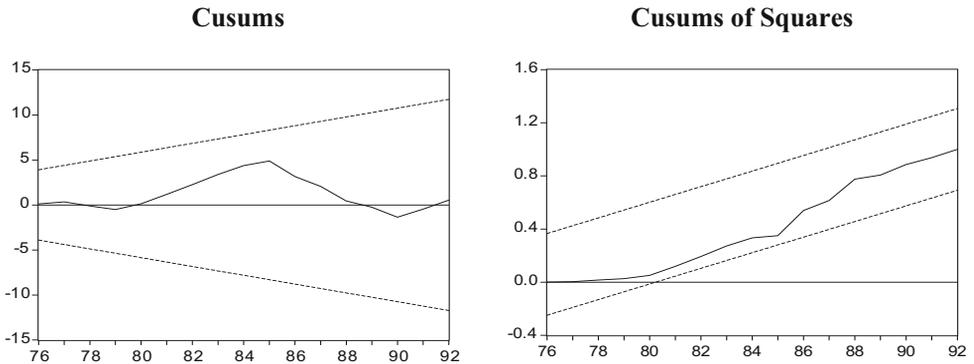
gramm signifikant von dieser Vorgabe abweichen, ist ein solches Ergebnis als Indiz gegen die Annahme stabiler Regressionsparameter zu werten.

Der Annahmereich der Nullhypothese wird durch ein Linienpaar festgelegt, welches parallel zum Erwartungswert verläuft. Die Parallelen sind durch

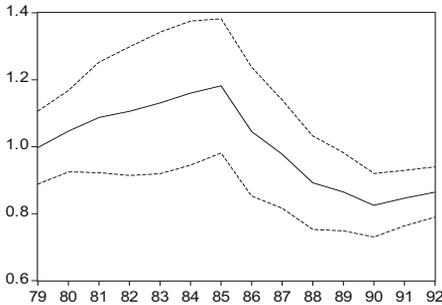
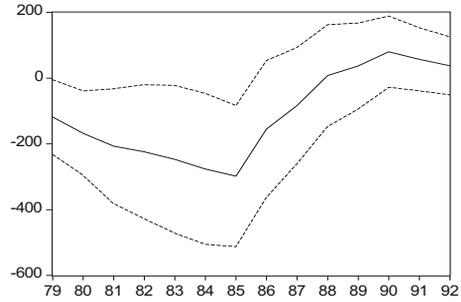
$$(2.8.13) \quad \pm c + (r - k)/(n - k)$$

identifiziert. Der Parameter c ergibt sich in Abhängigkeit vom gewählten Signifikanzniveau. Er ist bei einer Irrtumswahrscheinlichkeit von 0,05 gleich 1,115, wenn $n=100$ Beobachtungen vorliegen. Die Nullhypothese wird verworfen, sobald die Cusums of Squares den durch (2.8.13) identifizierten Annahmereich verlassen.

Beispiel 2.8. 1: Die Annahme der Parameterstabilität in der Konsumfunktion soll mit dem Cusum- und Cusum-of-Squares-Test überprüft werden. Bei einer graphischen Darstellung der Cusums und der Cusums of Squares ergibt sich



Beide Reihen verlaufen innerhalb der Signifikanzgrenzen, die für ein Signifikanzniveau von 5% ausgewiesen sind. Damit liefern die Tests kein Indiz auf Parameterinstabilität. Zusätzlich lassen sich die rekursiven Parameterschätzer für die Konstante und die marginale Konsumneigung betrachten,

Zeitvariable marginale Konsumquote**Zeitvariables absolutes Glied**

auf deren Basis die rekursiven Residuen berechnet sind, welche dann die Grundlage für die CUSUM-Tests bilden. Die durchgezogene Linie entspricht jeweils dem geschätzten Parameter, während die gestrichelten Linien den Bereich von ± 2 Standardabweichungen um den jeweiligen Schätzwert markieren. Die rekursiven Parameterschätzer weisen durchaus Fluktuationen auf. Darüber hinaus liegt die marginale Konsumneigung in Unterzeiträumen etwa bis 1985 eher über ihrem Maximalwert von 1. Führt man als ergänzende Prüfung einen Chow-Test im Zeitpunkt 1984 durch, ergibt sich ein Wert der Teststatistik von 3,88, der auf dem 5%-Niveau signifikant ist. Zum einen können die Parameter in der Konsumfunktion tatsächlich instabil sein und sich im Zeitablauf verschoben haben. Zum anderen kann eine Fehlspezifikation des Zusammenhangs für das Ergebnis verantwortlich sein.

2.8.3 RESET- und Harvey-Collier-Test

Das Problem variierender Regressionskoeffizienten, das z.B. der CUSUM- oder der CUSUMSQ-Test diagnostizieren kann, mag einerseits auf eine tatsächliche Verschiebung des Einflusses exogener Variablen deuten. Andererseits kann eine Widerlegung der Hypothese eines konstanten Parameterregimes aber auch mit einer Fehlspezifikation des Regressionsmodells begründet sein. Daher sind ergänzend Tests zu diskutieren, die auf den Modellzusammenhang abzielen.

Ein genereller Test auf die Modellspezifikation ist der RESET-Test von Ramsey (1969). Das Verfahren ist anwendbar, wenn überprüft wird, ob im gewählten Ansatz wichtige Erklärungsvariablen fehlen. Darüber hinaus lässt sich der funktionale Zusammenhang evaluieren.

Der **RESET-Test** basiert auf der Regression

$$(2.8.14) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}$$

in der \mathbf{Z} eine $n \times m$ -Matrix und $\boldsymbol{\gamma}$ ein $m \times 1$ -Parametervektor ist. Der ursprüngliche Modellzusammenhang wird mit $\mathbf{Z}\boldsymbol{\gamma}$ erweitert. Es wird getestet, ob die Variablen in der Matrix \mathbf{Z} gemeinsam signifikant sind. In einem solchen Fall ist die Prüfgröße bei normalverteilten Restgrößen allgemein F-verteilt, wobei m und $n-k$ die entsprechenden Freiheitsgrade bezeichnen. Sofern die Nullhypothese $\boldsymbol{\gamma}=\mathbf{0}$ verworfen wird, ist das ursprüngliche Modell $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ fehlspezifiziert.

Als Kandidaten für die \mathbf{Z} -Matrix werden Potenzen der geschätzten Werte der endogenen Variablen herangezogen, die man aus dem ursprünglichen Modell erhält. Damit hat man

$$(2.8.15) \quad \mathbf{Z} = (\hat{\mathbf{y}}^2, \hat{\mathbf{y}}^3, \hat{\mathbf{y}}^4, \dots).$$

Die Potenzen approximieren die potenziell unberücksichtigten Größen. Außerdem wird durch diese Strategie der funktionale Zusammenhang mit einbezogen. Die nicht potenzierten Schätzwerte der endogenen Variablen werden nicht in die \mathbf{Z} -Matrix aufgenommen, da sie sich als Linearkombination der Regressoren in \mathbf{X} ergeben.

Die funktionale Form des Modellzusammenhangs lässt sich schließlich durch das Verfahren von Harvey und Collier (1977) überprüfen. In der Nullhypothese wird eine lineare Beziehung zwischen y und dem j -ten Regressor vermutet, während der Zusammenhang unter der Alternativhypothese konvex oder konkav ist. Hierbei werden die Beobachtungen in einem ersten Schritt so umgruppiert, dass ihre neue Reihenfolge den aufsteigenden Werten des j -ten Regressors entspricht. Anschließend werden für das so transformierte Modell einperiodige Prognosefehler, mithin rekursive Residuen berechnet. Ist die wahre Beziehung zwischen y und x_j konvex, sind die $n-k$ Prognosefehler und ihr Mittelwert tendenziell positiv. Der Mittelwert ist negativ, wenn die Beziehung konkav ist. Dann würde eine lineare Prognose die tatsächliche Entwicklung überzeichnen.

Sofern die Nullhypothese zutrifft, ist der Erwartungswert der rekursiven Residuen gleich 0, so dass auch ihr absoluter Mittelwert relativ klein sein sollte. Betragsmäßig

hohe Mittelwerte führen dagegen zu einer Ablehnung der Linearitätshypothese. Da die Standardabweichung der rekursiven Residuen durch (2.8.10) gegeben ist, erhält man die Teststatistik

$$(2.8.16) \quad \mathcal{Q} = \frac{\sum_{r=k+1}^n \tilde{u}_r / (n-k)}{\left(\sum_{r=k+1}^n (\tilde{u}_r - \bar{\tilde{u}})^2 / (n-k-1) \right)^{1/2}},$$

die t-verteilt mit $n-k-1$ Freiheitsgraden ist. In der Anwendung empfiehlt es sich, den Harvey-Collier Test für verschiedene ausgewählte Regressoren durchzuführen.

Beispiel 2.8. 2: Ob die Konsumfunktion fehlspezifiziert ist, wird mit dem RESET-Test überprüft. Dazu werden Potenzen der geschätzten Werte der endogenen Variablen als weitere Regressoren einbezogen. Bei der zweiten sowie zweiten und dritten Potenz ergeben sich die empirischen F-Werte von 2,70 und 6,88. Der letztgenannte Wert ist bereits auf dem 1%-Niveau signifikant. Damit erhärtet sich der Eindruck, dass die Konsumfunktion fehlspezifiziert ist. Denkbar wäre die Verwendung einer alternativen Einkommensgröße wie z.B. das permanente Einkommen (Friedman, 1986) oder des Lebenseinkommens (Modigliani, 1986); alternativ käme eine Einbeziehung latenter Konstrukte wie z.B. das Konsumklima in Betracht (s. hierzu Kosfeld, 1996). ♦

2.8.4 Jarque-Bera-Test

Abschließend soll eine **Test von Jarque und Bera** (1980) dargestellt werden, mit dem sich die Annahme normalverteilte Störterme überprüfen lässt. Diese Annahme wird in verschiedenen Hypothesentests vorausgesetzt. Der Test basiert auf der Schiefe SK und der Wölbung K der Verteilung der Residuen. Dabei werden Abweichungen von der Normalverteilung in Form dieser beiden Größen gemessen. Tatsächlich impliziert die Normalverteilung einen Schiefeparameter von 0 und eine Wölbung von 3. Als Prüfgröße wird

$$(2.8.17) \quad JB = \frac{n-k}{6} \left(SK^2 + \frac{1}{4} (K-3)^2 \right)$$

verwendet. Dabei bezeichnet n die Anzahl der Beobachtungen und k die Anzahl der Regressoren, die zur Schätzung des Modells und damit zur Berechnung der Residuen benötigt werden. Die Teststatistik ist unter der Nullhypothese eines normalverteilten Störprozesses asymptotisch Chi-quadrat-verteilt mit 2 Freiheitsgraden.

Beispiel 2.8. 3: Die Residuen der Geldnachfragefunktion haben einen Schiefekoeffizienten von 0,226 und eine Wölbung von 2.526. Einsetzen dieser Werte in die JB-Statistik ergibt einen Wert der Prüfgröße von 0,36, der deutlich unterhalb des kritischen Wertes der Chi-Quadrat-Verteilung bei 2 Freiheitsgraden liegt. Es ist davon auszugehen, dass die Residuen der Geldnachfragefunktion normalverteilt sind. ♦

Aufgaben

- (2.8.1) Testen Sie die Geldnachfragefunktion (Daten siehe Beispiel 2.1.2) auf der Basis des CUSUM-Tests auf Parameterinstabilität ($\alpha=0,05$)!
- (2.8.2) Inwiefern lässt sich der RESET-Test als Spezifikationstest interpretieren?
- (2.8.3) Wenden Sie den RESET-Test auf die Geldnachfragefunktion (Daten siehe Beispiel 2.1.2) an ($\alpha=0,05$)!
- (2.8.4) Lässt sich eine Normalverteilung der Störgröße der Keynes'schen Konsumfunktion (Daten siehe Beispiel 2.1.1) auf dem 1%-Niveau aufrechterhalten?

2.11 Robuste Regression

2.11.1 Begriff der Robustheit

Die Optimalitätseigenschaften ökonometrischer Schätzmethoden sind nur dann gegeben, wenn die zugrunde liegenden Modellannahmen in der Realität erfüllt sind. Insbesondere muss für inferenzstatistische Zwecke in der Regel eine Normalverteilung der Störgrößen unterstellt werden, was unter bestimmten Bedingungen durch den Zentralen Grenzwertsatz der Statistik begründet werden kann. Ob seine Gültigkeit allerdings vorausgesetzt werden kann, hängt vor allem davon ab, ob die Störgrößen als gleichwertig und unabhängig voneinander angesehen werden können. Das Problem der Autokorrelation verdeutlicht, dass sich die Unabhängigkeitsannahme keinesfalls von vornherein als gegeben unterstellen lässt. Ausreißer, die z.B. durch Streiks, Krisen, Maßnahmen, Sondereinflüsse oder grobe Fehler auftreten, können eine Dominanz bestimmter Störgrößen zur Folge haben, die der Vorstellung einer Gleichwertigkeit von Zufallseinflüssen im Wege stehen. Die Verteilung der Störvariablen wird dann oftmals "schwerere Enden" (heavier tails) haben als im Falle einer Normalverteilung. Optimalitätseigenschaften wie z.B. das Gauß-Markow-Theorem gelten dann nur noch für eine eng begrenzte Klasse von Schätzern, nicht jedoch mehr generell, wie es unter einer Normalverteilung der Fall wäre.

In der Praxis werden ökonomische Daten nie vollständig mit den Modellannahmen konform sein. Allein aufgrund von Erhebungsfehlern ist nicht zu erwarten, dass ein Verteilungsmodell wie z.B. die Normalverteilung ohne weiteres unterstellt werden kann. Man bezeichnet die Daten dann als verschmutzt oder kontaminiert. Kritisch ist die Frage einzuordnen, inwiefern Kontaminationen eine Inferiorität ökonometrischer Verfahren hervorrufen können. Ein ökonometrisch-statistisches Verfahren kann, muss aber nicht notwendig sensitiv auf Datendefekte und Modellabweichungen reagieren.

Ein Schätz- oder Testverfahren heißt **robust**, wenn es **bei Modellabweichungen** weiterhin im Kern **zuverlässig** arbeitet. Dies ist aber gerade bei klassischen Schätzverfahren häufig nicht gegeben. Wie Tukey (1960) gezeigt hat, kann bei einer unmodifizierten Kleinst-Quadrate-Schätzung ein geringer Ausreißeranteil ausreichen, um die Schätzung inferior werden zu lassen. Ein robustes Schätzverfahren orientiert sich dagegen an der Masse der Daten (Wahrendorf, 1980; Brachinger, 1981; Hampel et al., 1986). Der Ein-

fluss atypischer Beobachtungen ist dabei begrenzt und geht mit zunehmender Devianz gegen Null. Aus dieser Sicht ließe sich ein robustes Schätzverfahren als ein klassisches Schätzverfahren in Verbindung mit einer Ausreißeranalyse interpretieren.⁶⁷

Die Inferiorität eines statistischen Verfahrens bei Modelldefekten zeigt sich in Form von Effizienzverlusten. Klassische Schätz- und Testverfahren weisen vor allem bei Ausreißern oft hohe Effizienzverluste auf, da von ihnen unkontrollierte Effekte ausgehen können. Eine Schätzung wird dadurch leicht in die Richtung eines oder mehrerer Ausreißer "verzerrt", so dass sie nicht mehr die Struktur der Masse der Beobachtungswerte wiedergibt. Wenn man aus der Stichprobe die Ausreißer entfernen würde, könnte sich ein völlig unterschiedliches Bild ergeben. Ein robustes Verfahren zielt dagegen darauf ab, atypische Beobachtungswerte bei einer Parameterschätzung oder einem Hypothesentest in dem Sinne unter Kontrolle zu halten, dass die Ergebnisse durch sie nicht beliebig verzerrt werden können. Bei unterschiedlichen Stichproben liegen z.B. robuste Schätzer im Falle von Datenkontaminationen nicht so weit auseinander wie die klassischen Schätzer. Robuste Schätzer besitzen in solchen Situationen im Allgemeinen eine höhere Effizienz, was eine größere Zuverlässigkeit der Schätzung bedeutet.

Robuste Verfahren stellen jedoch nicht ein Allheilmittel dar, das ohne weiteres eine gute Datenanalyse gewährleistet. Vielmehr ist bei ihrer Anwendung oft ein Vergleich mit einer klassischen Schätzung wie z.B. der OLS-Methode sinnvoll, um eventuelle Modelldefekte aufspüren zu können. Wenn die Schätzungen zu ähnlichen Ergebnissen führen, wird man häufig von einer zuverlässigen Interpretationsbasis ausgehen können. Andernfalls lassen sich vielleicht außergewöhnliche Ereignisse identifizieren, auf die die Unterschiede zurückgeführt werden können. Es ist dann die Aufgabe des Ökonometrikers, eine entsprechende Bewertung vorzunehmen. Dabei kann im Einzelfall durchaus ein singuläres Ereignis im Vordergrund der ökonometrischen Analyse stehen, wenn es z.B. darauf ankommt, die Effekte einer wirtschaftspolitischen Maßnahme auf bestimmte ökonomische Variablen einzuschätzen. Wenn die Datenstruktur dagegen im Vordergrund steht, werden Ausreißer dagegen häufig problematisch sein, da sie in der Regel nicht als Gleichgewichtsdaten im Sinne der ökonomischen Theorie interpretiert werden können. Robuste Verfahren können dann als Bereicherung des ökonometrischen

⁶⁷ Ein Ausreißer kann einen verzerrenden Einfluss auf die Schätzung eines ökonometrischen Modells haben, es muss jedoch nicht notwendig so sein. Aus diesem Grunde stehen einflussreiche Beobachtungen im Blickpunkt des Interesses, die durch eine spezielle Form der Ausreißeranalyse, einer Einflussdiagnostik, ermittelt werden können. Sie basiert auf dem Konzept der Einflussfunktion, das von Hampel (1974) entwickelt worden ist. Zu einer Einflussdiagnostik speziell für die Regressionsanalyse und ökonometrische Modelle s. Belsley, Kuh und Welsch (1980).

Instrumentariums angesehen werden, sofern von ihnen sachgerecht Gebrauch gemacht wird. Aus diesem Grund stellen wir zwei alternative robuste Schätzverfahren für ökonomische Eingleichungsmodelle vor, die einen Eindruck über die Prinzipien und Arbeitsweisen robuster Verfahren vermitteln sollen.

2.11.2 Verallgemeinerte Maximum-Likelihood-Schätzung (M-Schätzung)

Gegeben ist das multiple Regressionsmodell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

das sich für die Periode t durch

$$y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$$

mit

$$\mathbf{x}_t = (1 \quad x_{2t} \quad x_{3t} \quad \cdots \quad x_{kt})'$$

darstellen lässt. Unter der Annahme einer Normalverteilung der Störvariablen u_t erhält man mit der gewöhnlichen Methode der kleinsten Quadrate (OLS-Methode) und der Maximum-Likelihood-Methode (ML-Methode) einen identischen Schätzer für den Parametervektor $\boldsymbol{\beta}$. Um einen robusten M-Schätzer für das multiple Regressionsmodell zu entwickeln, unterstellen wir zunächst einmal tatsächlich eine Normalverteilung der Störvariablen u_t . Ihre Dichtefunktion ist dann durch

$$f_t(u_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{u_t^2}{\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y}_t - \mathbf{x}'_t \boldsymbol{\beta})^2\right]$$

gegeben. Die ML-Schätzung basiert nun auf der Likelihood-Funktion

$$L(\boldsymbol{\beta}, \sigma^2) = \prod_{t=1}^n f_t(u_t) = \frac{1}{(2\pi)^{n/2} \sigma^n} \prod_{t=1}^n \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y}_t - \mathbf{x}'_t \boldsymbol{\beta})^2\right],$$

die in logarithmierter Form

$$\ln L(\boldsymbol{\beta}, \sigma^2) = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{t=1}^n (\mathbf{y}_t - \mathbf{x}'_t \boldsymbol{\beta})^2$$

lautet. Wenn wir der Einfachheit halber die Störvarianz σ^2 zunächst einmal als gegeben betrachten, erhält man den ML-Schätzer für β offenbar durch Minimierung der Funktion

$$(2.11.1) \quad L^*(\beta) = \sum_{t=1}^n (y_t - x_t' \beta)^2,$$

die von den quadrierten Differenzen zwischen den Beobachtungswerten y_t der endogenen Variablen und den Regressionswerten \hat{y}_t abhängig ist. Angenommen, die Datenbasis besteht nicht ausschließlich aus "guten" Beobachtungswerten, sondern sie ist mit Ausreißern kontaminiert.⁶⁸ Dann hätten atypische Beobachtungen, die konträr zur Masse der Daten liegen, einen außergewöhnlich großen Einfluss auf die Lage der Regressionsgeraden oder -ebene, da ein großes Abweichungsquadrat nur durch viele kleinere Abweichungsquadrate gewöhnlicher Beobachtungen ausgeglichen werden kann. Die Regressionskoeffizienten würden bei einer ML-Schätzung dadurch einseitig in die Ausreißerrichtung verzerrt werden.

Bei der **verallgemeinerten Maximum-Likelihood-Schätzung** (M-Schätzung) verzichtet man auf eine Optimalität im Falle einer exakten Normalverteilung zugunsten einer hohen Effizienz der Schätzung in einer breiten Modellumgebung. Man erreicht dies, indem man anstelle der quadrierten Abweichungen $(y_t - x_t' \beta)^2$ alternative Funktionen ρ der Störvariablen u_t betrachtet, die weniger empfindlich gegenüber Ausreißern sind. Die modifizierte Likelihood-Funktion (2.11.1) nimmt damit die Form

$$(2.11.2) \quad L^*(\beta) = \sum_{t=1}^n \rho(y_t - x_t' \beta)$$

an, aus der ein verallgemeinerter Maximum-Likelihood-Schätzer (M-Schätzer) für den Parametervektor gewonnen werden kann (vgl. hierzu Li, 1983, S. 291ff.; Heiler, 1980, S. 39ff.). Sofern die Funktion ρ differenzierbar ist, erhält man mit $\rho'(u) = \psi(u)$ einen **M-Schätzer** für β aus der Beziehung

⁶⁸ Modellmäßig kann man sich eine kontaminierte Datenbasis so vorstellen, dass der Anteil $1-\varepsilon$ der "guten" Beobachtungen aus einer Normalverteilung Φ mit einer Varianz von σ^2 und der Anteil ε der atypischen Beobachtungen aus einer Normalverteilung Φ mit einer Varianz von $k\sigma^2$ stammt:

$$F(u) = (1-\varepsilon) \cdot \Phi(u|0, \sigma^2) + \varepsilon \cdot \Phi(u|0, k \cdot \sigma^2), \quad k > 1.$$

Es handelt sich dabei um ein Modell einer "Mischverteilung", die Ausreißer erzeugen kann. Die Mischverteilung F repräsentiert eine Wahrscheinlichkeitsverteilung der Störgröße u mit einer größeren Wahrscheinlichkeitsmasse in den äußeren Bereichen (vgl. hierzu Wahrendorf, 1980, S. 3f., und Brachinger, 1981, S. 330f.).

$$(2.11.3) \quad \sum_{t=1}^n \psi(y_t - \mathbf{x}'_t \boldsymbol{\beta}) \mathbf{x}_t = \mathbf{0} ,$$

die die notwendige Bedingung für einen Extremwert wiedergibt. Für $\rho(u_t) = u_t^2/2$ ist $\psi(u_t) = u_t$, so dass durch (2.11.2) und (2.11.3) der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ definiert ist, der mit dem Maximum-Likelihood-Schätzer für $\hat{\boldsymbol{\beta}}_{ML}$ übereinstimmt, wenn die Störgrößen einer Normalverteilung folgen. Eine solche Wahl der ψ -Funktion erfolgt jedoch im Rahmen einer robusten Schätzung nur für die "guten" Daten. Für einen robusten M-Schätzer $\boldsymbol{\beta}_M^*$ ist daher eine geeignete Form der ψ -Funktion zu wählen.

Anschaulich kann ein M-Schätzer für den Parametervektor $\boldsymbol{\beta}$ als modifizierter Kleinst-Quadrate-Schätzer interpretiert werden. Wenn man

$$(2.11.4) \quad w_t = \frac{\psi(y_t - \mathbf{x}'_t \boldsymbol{\beta})}{y_t - \mathbf{x}'_t \boldsymbol{\beta}}$$

setzt, geht (2.11.3) in die Form

$$(2.11.5) \quad \sum_{t=1}^n w_t \cdot (y_t - \mathbf{x}'_t \boldsymbol{\beta}) \mathbf{x}_t = \mathbf{0}$$

über. Die Größen w_t lassen sich hierin als Gewichte der Beobachtungen einer Regression der endogenen Variablen y auf die exogenen Variablen x_1, x_2, \dots, x_k deuten. Der Unterschied gegenüber der Bestimmungsgleichung eines OLS-Schätzers für $\boldsymbol{\beta}$ besteht allein darin, dass die Gewichte w_t in (2.11.5) allgemein von 1 verschieden sein können. Ein robuster M-Schätzer $\boldsymbol{\beta}_M^*$ ist somit ein **gewichteter Kleinst-Quadrate-Schätzer** des Parametervektors $\boldsymbol{\beta}$, der jedoch iteriert werden muss, um die Gewichte w_t nach und nach zu verbessern. Bei einer Ein-Schritt-Lösung des Gleichungssystems

$$(2.11.6) \quad \sum_{t=1}^n w_t y_t \mathbf{x}_t = \sum_{t=1}^n w_t \mathbf{x}'_t \boldsymbol{\beta} \mathbf{x}_t ,$$

das sich durch Umformung von (2.11.5) ergibt, entsprechen die Gewichte w_t der Beobachtungen im Allgemeinen noch nicht denen einer robusten Anpassung. Man bezeichnet den Ein-Schritt-Schätzer für $\boldsymbol{\beta}$ als **W-Schätzer**. Der **M-Schätzer** kann dann als **vollständig iterierter W-Schätzer** interpretiert werden. Diese Methode zur Bestimmung eines M-Schätzers für den β -Faktor ist unter dem Namen "iteratively reweighted least-squares" bekannt.⁶⁹

⁶⁹ Sie geht auf Holland und Welsch (1977) zurück. Vgl. auch Li (1983), S. 306f.

Bisher ist die Residualvarianz σ^2 der Einfachheit halber als gegeben angenommen worden. Tatsächlich ist sie jedoch ebenfalls aus den ökonomischen Daten zu schätzen. Während der OLS-Schätzer für den Parametervektor β skaleninvariant ist,⁷⁰ gilt dies für die robusten Schätzfunktionen mit Ausnahme des "least absolute residuals"-Schätzers (LAR-Schätzer) nicht. Es sei σ^{*2} ein robuster Skalenschätzer für die Residualvarianz σ^2 . Dann ist der W-Schätzer für den Vektor β der Regressionskoeffizienten aus der Beziehung

$$\sum_{t=1}^n \psi\left(\frac{y_t - \mathbf{x}'_t \beta}{\sigma^*}\right) \mathbf{x}_t = \mathbf{0}$$

zu bestimmen, in der die Residuen

$$u_t = y_t - \mathbf{x}'_t \beta$$

im Unterschied zu (2.11.3) als Vielfaches des robusten Skalenschätzers σ^* gemessen werden. Für den Ein-Schritt-Schätzer β bietet sich z. B. der Median der absoluten Abweichungen (MAD) nach einer Adjustierung als robuster Skalenschätzer an. Eine Adjustierung ist erforderlich, um Niveauunterschiede zwischen dem MAD und der Standardabweichung auszugleichen.⁷¹ Der MAD ist einfach zu berechnen, und er ist aufgrund seines hohen Bruchpunkts von nahezu 50 % äußerst resistent gegenüber Ausreißern.⁷² Bei einer Iterierung des W-Schätzers kann die Residualvarianz σ^2 jeweils aus den Residuen der robusten Anpassung geschätzt werden.

Um den Einfluss von Ausreißern auf die Schätzung zu begrenzen, hat Huber (1964) die Verwendung der ψ -Funktion

$$\psi(u_t^*) = \begin{cases} u_t^* & \text{für } |u_t^*| \leq k \\ k \cdot \text{sgn}(u_t^*) & \text{für } |u_t^*| > k \end{cases}$$

vorgeschlagen⁷³, womit sich die Gewichtsfunktion

⁷⁰ Der Begriff der Skaleninvarianz stammt aus der Terminologie der robusten Statistik, in der die Varianz bzw. Standardabweichung als Skalenparameter bezeichnet werden.

⁷¹ Unter einer Standardnormalverteilung entspricht der MAD dem 0,6745-fachen der Standardabweichung.

⁷² Der Bruchpunkt gibt den Anteil von Ausreißern an, der ausreicht, um einen Schätzer über alle Grenzen verändern zu können. Vgl. Barnett und Lewis (1978), S. 69f.; Brachinger (1981), S. 333.

⁷³ Bei der Definition der ψ -Funktion ist von der Vorzeichenfunktion $\text{sgn}(u^*)$ Gebrauch gemacht worden, die bei positivem u gleich 1, bei negativem u gleich -1 und für u^* gleich null ebenfalls 0 ist.

$$w_t = \begin{cases} 1 & \text{für } |u_t^*| \leq k \\ k/|u_t^*| & \text{für } |u_t^*| > k \end{cases}$$

ergibt. Die Größe u_t^* steht darin für die standardisierten Residuen:

$$u_t^* = (y_t - \mathbf{x}'_t \boldsymbol{\beta}) / \sigma^* .$$

Beim Huber-Schätzer wird dem Problem Rechnung getragen, dass deviante Beobachtungen aufgrund ihrer großen Residuen unter Umständen bei einer OLS-Schätzung einen größeren Einfluss auf die Größenordnung der Regressionskoeffizienten haben als eine Vielzahl konformer Beobachtungen gleichzeitig. Dies erfolgt dadurch, dass von einer bestimmten Grenze k der standardisierten Residuen ab die Gewichte der Beobachtungen verringert werden. Auf diese Weise wird sichergestellt, dass die Regressionskoeffizienten nicht einseitig von außergewöhnlichen Beobachtungen bestimmt werden. Der Parameter k hat dabei die Funktion eines Steuerungsparameters. Je größer (kleiner) k ist, umso stärker (schwächer) muss die Devianz einer Beobachtung sein, um als Ausreißer eingestuft zu werden. Für $k \rightarrow \infty$ geht der M-Schätzer von Huber in den ML-Schätzer $\hat{\boldsymbol{\beta}}$ über.

Der Huber-Schätzer ist jedoch nur als eine Möglichkeit zu verstehen, einen M-Schätzer zu operationalisieren. Als Alternativen hierzu kommen der Hampel-Schätzer, Tukeys biweight und Andrews wave in Betracht, die sich von ihrer Struktur her ähnlich sind (s. hierzu z.B. Li, 1983, S. 293ff.). Sie unterscheiden sich von Hubers Schätzer dadurch, dass bei ihnen der Einfluss von Ausreißern mit zunehmender Entfernung von einem bestimmten Punkt ab stetig auf null zurückgeführt wird. Auf eine Diskussion der unterschiedlichen Varianten der M-Schätzer wird hier verzichtet, da bei ökonomischen Anwendungen unseres Erachtens nach ein anderer robuster Schätzansatz eine bessere Kontrolle von Ausreißern gewährleistet.

2.11.3 Die Reweighted-Least-Squares-Methode (RLS-Methode)

Anhand der verallgemeinerten Maximum-Likelihood-Methode lässt sich das Prinzip einer robusten Schätzung anschaulich erläutern. Die verallgemeinerte Maximum-Likelihood-Methode trägt Ausreißern in der abhängigen Variablen dadurch Rechnung, dass sie ihren Einfluss auf die Modellschätzung begrenzt oder ausschaltet. Eine Kontrolle

derartiger Ausreißer ist somit ohne weiteres gegeben. Allerdings können Ausreißer nicht nur in der endogenen Variablen auftreten, sondern ebenfalls in den exogenen Variablen. Vor solchen extremen Werten bietet die M-Schätzung keinen Schutz.⁷⁴ In dieser Situation ist die Reweighted-Least-Squares-Methode (RLS-Methode) von Rousseeuw (1984) bzw. Rousseeuw und Leroy (1987) vorzuziehen, die beiden Ausreißertypen gerecht wird.

Mit der **Reweighted-Least-Squares-Methode** wird ein Schätzer für den Parametervektor β durch Minimierung der Kriteriumsfunktion

$$(2.11.7) \quad W(\beta) = \sum_{t=1}^n w_t (y_t - x_t' \beta)$$

bestimmt (Rousseeuw und Leroy, 1987, S. 14ff. und S. 44ff.). Die Gewichte w_t der Beobachtungswerte hängen von ihrer Lage im k -dimensionalen Variablenraum ab. Je mehr sich eine Beobachtung von der Korrelationsstruktur der Masse der Daten entfernt, um so eher wird ihr Gewicht bei Null liegen. "Gute" Beobachtungen erhalten dagegen ein Gewicht von Eins, sofern eine Normierung der Gewichte auf dem Intervall $[0; 1]$ erfolgt ist.

Prinzipiell lässt sich die RLS-Methode mit einem stetigen Gewichtungsschema durchführen. Wenn jedoch Ausreißer aus einer robusten Modellschätzung identifiziert und ausgeschlossen werden sollen, dann bietet sich eine 0,1-Gewichtung an. Rousseeuw und Leroy (1987, S. 17) schlagen vor, eine Beobachtung mit 1 zu gewichten, sofern ihr Residuum nach einer Standardisierung absolut die Schranke 2,5 nicht überschreitet; andernfalls erhält sie das Gewicht 0:

$$(2.11.8) \quad w_t = \begin{cases} 1 & \text{für } |u_t^*| \leq 2,5 \\ 0 & \text{für } |u_t^*| > 2,5 \end{cases} .$$

Die Verwendung des Gewichtungsschemas (2.11.8) entspricht einer "harten" Verwerfungsregel für Ausreißer. Eine "glatte" Verwerfungsregel kann z.B. durch

⁷⁴ Der Bruchpunkt (breakdown point), der den Anteil an Ausreißern misst, den ein Schätzer gerade noch verkraften kann, ohne beliebig verzerrt werden zu können, beträgt bei der M-Schätzung $1/n$ (vgl. Rousseeuw und Leroy, 1987, S. 13f.). Dieser niedrige Wert des Bruchpunkts resultiert aus den Problemen einer Kontrolle der Regressoren, deren Werte in der Ökonometrie im Unterschied z.B. zur Biometrie nicht vorgegeben sind, sondern beobachtet werden. Während dort ein robustes Verfahren nicht notwendig auch auf eine Begrenzung des Einflusses der Regressoren abstellen muss, ist es bei Beobachtungswerten zur Vermeidung unerwünschter Effekte wichtig, alle in die Regression eingehenden Variablen unter Kontrolle zu halten.

$$(2.11.9) \quad w_t = \begin{cases} 1 & \text{für } |u_t^*| \leq 2 \\ (0,1) & \text{für } 2 < |u_t^*| \leq 3 \\ 0 & \text{für } |u_t^*| > 3 \end{cases}$$

definiert werden. Beobachtungen, deren standardisierte Residuen im Intervall zwischen 2 und 3 liegen, werden dabei nicht unmittelbar als Ausreißer verworfen, sie erhalten jedoch ein geringeres Gewicht als die "guten" Daten.

Zunächst muss jedoch bei unbekanntem Gewichten eine Ausgangslösung gefunden werden, um die Residuen und einen Skalenschätzer berechnen zu können. Die OLS-Methode ist hierzu keine geeignete Wahl, da die Regressionsgerade in Richtung der Ausreißer zu liegen tendiert, so dass ihre Residuen in ihrer Größenordnung unterschätzt werden. In der RLS-Methode wird eine Ausgangslösung durch Minimierung des Medians der quadrierten Residuen ermittelt (Rousseeuw und Leroy, 1987, S. 44ff.):

$$(2.11.10) \quad M(\beta) = \underset{\beta}{\text{med}} \hat{u}_t^2 \xrightarrow{!} \text{Min}.$$

Geometrisch wird mit der **LMS-Methode (least-median of squares)** der schmalste Streifen bestimmt, in dem gerade die Hälfte der Beobachtungen liegt. Die LMS-Regressionsgerade liegt genau in der Mitte dieses Streifens. Die LMS-Methode erweist sich als äußerst robust und besitzt den höchstmöglichen Bruchpunkt (Rousseeuw, 1984). Sie stellt jedoch nicht die endgültige Lösung dar, da ihre asymptotische Effizienz unbefriedigend ist. Außerdem sind keine Statistiken verfügbar, die eine Bestimmung von Konfidenzintervallen und eine Durchführung von Signifikanztests erlauben.

Der vorläufige Skalenschätzer $\hat{\sigma}^{*(0)}$ basiert auf der Zielfunktion der LMS-Methode:

$$(2.11.11) \quad \hat{\sigma}^{*(0)} = C \cdot \sqrt{\underset{t}{\text{med}} \hat{u}_t^2}.$$

Der Korrekturfaktor C stellt eine unverzerrte Schätzung des Standardfehlers unter einer Normalverteilung sicher.⁷⁵ Nach Bestimmung der Gewichte w_t z.B. nach Regel (2.11.8) oder (2.11.9) unter Verwendung von $\hat{\sigma}^{*(0)}$ lässt sich ein neuer Skalenschätzer der Form

⁷⁵ Der Korrekturfaktor C ist durch $C = 1,4826 \cdot \left(1 + \frac{5}{n-p}\right)$ gegeben, wobei der reziproke Wert des ersten

Faktors dem Durchschnitt des MAD (median absolute deviation) bei einer Normalverteilung entspricht. Vgl. Rousseeuw und Leroy (1987), S. 44; und Li (1985), S. 302.

$$(2.11.12) \quad \hat{\sigma}^{*(1)} = \sqrt{\frac{\sum_{t=1}^n w_t \cdot \hat{u}_t^2}{\sum_{t=1}^n w_t - p}}$$

bestimmen. Die Beobachtungen werden dann mit der auf der Grundlage des Skalenschätzers $\hat{\sigma}^{*(1)}$ ermittelten Gewichtsfunktion w_t in die RLS-Schätzung einbezogen. Der endgültige Skalenschätzer σ^* ist ebenfalls von der Bauart (2.11.12), jedoch mit Gewichten und Residuen, die aus der RLS-Methode resultieren. Bei Verwendung einer "harten" Verwerfungsregel wie z.B. (2.11.8) lässt sich die RLS-Methode als Anwendung der gewöhnlichen Methode der kleinsten Quadrate auf einen von Ausreißern bereinigten Datensatz interpretieren.

Welche Verwerfungsregel für Ausreißer jedoch immer gewählt wird: Vorteilhaft ist, dass bei der RLS-Methode zur OLS-Methode vergleichbare Statistiken verfügbar sind. Bei der Berechnung der entsprechenden Statistiken ist jedoch die Gewichtsfunktion zu berücksichtigen. Die inferenzstatistischen Konzepte gelten allerdings nicht mehr als exakt, da die zugrunde liegende Verteilungstheorie ungleich komplizierter ist als unter den Voraussetzungen der OLS-Methode. In Monte-Carlo-Studien hat sich jedoch gezeigt, dass die Approximationen hinreichend gut sind, wodurch zuverlässige Aussagen ermöglicht werden.

Ein Beispiel soll die Ausreißereffekte illustrieren, die bei einer ökonometrischen Schätzung eines Modells der Kapitalmarkttheorie auftreten können.⁷⁶ Unseres Erachtens kommt einer robusten Schätzung bei Finanzmarktdaten eine besondere Relevanz zu, da sie häufig hochgradig volatil sind. Da einer Ausgangslösung mittels der LMS-Methode effizient allein mit einem Rechner ermittelt werden kann, verzichten wir auf eine Illustration der Rechentechnik. Vielmehr steht hier die Interpretation des Schätzergebnisses im Mittelpunkt des Interesses.

Beispiel 2.11. 1: In der Kapitalmarkttheorie wird bei Finanzanlagen zwischen einem systematischen und unsystematischen Risiko unterschieden. Das unsystematische Risiko kann als wertpapierspezifisches Risiko durch Diversifikation eliminiert werden. Bewertungsrelevant ist allein das systematische Risiko, das die Sensitivität eines Wertpapiers in Bezug auf die Marktlage widerspiegelt.

⁷⁶ Zur Anwendung robuster Schätzverfahren in der Kapitalmarkttheorie s. Kosfeld (1996), S. 78ff.

Das Marktmodell von Sharpe (1963) geht in diesem Zusammenhang von einer linearen Beziehung zwischen der Rendite r_{jt} des Wertpapiers j und der Rendite r_{Mt} des "Marktportefeuilles" aus:

$$(2.11.13) \quad r_{jt} = \alpha_j + \beta_j \cdot r_{Mt} + u_t .$$

Der Störterm u_t erfasst hierbei das wertpapierspezifische Risiko, das in dem Kapitalmarktmodell als Residualgröße mit einem Erwartungswert von Null und konstanter Varianz aufgefasst wird. Das systematische Risiko wird dagegen durch den β -Faktor erfasst, der für das Portfoliomanagement eine entscheidungsrelevante Größe darstellt.

Bei der ökonometrischen Schätzung des Marktmodells muss das "Marktportefeuille" geeignet operationalisiert werden. Hier soll das Marktmodell für die Karstadt-Aktie ökonometrisch geschätzt werden, wobei das Marktportefeuille durch den Deutschen Aktienindex (DAX) erfasst wird. Als Ursprungsdaten werden die börsentäglichen Kursnotierungen der Karstadt-Aktie und der Indexwerte des DAX an der Frankfurter Wertpapierbörse für den Zeitraum vom 17.06.1994 bis zum 18.07.1994 verwendet:

Tag	DAX-Index	Aktienkurs Karstadt
17.06.	2058,09	595,00
20.06.	1968,82	580,00
21.06.	1983,27	589,00
22.06.	1994,42	587,00
23.06.	2022,10	584,00
24.06.	2005,31	570,00
27.06.	1988,60	557,00
28.06.	2018,26	552,00
29.06.	2046,30	584,00
30.06.	2025,34	575,00
01.07.	2036,52	573,00
04.07.	2054,40	578,00
05.07.	2032,69	580,00
06.07.	2035,70	579,00
07.07.	2043,85	576,00
08.07.	2050,85	572,50
11.07.	2065,66	580,00
12.07.	2048,05	583,00
13.07.	2054,00	587,00
14.07.	2055,62	568,00
15.07.	2093,61	560,50
18.07.	2098,19	565,00

Quelle: Handelsblatt (diverse Ausgaben)

Aus den Kurznotierungen K_{jt} eines Wertpapiers oder Portefeuilles j lassen sich die Tagesrenditen r_{jt} bestimmen:⁷⁷

$$r_{jt} = \frac{K_{jt} - K_{j,t-1}}{K_{j,t-1}}$$

Auf diese Weise erhält man die Daten, auf die sich das Marktmodell (2.11.13) bezieht:

Tag	Tagesrendite r_{Mt} des DAX	Tagesrendite r_{Kt} der Karstadt-Aktie
20.06.	-0,0252	-0,0434
21.06.	0,0155	0,0073
22.06.	-0,0034	-0,0196
23.06.	-0,0051	0,0139
24.06.	-0,0240	-0,0083
27.06.	-0,0228	-0,0083
28.06.	0,0269	0,0149
29.06.	0,0210	0,0139
30.06.	-0,0154	-0,102
01.07.	-0,0035	0,0055
04.07.	0,0087	0,0088
05.07.	0,0035	-0,1060
06.07.	-0,0017	0,0015
07.07.	-0,0052	0,0040
08.07.	-0,0061	0,0034
11.07.	0,0131	0,0072
12.07.	0,0052	-0,0085
13.07.	0,0069	0,0029
14.07.	-0,0324	0,0008
15.07.	-0,0132	0,0185
18.07.	0,0080	0,0022

Wenn man das Sharpsche Marktmodell mit der OLS-Methode schätzt, erhält man folgendes Ergebnis (t-Werte in Klammern):

$$\hat{r}_{Kt} = 0,00085 + 0,44527 \cdot r_{Mt}$$

(0,310) (2,543)

$$R^2 = 0,254, \quad F = 6,469$$

Da die F-Statistik mit 6,469 den kritischen Wert $F_{0,95;1;19} = 4,38$ übersteigt, ergibt sich bei einem Signifikanzniveau von 5% ein statistisch gesicherter Zusammenhang zwischen den Tagesrenditen der Karstadt-Aktie und den Tagesrenditen des DAX. Wegen

⁷⁷ Eine Kursbereinigung im Falle von Dividendenzahlungen, Aktiensplits usw. ist hierbei nicht berücksichtigt.

$F = 6,469 < F_{0,99;1;19} = 8,18$ kann der Zusammenhang jedoch nicht als hochsignifikant bezeichnet werden. Die Renditeschwankungen der Karstadt-Aktie können zu 25,4 % durch die Marktrendite "erklärt" werden, die anhand der Indexwerte des DAX gemessen wird. Während sich der β -Faktor auf dem 5 %-Niveau als signifikant erweist, ist das absolute Glied nicht gegen Null gesichert.

Bei der OLS-Schätzung des Marktmodells werden jedoch Ausreißer berücksichtigt, die im börsentäglichen Handel aufgrund verschiedenartigster Informationen zustande kommen können. Sie können sich im Nachhinein als korrekturbedürftig erweisen, wenn sie sich als nicht bewertungsrelevant herausstellen. Hierzu zählen auch Saisonalitäten und Anomalien, die für den deutschen Aktienmarkt nachgewiesen werden konnten (s. z.B. Frantzmänn, 1987, S. 66ff.; Krämer und Runde, 1992). Da Ausreißer unkontrollierte Einflüsse auf die Wertpapier-Portefeuille-Beziehung haben können, soll eine robuste Schätzung durchgeführt werden, die die Struktur zwischen der Masse der Daten offen legt. Die Schätzergebnisse der Reweighted-Least-Square-Methode (RLS-Methode),

$$\hat{r}_{Kt} = 0,00407 + 0,34459 \cdot r_{Mt}$$

(3,105)
(4,290)

$$R^2 = 0,568, \quad F = 18,407$$

zeigen nicht unbeträchtliche Unterschiede zu den Ergebnissen der OLS-Schätzung. Bei fünf Ausreißern hat die RLS-Methode eine Repräsentation von 76,2 %⁷⁸. Offenbar tritt die Relevanz des Marktmodells bei der OLS-Methode aufgrund von Ausreißereffekten nicht klar genug hervor. Bei einer Kontrolle des Ausreißereinflusses steigt der Erklärungsgehalt des Marktmodells von 25,4 % auf 56,8 % an. Der Gesamtzusammenhang ist bei einem Signifikanzniveau von 1 % statistisch gesichert. In der robusten Schätzung ist das absolute Glied auf dem 5 %-Niveau signifikant, während der β -Faktor auf dem 1 %-Niveau signifikant ist. Die Sensitivität der Karstadt-Aktie gegenüber den allgemeinen Markteinflüssen, die durch den β -Faktor gemessen wird, muss dagegen bei einer Bereinigung der Ausreißereffekte nach unten korrigiert werden. ♦

⁷⁸ Ausreißer treten an folgenden Börsentagen auf: 20.6., 22.6., 5.7., 12.7., 15.7. Die ersten beiden Ausreißer kommen durch hohen Kurseinbruch bei der Karstadt-Aktie im Vergleich zum DAX zustande, während bei den letzten drei Ausreißern eine inverse Beziehung zwischen den Renditen der Karstadt-Aktie und des DAX besteht.

Aufgaben

- (2.11.1) Was versteht man unter einem robusten Verfahren?
- (2.11.2) Warum kann ein robustes Schätzverfahren bei der Existenz von Ausreißern gegenüber der OLS-Methode den Vorzug zu geben sein?
- (2.11.3) Erläutern Sie die Grundidee der verallgemeinerten Maximum-Likelihood-Schätzung (M-Schätzung)!
- (2.11.4) Nach welchem Prinzip arbeitet die Reweighted-Least-Squares-Methode (RLS-Methode)?

Anhang

Tabelle A2: Quantile der Chi-Quadrat-Verteilung bei v Freiheitsgraden

F(x)	0.005	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99	0.995
v										
1	0.000	0.0002	0.0010	0.004	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.010	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.6
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5

Tabelle A3: Quantile der t-Verteilung bei v Freiheitsgraden

F(x)	0.9000	0.9500	0.9750	0.9900	0.9950	0.9995
v						
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Tabelle A4: Quantile der F-Verteilung bei v_1 und v_2 Freiheitsgraden

F(x)	v_2	v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
0.950 0.990	1		161 40524999	200 54035625	216 57645859	225 59285981	230 60236056	234 61066157	237 62096261	239 63136340	241 6406366	242 6499	244 6592	246 6685	248 6778	250 6871	252 6964	253 7057	254 7150
0.950 0.990	2		18.5 98.5	19.0 99.0	19.2 99.2	19.2 99.3	19.3 99.3	19.3 99.4	19.4 99.4	19.4 99.4	19.4 99.4	19.4 99.4	19.4 99.4	19.5 99.5	19.5 99.5	19.5 99.5	19.5 99.5	19.5 99.5	19.5 99.5
0.950 0.990	3		10.1 34.1	9.55 30.8	9.28 29.5	9.12 28.7	9.01 28.2	8.94 27.9	8.89 27.7	8.85 27.5	8.81 27.3	8.79 27.2	8.74 27.1	8.70 26.9	8.66 26.7	8.62 26.5	8.57 26.3	8.55 26.2	8.53 26.1
0.950 0.990	4		7.71 21.2	6.94 18.0	6.59 16.7	6.39 16.0	6.26 15.5	6.16 15.2	6.09 15.0	6.04 14.8	6.00 14.7	5.96 14.5	5.91 14.4	5.86 14.2	5.80 14.0	5.75 13.8	5.69 13.7	5.66 13.6	5.63 13.5
0.950 0.990	5		6.61 16.3	5.79 13.3	5.41 12.1	5.19 11.4	5.05 11.0	4.95 10.7	4.88 10.5	4.82 10.3	4.77 10.2	4.74 10.1	4.68 9.89	4.62 9.72	4.56 9.55	4.50 9.38	4.43 9.20	4.40 9.11	4.37 9.02
0.950 0.990	6		5.99 13.7	5.14 10.9	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.00 7.72	3.94 7.56	3.87 7.40	3.81 7.23	3.74 7.06	3.70 6.97	3.67 6.88
0.950 0.990	7		5.59 12.2	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 6.99	3.73 6.84	3.68 6.72	3.64 6.62	3.57 6.47	3.51 6.31	3.44 6.16	3.38 5.99	3.30 5.82	3.27 5.74	3.23 5.65
0.950 0.990	8		5.32 11.3	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.18	3.44 6.03	3.39 5.91	3.35 5.81	3.28 5.67	3.22 5.52	3.15 5.36	3.08 5.20	3.01 5.03	2.97 4.95	2.93 4.86
0.950 0.990	9		5.12 10.6	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.61	3.23 5.47	3.18 5.35	3.14 5.26	3.07 5.11	3.01 4.96	2.94 4.81	2.86 4.65	2.79 4.48	2.75 4.40	2.71 4.31
0.950 0.990	10		4.96 10.0	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.20	3.07 5.06	3.02 4.94	2.98 4.85	2.91 4.71	2.84 4.56	2.77 4.41	2.70 4.25	2.62 4.08	2.58 4.00	2.54 3.91
0.950 0.990	12		4.75 9.33	3.89 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.91 4.64	2.85 4.50	2.80 4.39	2.75 4.30	2.69 4.16	2.62 4.01	2.54 3.86	2.47 3.70	2.38 3.54	2.34 3.45	2.30 3.36
0.950 0.990	15		4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.71 4.14	2.64 4.00	2.59 3.89	2.54 3.80	2.48 3.67	2.40 3.52	2.33 3.37	2.25 3.21	2.16 3.05	2.11 2.96	2.07 2.87
0.950 0.990	20		4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.51 3.70	2.45 3.56	2.39 3.46	2.35 3.37	2.28 3.23	2.20 3.09	2.12 2.94	2.04 2.78	1.95 2.61	1.90 2.52	1.84 2.42
0.950 0.990	30		4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.33 3.30	2.27 3.17	2.21 3.07	2.16 2.98	2.09 2.84	2.01 2.70	1.93 2.55	1.84 2.39	1.74 2.21	1.68 2.11	1.62 2.01
0.950 0.990	60		4.00 7.08	3.15 4.98	2.76 4.13	2.53 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.92 2.50	1.84 2.35	1.75 2.20	1.65 2.03	1.53 1.84	1.47 1.73	1.39 1.60
0.950 0.990	120		3.92 6.85	3.07 4.79	2.68 3.95	2.45 3.48	2.29 3.17	2.18 2.96	2.09 2.79	2.02 2.66	1.96 2.56	1.91 2.47	1.83 2.34	1.75 2.19	1.66 2.03	1.55 1.86	1.43 1.66	1.35 1.53	1.25 1.38
0.950 0.990	∞		3.84 6.63	3.00 4.61	2.60 3.78	2.37 3.32	2.21 3.02	2.10 2.80	2.01 2.64	1.94 2.51	1.88 2.41	1.83 2.32	1.75 2.18	1.67 2.04	1.57 1.88	1.46 1.70	1.32 1.47	1.22 1.32	1.00 1.00

Tabelle A5: Durbin-Watson-Statistik

a) Signifikanzniveau 0,05

n	k = 2		k = 3		k = 4		k = 5		k = 6	
	DL	dU								
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,73	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,79
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,78
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,77
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

b) Signifikanzniveau 0,01

n	k = 2		k = 3		k = 4		k = 5		k = 6	
	d_u	d_o								
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,90
17	0,87	1,10	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,80	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
21	0,97	1,16	0,89	1,27	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,91	1,28	0,83	1,40	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,40	0,77	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,80	1,53	0,72	1,66
25	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,10	1,24	1,04	1,32	0,97	1,41	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,05	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,20	1,48	1,16	1,53	1,11	1,58
50	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,60	1,39	1,63
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,42	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

n bezeichnet den Beobachtungsumfang, k die Anzahl der erklärenden Variablen (einschl. Scheinvariable).

Quelle: Durbin, Watson (1951)

Tabelle A6: Dickey-Fuller-Verteilung

ohne Konstante, ohne Trend

$\frac{n}{k}$	10	15	20	25	30	35	40	45	50	60	80	100	150	200	300	500	∞
1	-1,9791	-1,9658	-1,9592	-1,9552	-1,9526	-1,9507	-1,9493	-1,9481	-1,9473	-1,9459	-1,9443	-1,9433	-1,9420	-1,9413	-1,9406	-1,9401	-1,9393
2	-4,0242	-3,7754	-3,6585	-3,5907	-3,5466	-3,5155	-3,4925	-3,4747	-3,4606	-3,4396	-3,4137	-3,3983	-3,3779	-3,3678	-3,3577	-3,3497	-3,3378
3	-4,7122	-4,3593	-4,1940	-4,0984	-4,0362	-3,9925	-3,9601	-3,9351	-3,9153	-3,8858	-3,8494	-3,8278	-3,7992	-3,7850	-3,7709	-3,7597	-3,7430
4	-5,3902	-4,9122	-4,6912	-4,5643	-4,4821	-4,4246	-4,3821	-4,3494	-4,3235	-4,2851	-4,2377	-4,2096	-4,1726	-4,1543	-4,1361	-4,1216	-4,1000

mit Konstante, ohne Trend

$\frac{n}{k}$	10	15	20	25	30	35	40	45	50	60	80	100	150	200	300	500	∞
1	-3,2195	-3,0818	-3,0199	-2,9850	-2,9627	-2,9472	-2,9358	-2,9271	-2,9202	-2,9101	-2,8976	-2,8903	-2,8807	-2,8760	-2,8713	-2,8676	-2,8621
2	-4,0242	-3,7754	-3,6585	-3,5907	-3,5466	-3,5155	-3,4925	-3,4747	-3,4606	-3,4396	-3,4137	-3,3983	-3,3779	-3,3678	-3,3577	-3,3497	-3,3378
3	-4,7122	-4,3593	-4,1940	-4,0984	-4,0362	-3,9925	-3,9601	-3,9351	-3,9153	-3,8858	-3,8494	-3,8278	-3,7992	-3,7850	-3,7709	-3,7597	-3,7430
4	-5,3902	-4,9122	-4,6912	-4,5643	-4,4821	-4,4246	-4,3821	-4,3494	-4,3235	-4,2851	-4,2377	-4,2096	-4,1726	-4,1543	-4,1361	-4,1216	-4,1000

mit Konstante, mit Trend

$\frac{n}{k}$	10	15	20	25	30	35	40	45	50	60	80	100	150	200	300	500	∞
1	-3,9948	-3,7611	-3,6591	-3,6027	-3,5670	-3,5426	-3,5247	-3,5112	-3,5005	-3,4849	-3,4659	-3,4548	-3,4403	-3,4332	-3,4263	-3,4207	-3,4126
2	-4,8736	-4,4759	-4,2896	-4,1818	-4,1117	-4,0624	-4,0258	-3,9977	-3,9753	-3,9421	-3,9010	-3,8766	-3,8444	-3,8284	-3,8125	-3,7998	-3,7810
3	-5,4530	-4,9793	-4,7533	-4,6213	-4,5347	-4,4736	-4,4281	-4,3930	-4,3650	-4,3233	-4,2717	-4,2409	-4,2000	-4,1797	-4,1595	-4,1434	-4,1194
4	-6,0749	-5,4830	-5,2033	-5,0407	-4,9345	-4,8597	-4,8041	-4,7613	-4,7272	-4,6765	-4,6137	-4,5764	-4,5269	-4,5024	-4,4780	-4,4585	-4,4295

n bezeichnet den Beobachtungsumfang, k die Anzahl der in den Kointegrationstest einbezogenen Variablen (ohne Absolutglied).

Anhang B: Daten

Tabelle B1: Daten für Energienachfragemodelle

obs	GASV	VEINKR	GASPR	FERNWPR
1980	10.00	1285.70	0.92	0.90
1981	10.60	1286.00	1.04	1.04
1982	10.40	1254.40	1.15	1.08
1983	11.10	1246.20	1.11	1.11
1984	11.90	1275.30	1.08	1.10
1985	13.80	1298.70	1.11	1.11
1986	13.70	1357.10	1.05	1.14
1987	13.70	1408.40	0.84	1.07
1988	12.20	1450.90	0.80	1.02
1989	12.90	1485.10	0.80	1.00
1990	13.60	1590.00	0.82	1.01
1991	13.80	1671.10	0.85	1.02
1992	13.60	1699.10	0.83	1.00
1993	13.60	1691.40	0.80	0.97
1994	13.80	1681.70	0.78	0.95
1995	13.70	1700.50	0.71	0.94

Legende:

GASV Erdgasverbrauch (real)

VEINKR Verfügbares Einkommen (real)

GASPR Relativer Gaspreis (Bezugsbasis: Preisindex des Privaten Verbrauchs)

FERNWPR Relativer Fernwärmepreis (Bezugsbasis: Preisindex des Privaten Verbrauchs)

Quelle: Statistisches Bundesamt; eigene Berechnungen

Tabelle B2: Daten zur Verkehrsmittelwahl

Person	Verkehrsmittel	Entfernung zum Arbeitsplatz (km)	Berufliche Stellung
1	0	6	0
2	0	3	1
3	1	5	1
4	1	8	1
5	1	2	0
6	0	8	0
7	1	10	0
8	0	5	0
9	1	6	1
10	1	8	1
11	1	3	0
12	1	8	1
13	0	2	0
14	0	1	0
15	1	4	0
16	1	8	1
17	0	2	1
18	0	4	0
19	1	7	0
20	1	4	1
21	1	9	1
22	0	4	0
23	0	2	0
24	1	6	1
25	0	6	0
26	1	8	1
27	0	4	0
28	1	10	1
29	1	12	0
30	0	4	0

Literaturverzeichnis

Das Verzeichnis enthält neben der im Text zitierten Literatur eine Auswahl von deutschsprachigen Lehrbüchern zur Ökonometrie.

Aitken, A.C. (1935), On Least Squares and Linear Combinations of Observations, Proceedings of the Royal Statistical Society of Edinburgh, 55, S. 42-48.

Almon, S. (1965), Distributed Lag between Capital Appropriations and Expenditures, *Econometrica*, 33, S. 178-196.

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, Mass.

Anderson, T.W. und Rubin, H. (1949), Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *Annals of Mathematical Statistics*, 20, S. 46-63.

Assenmacher, W. (1990), *Einführung in die Ökonometrie*, 3. Aufl., München.

Assenmacher, W. und Braun, G.E. (1981), Das Einfachheitspostulat in Wissenschaftstheorie und Ökonometrie, *Statistische Hefte*, S. 152-175.

Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, Wiley & Sons, New York.

Banerjee, A. (1999), Panel Data Unit Root Tests and Cointegration: An Overview, *Oxford Bulletin of Economics and Statistics*, Special Issue, S. 607-629.

Banerjee, A., Dolado, J. J., Mestre, R. (1998), Error-Correction Mechanism Tests for Cointegration in a Single-Equation framework, *Journal of Time Series Analysis*, 19, S. 267-283.

Banerjee, A.; Dolado, J. J.; Hendry, D. F.; Smith, G. W. (1986), Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence, *Oxford Bulletin of Economics and Statistics*, Vol. 48, S. 253-278.

Banerjee, A.; Dolado, J. J.; Mestre, R. (1998): Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework, *Journal of Time Series Analysis*, 19, S. 267-283.

Barnett, V. und Lewis, T. (1978), *Outliers in Statistical Data*, London.

Basman, R.L. (1963), A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three Equations Case, *Journal of the American Statistical Association*, 58, S. 161-171.

Basman, R.L. (1961), A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Over Identified Cases, *Journal of the American Statistical Association*, 56, S. 619-636.

Belsley, D.A.; Kuh, E. und Welsch, R.E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York.

- Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics*, Vol. 31, S. 307-327.
- Brachinger, H.-W. (1981), Robuste Punktschätz-Verfahren, *Allgemeines Statistisches Archiv*, S. 325-349.
- Breusch, T. S. (1978), Testing for Autocorrelation in Dynamic Linear Models, *Australian Economic Papers*, Vol. 17, S. 334-355.
- Breusch, T. S., Pagan, A. R. (1980), The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics, *Review of Economic Studies*, Vol. 47, S. 239-253.
- Breusch, T.S. und Pagan, A.R. (1980), A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, 47, S. 1287-1294.
- Brown, R. L., Durbin, J. und Evans, J. M. (1975), Techniques for Testing the Constancy of Regression Relationships over Time, *Journal of the Royal Statistical Society, Series B*, Vol. 37, S. 149-163.
- Brown, T.M. (1952), Habit Persistence and Lags in Consumer Behaviour, *Econometrica*, 20, S. 355-371.
- Cagan, P. (1956), The Monetary Dynamics of Hyperinflations, in: Friedman, M. (ed.), *Studies in the Quantity Theory of Money*, S. 25-117.
- Chan, K.S. (1993), Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *Annals of Statistics*, 21, S. 520-533.
- Chow, G.C. (1960), Test of Equality between Sets of Coefficients in Two Linear Regressions, *Econometrica*, 28, S. 591-605.
- Cochrane, D. und Orcutt, G.H. (1949), Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms, *Journal of the American Statistical Association*, 44, S. 32-61.
- Cochrane, J. (1988), How Big is the Random Walk Component in GNP, *Journal of Political Economy*, 96, S. 893-920.
- Copas, J.B. (1966), Monte Carlo Results for Estimation in a Stable Markov Time Series, *Journal of Royal Statistical Society, A* 129, S. 110-116.
- Cragg, J.G. (1967), On the Relative Small-Sample Properties of Several Structural-Equation Estimators, *Econometrica*, 35, S. 89-110.
- Darnell, A.C. (1994), *A Dictionary of Econometrics*, Eldershot, Hants, England.
- Dickey, D.A. und Fuller, W.A. (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, Vol. 74, S. 427-431.

- Dickey, D.A. und Fuller, W.A. (1981), Likelihood Ratio Test Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, 55, S. 251-256.
- Durbin, J. (1970), Testing for Serial Correlation in Least Squares Regression when some of the Regressors are lagged dependent Variables, *Econometrica*, 38, S. 410-421.
- Durbin, J. und Watson, G.S. (1950), Testing for Serial Correlation in Least-Squares Regression I, *Biometrika*, 37, S. 409-428.
- Durbin, J. und Watson, G.S. (1951), Testing for Serial Correlation in Least-Squares Regression II, *Biometrika*, 38, S. 159-178.
- Eckey, H.-F., Kosfeld, R. und Dreger, C. (2002), *Statistik, Grundlagen – Methoden – Beispiele*, 3. Aufl., Gabler, Wiesbaden.
- Ehrlicher, W. (1972), Geldtheorie, in: Ehrlicher, W. et al. (Hrsg.), *Kompandium der Volkswirtschaftslehre*, S. 339-407.
- Ehrlicher, W.; Esenwein-Rothe, I.; Jürgensen, H. und Rose, K. (1972), *Kompandium der Volkswirtschaftslehre*, Göttingen.
- Enders (2001), Improved Critical Values for the Enders-Granger Unit Root test, *Applied Economics Letters*, 8, S. 257-261.
- Enders, W., Granger, C.W.J. (1998), Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates, *Journal of Business and Economic Statistics*, 16, S. 304-311.
- Engle, R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50, S. 987-1008.
- Engle, R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, Vol. 50, S. 987-1007.
- Engle, R.F., Lilien, D., Robins, R. (1987), Estimating Time Varying Risk Premia in the Term Structure. The ARCH-M Model, *Econometrica*, Vol. 55, S. 391-407.
- Engle, R.F.; Granger, C.W.J. (1987), Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, S. 251-276.
- Farrar, D.E. und Glauber, R.R. (1967), Multicollinearity in Regression Analysis (The Problem Revisited), *Review of Economics and Statistics*, 49, S. 92-107.
- Fisz, M. (1976), *Wahrscheinlichkeitsrechnung und mathematische Statistik*, Berlin.
- Fomby, T.B.; Hill, R.C. und Johnson, S.R. (1984), *Advanced Econometric Methods*, New York.
- Frantzmann, H.-J. (1987), *Saisonalitäten und Bewertung am deutschen Aktien- und Rentenmarkt*, Frankfurt/M.

- Friedman, M. (1957), *A Theory of the Consumption Function*, Princeton.
- Frisch, R. (1933), Editorial, *Econometrica*, 1, S. 1-4.
- Frisch, R. und Waugh, F.V. (1933), Partial Time Regressions as Compared with Individual Trends, *Econometrica*, 1, S. 387-401.
- Fuller, W.A. (1976), *Introduction to Statistical Time Series*, New York.
- Gaab, W. (1974), *Schätzung verteilter Lags*, Meisenheim/Glan.
- Godfrey, L. (1978), Testing against General Autoregressive and Moving Error Models when the Regressors include Lagged Dependent Variables, *Econometrica*, Vol. 46, S. 1293-1302.
- Goldberger, A.S. (1964), *Econometric Theory*, New York/London.
- Goldfeld, S. M. (1973), The Demand for Money Revisited, *Brookings Papers on Economic Activity*, S. 577-638.
- Goldfeld, S.M. und Quandt, R.E. (1965), Some Tests of Homoscedasticity, *Journal of the American Statistical Association*, 60, S. 539-547.
- Granger, C.W.J. und Newbold, P. (1974), Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, S. 111-120.
- Gregory, A.E.; Hansen, B.E. (1996a): Residual-based Tests for Cointegration in Models with Regime Shift, *Journal of Econometrics*, 70, S. 99-126.
- Gregory, A.E.; Hansen, B.E. (1996b): Tests for Cointegration in Models with Regime and Trend Shifts, *Oxford Bulletin of Economics and Statistics*, Vol. 58, S. 555-560.
- Grenzdörfer, K. (1969), Vergleich einiger in der Ökonometrie verwendeter Schätzverfahren mittels Simulation von Drei-Gleichungssystemen, Würzburg.
- Grohmann, H. (1985), Vom theoretischen Konstrukt zum statistischen Begriff - Das Adäquationsproblem, *Allgemeines Statistisches Archiv*, S. 1-15.
- Haavelmo, T. (1943), The Statistical Implications of a System of Simultaneous Equation, *Econometrica*, 11, S. 1-12.
- Hampel, F.R. (1974), The Influence Curve and its Role in Robust Estimation, *Journal of the American Statistical Association*, 69, S. 383-393.
- Hampel, F.R.; Rochetti, E.M.; Rousseeuw, P.J. und Stahel, W.A. (1986), *Robust Statistics*, New York.
- Hardes, H.-D. und Mertes, J. (1991), *Grundzüge der Volkswirtschaftslehre*, 3. Aufl., München.

- Hartwig, H. (1956), *Naturwissenschaftliche und sozialwissenschaftliche Statistik*, Zeitschrift für die gesamte Staatswissenschaft, 112, S. 252-266.
- Harvey, A.; Ruiz, E. und Shepard, N. (1994), *Multivariate Stochastic Variance Models*, Review of Economic Studies, 61, S. 247-264.
- Harvey, A.C. (1975), *Contribution to the Discussion of the Paper by Brown, Durbin and Evans*, Journal of the Royal Statistical Society, Series B, Vol. 37, S. 179-180.
- Harvey, A.C. (1981), *Time Series Models*, Oxford.
- Harvey, A.C. (1990), *The Econometric Analysis of Time Series*, 2nd ed., New York.
- Harvey, A.C. und Collier, P. (1977), *Testing for Functional Misspecification in Regression Analysis*, Journal of Econometrics, Vol. 6, S. 103-119.
- Hassler, U. (2001): *(Co-)Integration Testing under Structural Breaks – A Survey with Special Emphasis on the German Unification*, in: Galler, H.-P., Pohl, R. (Hrsg.): *Implikationen der Währungsunion für makroökonomische Modelle*, Nomos, Baden-Baden. S. 113-126.
- Hausman, J. A. (1978), *Specification Tests in Econometrics*, Econometrica, Vol. 46, S. 1251-1271.
- Heiler, S. (1980), *Robuste Schätzung im Linearen Modell*, in: Nowak, H. und Zentgraf, R. (Hrsg.), *Robuste Verfahren*, S. 34-55, Berlin.
- Hildreth, C. und Lu, J.Y. (1960), *Demand Relations with Autocorrelated Disturbances*, Michigan State University, Agricultural Experiment Station, Technical Bulletin 276.
- Holland, P.,W. und Welsch, R.E. (1977), *Robust Regression Using Iteratively Reweighted Least-Squares*, Communications in Statistics, A 6, S. 813-828.
- Hood, W.C. und Koopmans, T.C. (eds.) (1953), *Studies in Econometric Method*, New York/London.
- Huber, P. (1964), *Robust Estimation of a Location Parameter*, Annals of Statistics, 35, S. 73-101.
- Hujer, R. und Cremer, R. (1978), *Methoden der empirischen Wirtschaftsforschung*, München.
- Hurwicz, L. (1950), *Least Squares Bias in Time Series*, in: Koopmans, T.C. (ed.), *Statistical Inference in Dynamic Economic Models*, New York/London.
- Im, K.S., Pesaran, M.H. und Shin, Y. (2003), *Testing for Unit Roots in Heterogeneous Panels*, Journal of Econometrics, 115, S. 53-74.
- Jarque, C.M. und Bera, A.K. (1980), *Efficient Tests for Normality, Homoscedasticity and Serial Independence in Regression Residuals*, Economic Letters, 6, S. 255-259.

- Johansen, S. (1988), *Statistical Analysis of Cointegration Vectors*, *Journal of Economic Dynamics and Control*, Vol. 12, S. 231-254.
- Johansen, S. (1995), *Likelihood based Inference on Cointegration in the Vector Autoregressive Model*.
- Jorgenson, D.W. (1966), *Rational Distributed Lag Function*, *Econometrica*, 34, S. 135-149.
- Judge, G.J., Griffith, W.E., Hill, R.C., Lütkepohl, H. und Lee, T.-C. (1985), *The Theory and Practice of Econometrics*, New York.
- Kabe, D.G. (1963), *A Note on the Exact Distribution of the GCL-Estimator in Two Leading Over-Identified Cases*, *Journal of the American Statistical Association*, 58, S. 535-537.
- Kabe, D.G. (1964), *On the Exact Distributions of the GCL-Estimator in a Leading Three Equations Case*, *Journal of the American Statistical Association*, 59, S. 881-894.
- Klein, L.R. (1962), *An Introduction to Econometric*, Engle-Wood Cliffs, New Jersey.
- Koenker, R.W. (1982), *Robust Methods in Econometrics*, *Econometric Reviews*, 1, S. 213-290.
- Koopmans, T.C. (ed.) (1950), *Statistical Inference in Dynamic Economic Models*, New York/London.
- Kosfeld, R. (1986), *Theoretische und numerische Aspekte in der Maximum-Likelihood-Faktorenanalyse*, Frankfurt/M.
- Kosfeld, R. (1996), *Kapitalmarktmodelle und Aktienbewertung – Eine statistisch-ökonomische Analyse*, Wiesbaden.
- Kosfeld, R. (1996), *Konsumklima und Verbrauchskonjunktur*, *Jahrbuch der Absatz- und Verbrauchsforschung*, 43, S. 298–317.
- Kosfeld, R. und Robé, S. (2001), *Testing for Nonlinearities in German Bank Stock Returns*, *Empirical Economics* 26, S. 581–597.
- Koyck, L.M. (1954), *Distributed Lags and Investment Analysis*, Amsterdam.
- Krämer, W. (1980), *Eine Rehabilitation der Gewöhnlichen Kleinst-Quadrate-Methode als Schätzverfahren in der Ökonometrie*, Frankfurt (M.).
- Krämer, W. (1991), *Modellspezifikationstests in der Ökonometrie*, *RWI-Mitteilungen*, S. 285-302.
- Krämer, W. und Runde, R. (1992), *Kalendereffekte auf Kapitalmärkten*, in: Bühler, W.; Hax, H. und Schmidt, R. (Hrsg.), *Empirische Kapitalmarktforschung*, Düsseldorf, S. 87-98.

- Kremers, J. J. M., Ericsson, N. R., Dolado, J. J. (1992), The Power of Cointegration Tests, *Oxford Bulletin of Economics and Statistics*, 54, S. 325-348.
- Krug und Nourney (1987), *Wirtschafts- und Sozialstatistik: Gewinnung von Daten*, 2. Aufl., München.
- Kwiatkowski, D., Phillips, P.C.B. Schmidt, P. and Y. Shin, Y. (1992), Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root? *Journal of Econometrics*, 44, S. 159-78.
- Levin, A.; Lin, Chien-Fu; Chu, C.J. (2002), Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, *Journal of Econometrics*, 108, S. 1-24.
- Li, G. (1985), Robust Regression, in: Hoaglin, D.C., Mosteller, F. und Tukey, J.W. (eds.), *Exploring Data Tables, Trends, and Shapes*, S. 281-340, New York.
- Ljung, G. M. und Box, G. E. P. (1978), On a Measure of Lack of Fit in Time Series Models, *Biometrika*, Vol. 65, S. 297-303.
- MacKinnon, J.G. (1991),. Critical Values for Cointegrating Test in Long-Run Economic Relations, in: R. F. Engle and C. W. J. Granger (eds.), *Readings in Cointegration Oxford*: Oxford University Press.
- Maddala, G. S. und Mount, T. D. (1973), A Comparative Study of alternative Estimators for Variance Components Models used in econometric Applications, *Journal of the American Statistical Association*, Vol. 68, S. 324-328.
- Mariano, R.S. (1972), The Existence of Moments of the Ordinary Squares and Two-Stage-Least-Squares Estimators, *Econometrica*, 40, S. 643-652.
- Mariano, R.S. (1982), Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous-Equations Case, *International Economic Review*, 23, S. 503-534.
- Mariano, R.S. und Sawa, T. (1972), The Exact Finite Sample Distribution of the Limited Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables, *Journal of the American Statistical Association*, S. 159-163.
- Menges, G. (1959), Zur stochastischen Grundlegung der Ökonometrie, *Zeitschrift für die gesamte Staatswissenschaft*, 115, S. 611-625.
- Menges, G. (1961), *Ökonometrie*, Wiesbaden.
- Modigliani, F. (1986), Life Cycle, Individual Thrift, and the Wealth of Nations, *American Economic Review*, 76, S. 297-313.
- Mood, A.M.; Graybill, F.A. und Boes, D.C. (1974), *Introduction to the Theory of Statistics*, Tokyo.
- Nagar, A.L (1960), A Monte Carlo Study of Alternative Simulatneous-Equation Estimators, *Econometrica*, 28, S. 573-590.

- Nerlove, M. (1958), Distributed Lags and Demand Analysis for Agricultural and Other Commodities, Agriculture Handbook 141, U.S. Department of Agriculture.
- Newey, W.K. and West, K.D. (1987), A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, S. 703-708.
- Nowak, H. und Zentgraf, R. (Hrsg.) (1980), Robuste Verfahren, Berlin.
- O'Connell, Paul G.J. (1998), The Overevaluation of Purchasing Power Parity, *Journal of International Economics*, 44, S. 1-19.
- Owen, A.D. (1976), A Proof that both the Bias and the Mean Square Error of the Two-Stage-Least Squares Estimator are Monotonically Non-Decreasing Functions of Sample Size, *Econometrica*, 44, S. 409-411.
- Perron, P. (1989): The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 57, S. 1361-1401.
- Perron, P. (1997), "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", *Journal of Econometrics*, 80, S. 355-385.
- Perron, P.; Vogelsang, T.J. (1993): The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis: Erratum, *Econometrica*, 61, S. 248-249.
- Phillips, P.C.B. (1980), The Exact Distribution of Instrumental Variable Estimators in an Equation Containing $n+1$ Endogenous Variables, *Econometrica*, 48, S. 861-878.
- Pindyck, R.S. und Rubinfeld, D.L. (1981), *Econometric Models and Economic Forecasts*, 2nd Ed., New York.
- Ramsey, J. B. (1969), Tests for Specification Errors in Classical Linear Least Squares Analysis, *Journal of the Royal Statistical Society, Series B*, Vol. 31, S. 350-371.
- Richardson, D.H. (1968), The Exact Distribution of a Structural Coefficient Estimator, 63, S. 1214-1226.
- Rinne, H. (1976), *Ökonometrie*, Stuttgart.
- Ronning, G. (1991), *Mikroökometrie*, München.
- Rothenberg, T.J. und Leenders, C.T. (1964), Efficient Estimation of Simultaneous Equation Systems, *Econometrica*, 32, S. 57-76.
- Rousseeuw, P. (1984), Least Median of Squares Regression, *Journal of the American Statistical Association*, 79, S. 871-880.
- Rousseeuw, P. und Leroy, A. (1987), *Robust Regression and Outlier Detection*, New York.
- Said, S. E., Dickey, D. A. (1984), Testing for Unit Roots in Autoregressive Moving-Average Models with unknown Order, *Biometrika*, Vol. 71, S. 599-607.

- Sawa, T. (1969), The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage-Least-Squares Estimators, *Journal of the American Statistical Association*, 64, S. 923-937.
- Sawa, T. (1972), Finite-Sample Properties of the k-Class Estimators, *Econometrica*, 40, S. 653-680.
- Schneeweiß, H. (1990), *Ökonometrie*, 4. Aufl., Würzburg.
- Schönfeld, P. (1969), *Methoden der Ökonometrie*, Bd. I: Lineare Regressionsmodelle, München.
- Schönfeld, P. (1971), *Methoden der Ökonometrie*, Bd. II: Stochastische Regressoren und simultane Gleichungen, München.
- Schwert, G.W. (1989), Tests for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics*, Vol. 7, S. 147-159.
- Seibt, H. (1968), *Schätzung und Prognose bei simultanen Gleichungsmodellen: Eine Monte Carlo-Untersuchung*, Göttingen.
- Sharpe, W.F. (1963), A Simplified Model for Portfolio Analysis, *Management Science*, 9, S. 277-293.
- Siebke, J. und Willms, M. (1972), Zinsniveau, Geldpolitik und Inflation, *Kredit und Kapital*, 5, S. 171-205.
- Sims, C.A. (1972), Money, Income and Causality, *American Economic Review*, 62, S. 545ff.
- Smith, V.K. (1970), A Monte Carlo Experiment with a Large Macro-Econometric Model, *Western Economic Journal*, S. 377-384.
- Solow, R.M. (1960), On a Family of Lag Distributions, *Econometrica*, 28, S. 393-406.
- Stock, J. H. (1987), Asymptotic Properties of LeastSquares Estimators of Co-Integrating Vectors, *Econometrica*, 55, S. 1035-1056.
- Stöwe, H. (1977), *Ökonometrie*, Meisenheim/Glan.
- Summers, R. (1965), A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators, *Econometrica*, 33, S. 1-41.
- Takeuchi, K. (1970), On the Exact Moment of the OLS, TSLS and the Instrumental Variables Method Estimator, *International Economic Review*, S. 1-12.
- Theil, H. (1961), *Economic Forecasts and Policy*, 2nd ed., Amsterdam.
- Theil, H. (1966), *Applied Economic Forecasting*, Amsterdam.

- Tinbergen, J. (1939), Econometric business cycle research, *Review of Economic Studies*, S. 73-90.
- Tukey, J.W. (1960), A Survey of Sampling from Contaminated Distributions, in: Olkin (ed.), *Contributions to Probability and Statistics*, S. 448-485, Stanford.
- Ullah, A. und Nagar, A.L. (1974), The Exact Mean of the Two-Stage-Least-Squares Estimator of the Structural Parameters in a Equation Having Three Endogenous Variables, *Econometrica*, S. 749-758.
- Wagner, H.W. (1958), A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations, *Econometrica*, 26, S. 117-133.
- Wahrendorf, J. (1980), Robuste Statistik: Eine einführende Übersicht, in: Nowak, H. und Zentgraf, R. (Hrsg.), *Robuste Verfahren*, S. 1-13, Berlin.
- Wald, A. (1936), Berechnung und Ausschaltung von Saisonschwankungen, in: *Beiträge zur Konjunkturforschung*, Österreichisches Institut für Konjunkturforschung, Wien.
- Westerhoff, H.-D. (1976), Ein Beitrag zur Überprüfung investitionstheoretischer Hypothesen, *RWI-Papiere*, Nr. 4, Essen.
- White, H. (1980), A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica*, Vol. 48, S. 817-838.
- Wold, H. (1954), Causality and Econometrics, *Econometrica*, 22, S. 162-177.
- Zellner, A. und Theil, H. (1962), Three Stages Least Squares: Simultaneous Estimation of Simultaneous Equations, *Econometrica*, 30, S. 54-78.

Stichwortverzeichnis

Abzählkriterium	331,334
adaptive Erwartung	142, 154, 156
Modell der adaptiven ~	154
ADF-Test	225ff., 245f., 251, 254ff., 262, 303
Aggregationsproblem	8, 17, 94
Akzelerator	356
Almon-Verfahren	148, 170ff., 175
Angebotsfunktion	14f., 323, 325ff., 332f.
ANOVA-Tabelle	78f., 82
ARCH-Modelle	264f., 267, 270
ARCH-M-Modell	269f.
ARCH-Prozess	266f.
Ätialprinzip	4f., 17
Ausreißer	10, 271ff., 283f.
Autokorrelation	21f., 98f., 113ff., 126, 131, 133, 135, 138ff., 151, 156, 158f., 162ff., 217, 232, 236f., 238, 250f., 271, 312, 359, 365, 368, 388
~ erster Ordnung	114ff., 121, 132, 135, 138, 163f.
Autokorrelationskoeffizient	113ff., 121, 135, 139, 162f., 294
Autoregressiver Prozeß	113, 120, 132, 228, 237, 264, 267, 372
Bestimmtheitsmaß	49ff.
korrigiertes ~	173
BLUE-Eigenschaft	40f., 46, 96, 298
Breusch-Godfrey-Test	119ff., 248, 251
Breusch-Pagan-Test	107ff., 127, 297
Chow-Test	184ff., 202, 217, 223
Cochran-Orcutt-Verfahren	138, 140
CUSUMSQ-Test	217, 223
CUSUM-Test	220f., 223, 226
Determinationskoeffizient	49ff., 78ff., 130, 140, 196
Dickey-Fuller-Test	234f., 239, 247, 250, 259f., 302
differenz-stationäre Prozesse	232f.
Distributed lag-Modelle	143ff., 170f., 175, 228
Dummy-Variablen	84, 176ff., 188, 192, 197, 202

Durbin-h-Test	162f., 166, 175
Durbin-Watson	113ff., 136ff., 162ff.
~-Statistik	115ff., 136ff., 162ff.
~Test	113ff., 140f., 152 165
E ffekte	
branchenspezifische ~	299f.
feste ~	287, 291, 295ff.
gruppenindividuelle ~	287ff.
individuelle ~	287ff., 290f., 293f., 295f., 298
zeitindividuelle	288
zufällige	286f., 293, 295, 297ff.
Effizienz	35, 41, 44, 45, 170, 272, 274
asymptotische ~	61, 279, 382
Eingleichungsmodelle	2, 10, 19, 203, 307, 309, 311, 338, 389
ökonometrische ~	11, 203, 273, 305f., 364
Einheitswurzeltest	234, 239ff., 245ff., 251, 253, 259ff., 263, 301, 304
Einkommenselastizität	38, 71
Zinselastizität der ~	2, 39, 76
Engle-Granger-Methode	246
~ zum Test auf Kointegration	246
~-Verfahren	255, 259, 263
Erwartungstreue	40, 42, 61f., 151f., 159
asymptotische ~	61
F eedback-Systeme	378
Fehlerkorrekturmodell	243, 247ff., 256
Finale Form	316
F-Test	79ff., 85, 91f., 234, 291, 299, 379
G ARCH	264, 268ff., 389
~Modell	268, 270
~Prozess	268f.
Geldnachfrage	2, 35, 38, 70f., 164, 185, 215
~funktion	2, 35f., 38, 70f., 105f., 110ff., 129, 131, 164ff., 186, 226
Gleichgewicht	153, 241, 243, 388
langfristiges ~	242, 249, 262
Gleichung	
Normal~	25, 39, 352
Struktur~	308, 326, 335ff., 345ff., 353, 357ff., 363

Goldfeld-Quandt-Test	102f., 105ff., 110, 127, 129, 141
Granger-Kausalität	378ff.
Granger-Repräsentationstheorem	243
Gütekriterium	40, 210, 380, 390
H arvey-Collier-Test	217, 223, 225
Hauptkomponentenanalyse	89
Hausmann	298
~Test	300
Heteroskedastizität	98ff., 103ff., 108, 110, 126f., 129, 131, 140f., 189, 193, 217, 264f., 304, 388
Test auf ~	100, 102, 111
Homoskedastizität	20, 22, 100, 103, 104f6f., 151f., 190, 312
Huber-Schätzer	277
i dempotent	63, 295, 348
Identifikationsproblem	155, 313, 321f., 324, 336, 373
Identifizierbarkeit	321ff., 326f., 329ff.
Inferenzstatistik	57, 130
Instrumentenschätzer	353f.
Instrumentvariablen	167ff., 313, 331
Methode der ~	162, 167, 352
Integration	252
Test auf ~	234
Integrationsgrad	234f., 235, 241ff., 257f., 263, 303
integrierte Prozesse	234, 263
IPS-Test	303
J arque-Bera-Test	217, 225, 251
Johansen-Ansatz	375
K ausalität	5, 364, 378f.
Granger-~	378ff.
Kausalprinzip	5, 17
Kleinst-Quadrate	24, 39, 55, 101, 122f., 129, 140, 228, 247, 252, 258f., 261, 263, 271, 275, 347, 349, 351, 353f.

Kointegration	227, 234, 241ff., 252, 259ff., 263
kointegrierender Vektor	241, 247
Konsistenz	41, 44, 47, 61, 126, 158, 160ff., 169, 348ff., 364, 376, 382f.
Konsum	
~funktion	4ff., 8, 10, 12, 16, 19, 30ff., 52, 64, 66, 69, 75, 80f., 117, 119, 121, 138f., 175, 202, 208, 222, 223, 225f., 305, 308, 355, 357, 388
~-neigung	3, 8f., 30, 32f., 66, 70, 179f., 222f., 249, 285
Korrelationsanalyse	89f., 97
einfache ~	89
Korrelationskoeffizient	49, 52f., 55, 83, 89f., 344
Kovarianzmatrix	
~bei Autokorrelation	131
~bei Heteroskedastizität	126
~des OLS-Schätzers	44, 86
Koyck-Modell	148, 141, 153, 155f.
KPSS-Test	237ff., 246
Lag	
~-Gewichte	145f., 148
~-Modell	143, 147f., 150ff., 170f., 175, 228
~-Operator	146f., 150, 232, 233, 236, 268, 317f.
mittlerer ~	149
~-Verteilung	148f., 154f., 170
Likelihood-Funktion	59f., 273f., 367ff.
Likelihood-Verhältnis-Test	195
Lineares Wahrscheinlichkeitsmodell	188
Ljung-Box-Test	119, 121, 248
LLC-Test	301ff.
LMS-Methode	279f.
Logit-Modelle	188, 190ff., 195ff., 199, 201f.
log-odds ratio	192
LSDV-Schätzer	290, 292, 298
Marktmodell	13ff., 281ff., 323f., 326ff., 332f.
~ von Sharpe	281
Maximum-Likelihood	57ff., 74, 82, 107f., 193, 198, 202, 273ff., 277, 284, 332, 335, 364, 377
Maximum-Likelihood- ~ Methode	57, 74, 193

~ bei beschränkter Information	3, 16, 364
~ bei voller Information	3, 16, 335, 364
~ Schätzer	59ff., 82, 108, 198, 275, 364, 377
Median der absoluten Abweichungen	276
Methode	
~ der Instrumentvariablen	162, 167, 352
~ der kleinsten Quadrate	11, 24f., 57, 74, 83, 122, 130, 141, 163, 167, 170, 179, 273, 280, 331, 335, 345, 354, 357, 361f., 364
dreistufige ~	332, 357
verallgemeinerte ~	122, 162, 193, 295, 341, 357
zweistufige ~	332, 345, 357, 370
Modell	
~-defekte	98f., 141, 272, 388
dynamisches ~	142
interdependentes ~	12f., 16, 305, 311, 335
lineares ~	19, 316
ökonometrisches ~	7, 54, 272, 310, 364, 388
rekursives ~	13, 310, 311
statisches ~	142
Moving-Average	121f., 133, 229f., 233, 236, 238
M-Schätzung	273f., 277, 284
Multikollinearität	26, 83f., 86ff., 99, 147, 151, 173, 245, 249, 335
Aufdeckung von ~	89, 97
Auswirkungen der ~	86
Begriff der ~	83
Überwindung von ~	94, 97
multipler Korrelationskoeffizient	49
Multiplikator	3, 145
Nachfragefunktion	14, 55, 94, 141, 177ff., 323ff., 332f.
Newton-Raphson-Verfahren	194, 370
Nichtstationarität	229f., 232, 234f., 246, 252f., 257, 302f.
Ökonometrie	1ff., 10ff., 17, 56, 115, 170, 176, 178, 227f., 265, 313, 331, 383, 390
Panel	285, 287, 304
~einheiten	290, 283f., 301ff.
~modell	285ff., 291, 293, 296, 298f., 304
~regression	299
~schätzer	287
partielle Anpassung	142, 153f., 156, 164, 175

Pfeilschema	12ff.
Pooled-Regression-Modell	291
Probit-Modell	190, 192
Produktionsfunktion	2, 11, 17, 72, 73, 306
Cobb-Douglas ~	11, 17, 72, 73, 306
Prognose	3, 203ff., 208, 210, 212, 213, 215, 218, 220, 224, 264, 266
~-fehlermaße	210f., 215
Güte der ~	210
Intervall-~	202f., 209, 215
Punkt-~	203f., 209
Q-Statistik	121, 236, 251
Qualitative Regressoren	176f.
Qualitative Wahlhandlungsmodelle	187
R	
Random Walk-Prozeß	230, 235, 238
Rang	25f., 62, 83, 90, 181, 330f., 334, 350f.
Reduzierte Form	313
Regressand	11
Regressionskoeffizienten	20, 66, 70, 74, 77, 83, 90, 94, 96, 163f., 179, 355
Regressionsmodell	11, 19, 26, 42, 55ff., 62, 66, 83f., 89, 92, 95f., 98f., 107f., 122, 130, 143, 151f., 156, 158, 163, 167f., 175, 216f., 219f., 228, 273, 286f.
einfaches ~	31, 101
gepooltes ~	286, 304
multiples ~	55f., 57f., 61, 99f., 104, 107, 114, 124, 141, 157, 176, 189, 196, 235
RESET-Test	223ff., 251
Robuste Regression	271
Robustheit	75, 271, 379, 389
S	
Simulation	386
Stationarität	227f., 232, 234, 237, 240, 245f., 249, 256, 258, 262, 265, 268, 373f.
Kovarianz-~	227, 374
Mittelwert-~	227, 229f.
Nicht-~	229f., 232, 234f., 245f.
Varianz-~	227, 231, 240
Streuungszerlegung	50, 56, 82
Strukturbruch	94, 105, 183ff., 202, 217f., 220f., 252ff., 383
~-test	176, 183

Superkonsistenz	248
SURE	311, 339, 343f., 358f., 376, 381
~-Methode	311, 339, 344, 358, 376, 381
~-Schätzer	343
Threshold Autoregression (TAR)	259ff.
Toleranzkoeffizient	92f.
t-Test	70, 102, 164, 173, 179, 248
überidentifiziert	332, 335, 345, 350, 353f., 364
unteridentifiziert	332, 335
Variable	
Dummy~	84, 176ff., 188, 192, 197, 202
latente ~	12
prä determinierte ~	16, 156, 306f., 309ff., 313, 329, 335ff., 357
qualitative ~	176ff., 180, 182, 187, 192f., 197, 202
Schein~	20f., 26, 78, 81, 83, 89, 96, 118, 157, 182, 188, 198, 204, 307, 323
Variance-Components-Modell	296ff.
Variance-Ratio	240
~-Analyse	239
Varianzanalyse	77, 178
Varianzinflationsfaktor (VIF)	92f.
Vektorautoregression	336, 390
Verfahren	
~ der Differenzenbildung	96ff., 234f., 235
~ der Hilfsregressionen	91f., 97
~ der Variablenunterdrückung	95, 97
verteilte Verzögerungen	142f., 147f., 175
Wahlhandlungsmodelle	187ff., 191
Wendepunktfehler	213f.
Wendepunkt-Fehlerquote	214
White-Noise-Prozeß	120, 131, 228f., 231f, 236, 373
White-Test	111f., 141
Zentraler Grenzwertsatz	6, 58, 271