

Gerhard Arminger · Franz Müller

Lineare Modelle zur Analyse von Paneldaten

ZUMA-Publikationen

In der Reihe ZUMA-Publikationen erscheinen fundierte Monographien und Sammelbände zu wichtigen Fragen der Empirischen Sozialforschung. Alle diese Werke sind in engem Zusammenhang mit dem Forschungsprogramm des *Zentrums für Umfragen, Methoden und Analysen (ZUMA)* entstanden. Veröffentlicht werden sowohl eigene Untersuchungen als auch die Ergebnisse der Arbeit von Gastwissenschaftlern, Workshops und wissenschaftlichen Tagungen. Es entspricht der Aufgabenstellung des ZUMA, daß der Schwerpunkt der Bände im Bereich der Methoden der Empirischen Sozialforschung liegt. Dennoch werden auch andere Themen behandelt, etwa Fragen des gesellschaftlichen Wandels und der Sozialberichterstattung.

Wir hoffen, daß die ZUMA-Publikationen einen Beitrag zur Weiterentwicklung, aber auch zur Ergebniskumulation der Empirischen Sozialforschung leisten.

Max Kaase, Peter Ph. Mohler

Gerhard Arminger · Franz Müller

Lineare Modelle zur Analyse von Paneldaten

Westdeutscher Verlag

Der Westdeutsche Verlag ist ein Unternehmen der Verlagsgruppe Bertelsmann International.

Alle Rechte vorbehalten

© 1990 Westdeutscher Verlag GmbH, Opladen



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.

Umschlaggestaltung: Horst Dieter Bürkle, Darmstadt

ISBN-13: 978-3-531-12176-5 e-ISBN-13: 978-3-322-88758-0

DOI: 10.1007/978-3-322-88758-0

Vorwort

Dieses Buch ist eine Einführung in lineare Modelle zur Analyse von Paneldaten. In Paneluntersuchungen werden an einer (Zufalls-) Stichprobe von Elementen die gleichen Variablen zu mehreren Zeitpunkten erhoben. Im Unterschied zu Zeitreihen, die in der Regel an einer Makroeinheit beobachtet werden, streuen Paneldaten sowohl über die Stichprobenelemente als auch über die Zeit. Damit sind Paneldaten einerseits zur Beobachtung der Effekte dynamischer Prozesse sowie der Effekte von Interventionen geeignet, andererseits lassen sich im Rahmen linearer Modelle die Einflüsse nicht beobachteter Variablen zumindest partiell eliminieren. Diese Möglichkeiten erklären die zunehmende Verwendung von Paneldaten sowohl in der empirischen Sozial- und Wirtschaftsforschung als auch in der Epidemiologie und der medizinischen Statistik.

In diesem Band werden statische und dynamische lineare Modelle vorgestellt, die in erster Linie für direkt beobachtbare metrische abhängige Variable entwickelt wurden. Um die Verwendbarkeit dieser Modelle zu erhöhen, werden diese linearen Modelle mit faktorenanalytischen Meßmodellen angereichert, die die Einbettung von latenten Variablen in die herkömmlichen Modelle erlauben. Den Komplikationen, die durch die zeitliche Abhängigkeit der Variablen entstehen, wird durch die Einführung von Varianzkomponenten- und Autokorrelationsmodellen sowohl in den Meßmodellen als auch in den Strukturgleichungsmodellen für latente Variable Rechnung getragen. Im letzten Kapitel, das von Ulrich Küsters und Andreas Schepers verfaßt wurde, wird auf die Möglichkeiten und die Probleme der Übertragung linearer Modelle für metrische abhängige Variable auf zensierte, dichotome oder ordinale abhängige Variable, wie sie gerade in den Sozial- und Wirtschaftswissenschaften häufig auftreten, eingegangen.

Die mathematischen Voraussetzungen an den Leser entsprechen dem Grundkurs in Mathematik, der an den wirtschaftswissenschaftlichen Fakultäten geboten wird. Unbedingt erforderlich sind Kenntnisse der elementaren Matrizenrechnung, insbesondere der Multiplikation und Inversion von Matrizen. Eine Zusammenfassung der Matrizenrechnung findet man in Kemeny (1984). Die im Buch benötigten Statistikkenntnisse entsprechen ebenfalls dem Grundkurs Statistik wirtschaftswissenschaftlicher Fakultäten. Darüber hinaus sind elementare Kenntnisse der Kovarianzstrukturanalyse notwendig. Die Kenntnis der Kovarianzstrukturmodelle sowie der numerischen Berechnung der diversen Schätzer für diese Modelle mit LISREL kann durch die Lektüre des Lehrbuchs von Pfeifer und Schmidt (1987) erworben werden.

Die im Buch verwendeten LISREL- und LISCOMP-Programme wurden an folgenden Maschinen entwickelt und getestet:

IBM PS2/60 mit 80286/80287 Prozessor; Betriebssystem DOS 3.3.

IBM PS2/80 mit 80386/80387 Prozessor; Betriebssystem DOS 3.3.

VI

Unser Dank gebührt dem Zentrum für Umfragen, Methoden und Analysen e.V. (ZUMA) in Mannheim und dem Westdeutschen Verlag für die hervorragende Zusammenarbeit bei der Erstellung des Buches sowie der Firma IBM Deutschland GmbH, die uns für eine Übergangszeit einen IBM PS2/60 zum Austesten der Programme zur Verfügung stellte. Anregung und ständige Ermutigung, dieses Buch zu schreiben, verdanken wir Herrn Prof. Dr. Hartmut Esser, Universität zu Köln. Große Teile dieses Buches wurden in Seminaren und Workshops bei ZUMA, an der Technischen Universität Berlin sowie am Bundesgesundheitsamt in Berlin als Vorlesungsskripten verwendet und mit den Veranstaltungsteilnehmern diskutiert. Besonders bedanken möchten wir uns bei Dr. Ulrich Küsters, Prof. Dr. Franz Urban Pappi, Prof. Dr. Günter Rothe, Dipl. Oec. Andreas Schepers, Prof. Dr. Michael Sobel und Prof. Dr. Rolf Ziegler, die das Manuskript kritisch gelesen und uns zahlreiche Anregungen zur Verbesserung gegeben haben. Für alle verbleibenden Fehler sind wir natürlich verantwortlich. Unseren Lesern sind wir für kritische Hinweise und Korrekturen dankbar.

Wuppertal, den 5. Mai 1989

Gerhard Arminger und Franz Müller

Inhaltsverzeichnis

1	Statistische Modellbildung für Paneldaten	1
1.1	Allgemeine Bemerkungen	1
1.2	Fehlspezifikation im linearen Modell	2
1.3	Dynamik und unbeobachtete Heterogenität	4
1.4	Modellklassen zur Analyse von Paneldaten	6
1.4.1	Modelle ohne endogene Dynamik	6
1.4.2	Modelle mit endogener Dynamik	10
1.4.3	Modelle für dichotome und ordinale abhängige Variable	11
1.5	Wahl des Beispieldatensatzes und der Analyseprogramme	12
2	Schätzung von Kovarianzstruktur- modellen	15
2.1	Modellformulierung	15
2.2	Pseudo-Maximum-Likelihood-Schätzung	18
2.3	Schätzung aus Mittelwertmodellen	24
2.4	Das Instrumentvariablenprinzip	28
3	Beschreibung des Beispieldatensatzes	31

VIII

4	Behandlung fehlender Daten	37
4.1	Statistische Voraussetzungen	37
4.2	Schätzung der Erwartungswerte und der Kovarianzmatrix bei fehlenden Daten	41
5	Statische Modelle	57
5.1	Einfache statische Modelle	58
5.2	Statische Modelle mit unbeobachteten unabhängigen Variablen . .	63
5.2.1	Varianzkomponentenmodelle	63
5.2.2	Modelle mit unbeobachteten, korrelierten unabhängigen Variablen	68
5.3	Statische Modelle mit Autokorrelation	82
6	Statische Modelle mit latenten Variablen	89
6.1	Einfache statische Modelle mit latenten Variablen	91
6.2	Statische latente Variablenmodelle mit unbeobachteten unabhängigen Variablen	99
6.2.1	Latente Variablenmodelle mit Varianzkomponenten	99
6.2.2	Latente Variablenmodelle mit Differenzenbildung	104
6.3	Latente Variablenmodelle mit Autokorrelation der Fehler im Strukturgleichungsmodell	110
7	Dynamische Modelle	121
7.1	Einfache dynamische Modelle	122
7.2	Dynamische Modelle mit unbeobachteten unabhängigen Variablen	129
7.3	Analyse von Paneldaten mit ungleichen Wellenabständen	138

8	Dynamische Modelle mit latenten Variablen	141
8.1	Einfache dynamische latente Variablenmodelle	142
8.2	Dynamische latente Variablenmodelle mit unbeobachten unabhängigen Variablen	154
8.3	Erweiterung auf mehr als drei Wellen	168
9	Modelle zur Analyse von zensierten, dichotomen und ordinalen Paneldaten	173
9.1	Einleitung	173
9.2	Muthén's LISCOMP-Modell	174
9.2.1	Meßrelationen	175
9.2.2	Mittelwert- und Kovarianzstruktur	178
9.2.3	Einbettung des LISREL-Modells in das LISCOMP-Modell	181
9.3	Sequentialschätzung im LISCOMP-Modell	182
9.4	Empirisches Beispiel	188
9.5	Schlußbemerkungen	201
	Anhang: Datensatz	203
	Literatur	207
	Sachindex	215