Emerging Topics in Statistics and Biostatistics

Jeffrey R. Wilson
Elsa Vazquez-Arreola
(Din) Ding-Geng Chen

Marginal Models in Analysis of Correlated Binary Data with Time Dependent Covariates

Springer

Emerging Topics in Statistics and Biostatistics

Series Editor
(Din) Ding-Geng Chen, University of North Carolina, Chapel Hill, NC, USA

Editorial Board Members
Andriëtte Bekker, University of Pretoria, Pretoria, South Africa
Carlos A. Coelho, Universidade Nova de Lisboa, Caparica, Portugal
Maxim Finkelstein, University of the Free State, Bloemfontein, South Africa
Jeffrey R. Wilson, Arizona State University, Tempe, AZ, USA

More information about this series at http://www.springer.com/series/16213

Jeffrey R. Wilson • Elsa Vazquez-Arreola
(Din) Ding-Geng Chen

Marginal Models in Analysis of Correlated Binary Data with Time Dependent Covariates

Springer

Jeffrey R. Wilson
Department of Economics
W. P. Carey School of Business
Arizona State University
Chandler, AZ, USA

Elsa Vazquez-Arreola
School of Mathematical and Statistical Sciences
Arizona State University
Tempe, AZ, USA

(Din) Ding-Geng Chen
School of Social Work &
Department of Biostatistics
University of North Carolina
Chapel Hill, NC, USA

Department of Statistics
University of Pretoria
Pretoria, South Africa

ISSN 2524-7735 ISSN 2524-7743 (electronic)
Emerging Topics in Statistics and Biostatistics
ISBN 978-3-030-48903-8 ISBN 978-3-030-48904-5 (eBook)
https://doi.org/10.1007/978-3-030-48904-5

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

I dedicate this to my students, present and past. Their insight had a great deal to do with the materials covered in this book.

Jeffrey R. Wilson

I dedicate this to my family for their unconditional support.

Elsa Vazquez-Arreola

I dedicate this to my family for their support.

(Din) Ding-Geng Chen

Preface

In the analysis of correlated data, it is often desirable to evaluate the effect of the time-dependent covariates. However, the changing nature of time-dependent covariates may have delayed effects or feedback. If the relation goes unchecked, one can have a differential effect on the response, and the conventional models may not be appropriate.

The focus of this book is the modeling of correlated response data with time-dependent covariates. We have been accustomed to models for correlated data with time-independent covariates, but modeling correlated data with time-dependent covariates brings some added challenges. These include delayed effects, feedback between responses and covariates, and relation among the responses. This book is then the first book designed to address these challenges with a compilation of research and publications we developed in the past years to address the analysis of correlated data with time-dependent covariates.

Chandler, AZ, USA — Jeffrey R. Wilson
Tempe, AZ, USA — Elsa Vazquez-Arreola
Chapel Hill, NC, USA — (Din) Ding-Geng Chen
Pretoria, South Africa

Acknowledgments

The authors of this book owe a great deal of gratitude to many who helped in the completion of the book. We have been fortunate enough to work with a number of graduate students at Arizona State University: Many thanks to the staff in the Department of Economics and the computing support group in the W. P. Carey School of Business. We also gratefully acknowledge the professional support of Ms. Laura Aileen Briskman from Springer, who made the publication of this book a reality. A special thanks to Dr. Kyle Irimata, Dr. Katherine Irimata, and Dr. Trent Lalonde. To everyone involved in the making of this book, we say thank you!. This work is based on the research supported partially by the National Research Foundation of South Africa (Grant Number 127727) and the South African National Research Foundation (NRF) and South African Medical Research Council (SAMRC) (South African DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant Number 114613).

JRW
EVA
DC

About the Book

In this book, we focus on time-dependent covariates in the fit of marginal models. We use five data sets to demonstrate these models fitted throughout the book. This book consists of eight chapters, and they represent the model development from time-independent to time-dependent developed over the last few years of our research and teaching of statistics at the master's and PhD level at Arizona State University. The aim of this book is to concentrate on using marginal models with their developed theory and the associated practical implementation. The examples in this book are analyzed whenever possible using SAS, but when possible, the R code is provided. The SAS outputs are given in the text with partial tables. The completed data sets and the associated SAS/R programs can be found at the web address www.public.asu.edu/~jeffreyw.

We provide several examples to allow the reader to mimic some of the models used. The chapters in this book are designed to help guide researchers, practitioners, and graduate students to analyze longitudinal data with time-dependent covariates.

The book is timely and has the potential to impact model fitting when faced with correlated data analyses. In an academic setting, the book could serve as a reference guide for a course on time-dependent covariates, particularly for students at the graduate-level statistics or for those seeking degrees in related quantitative fields of study. In addition, this book could serve as a reference for researchers and data analysts in education, social sciences, public health, and biomedical research or wherever clustered and longitudinal data are needed for analysis.

The book is composed of different opportunities for readers. Those interested in quick read can go from Chaps. 1, 2, 5 to 7. While others who wish to know all the details of time-dependent covariates may read Chaps. 1, 2, 5 to 8. However, once the reader is familiar with Chaps. 1 and 2, they can move in different directions as illustrated below (Fig. 1).

When analyzing longitudinal binary data, it is essential to account for both the correlation inherent from the repeated measures of the responses and the correlation realized because of the feedback created between the responses at a particular time and the covariates at other times (Fig. 2). Ignoring any of these correlations can lead to invalid conclusions. Such is the case when the covariates are time-dependent and

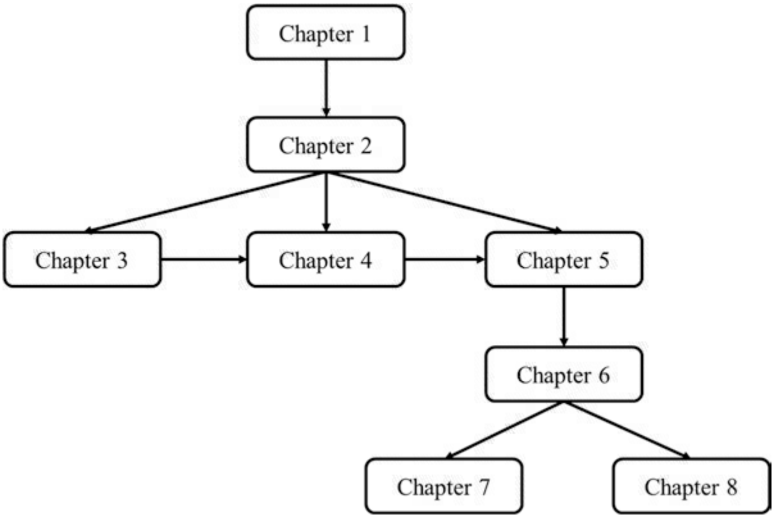

Fig. 1 Suggested system of chapter reading

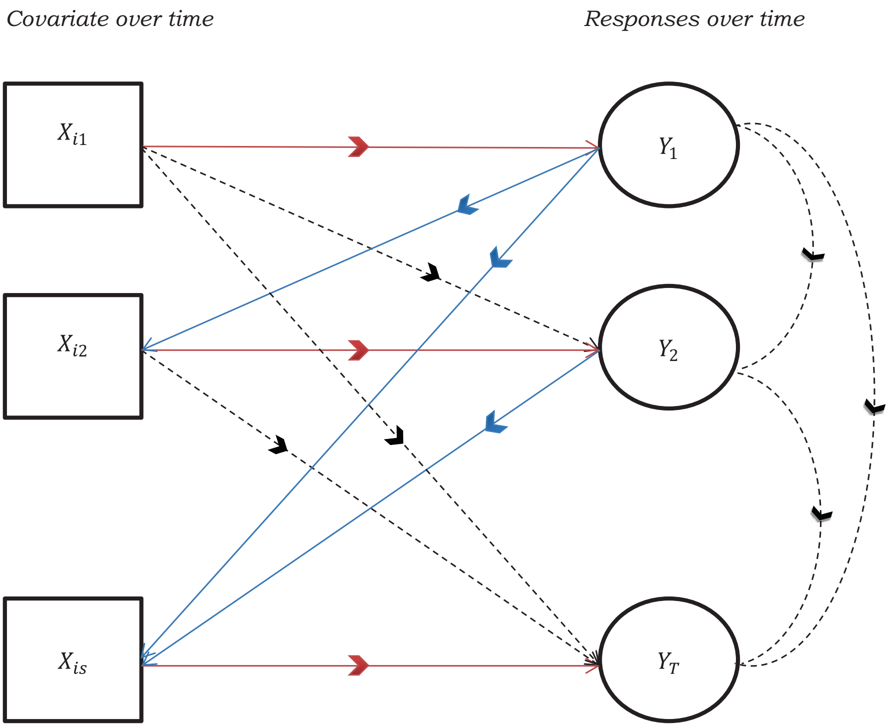

Fig. 2 Two types of correlation structures

the standard logistic regression model is used. Figure 2 describes two types of correlations: responses with responses and responses with covariates. We need a model that addresses both types of relationships. In Fig. 2, the different types of correlation presented are:

1. The correlation among the responses which are denoted by $y_1, \ldots, y_T$ as time t goes from 1 to T and
2. The correlation between response Y_t and covariate X_s:
   (a) When responses at time t impact the covariates in time t+s.
   (b) When the covariates in time t impact the responses in time t+s.

   These correlations regarding feedback from Y_t to the future X_{t+s} and vice versa are important in obtaining the estimates of the regression coefficients.

This book provides a means of modeling repeated responses with time-dependent and time-independent covariates. The coefficients are obtained using the generalized method of moments (GMM). We fit these data using SAs and SAS Macro and at times used R.

We welcome readers' comments, including notes on typos or other errors, and look forward to receiving suggestions for improvements to future editions. Please send comments and suggestions to Professor Jeffrey Wilson (email: jeffrey.wilson@asu.edu).

Contents

List of Figures

List of Tables