

The Work of Raymond J. Carroll

Marie Davidian • Xihong Lin
Jeffrey S. Morris • Leonard A. Stefanski
Editors

The Work of Raymond J. Carroll

The Impact and Influence of a Statistician

 Springer

Editors

Marie Davidian
Department of Statistics
North Carolina State University
Raleigh, NC, USA

Xihong Lin
Department of Biostatistics
Harvard School of Public Health
Boston, MA, USA

Jeffrey S. Morris
Department of Biostatistics
The University of Texas
MD Anderson Cancer Center
Houston, TX, USA

Leonard A. Stefanski
Department of Statistics
North Carolina State University
Raleigh, NC, USA

ISBN 978-3-319-05800-9 ISBN 978-3-319-05801-6 (eBook)
DOI 10.1007/978-3-319-05801-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014939953

© Springer International Publishing Switzerland 2014

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

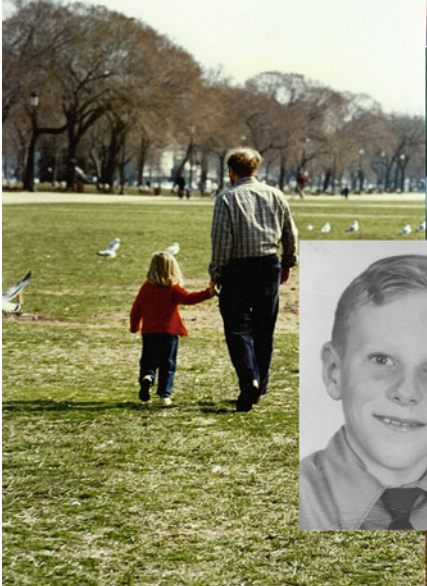
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Ray





Preface

Raymond J. Carroll's impact on statistics and numerous other fields of science is far-reaching and substantial. His vast catalog of work spans the spectrum from fundamental contributions to statistical theory to innovative methodological development to new insights in a number of subject matter areas. From the outset of his career, rather than taking the "safe" route of pursuing incremental advances, Ray has focused on tackling the most important statistical research challenges of our time, and in doing so it is fair to say that he has literally shaped and defined a host of areas of statistics, including weighting and transformation in regression, measurement error modeling, quantitative methods for nutritional epidemiology, and non- and semi-parametric regression. It is indisputable that Ray is one of the giants of the field, and we are honored to have had the opportunity to prepare this volume, which highlights some of his most influential work.

The book is organized into seven main parts, each focused on a key area in which Ray has made significant contributions. The seven subject areas reviewed in this book were chosen by Ray himself, as were the articles representing each area. Each part is focused around these key papers, and, for each, we asked distinguished researchers in the area to provide a commentary giving insight into not only the significance of the featured papers but also on Ray's impact on the area more broadly. The commentaries not only review Ray's work, but they also are filled with history and anecdotes that reflect the fact that Ray is also a really nice guy! Indeed, as former students and collaborators of Ray, we are pleased that the personality, generosity, friendship, and enthusiasm we know so well emerge throughout all of the commentaries, whose authors have almost all had the pleasure of working with Ray firsthand as we have. We are deeply grateful to these contributors, whose thoughtful, insightful commentaries provide an inspiring roadmap to Ray's achievements. Due to their extraordinary efforts, this book is a fitting tribute to a scholar and educator whose influence on not only science but also on the individual students, postdocs, and junior colleagues he has mentored is legendary.

Our elation with the authors who contributed their insights into Ray's work and personality is tempered by the death of George Casella. George provides an entertaining overview of Ray's work in a hodgepodge of "Other" areas. He was both a

close friend and colleague of Ray. We are grateful that George was able to contribute his personal reflections before his passing.

Putting together this volume was made even easier by Ray himself, and we cannot thank him enough. He provided us with extensive materials, including not only the list of articles around which the book is focused but also a detailed narrative of his own thoughts on his work, his biography, and other resources.

We would also like to acknowledge Jennifer Moy, a student at North Carolina State University, whose assistance in preparing Ray's complete bibliography was invaluable.

At the beginning of each commentary, the articles included in this volume that form the basis for the commentary are listed and are identified by acronyms in brackets; for example, "MEM" for Measurement Error Models." The second number in brackets is the number of citations reported by Google Scholar at the time Ray compiled the list (2011).

Of course, a book devoted to the contributions of Raymond Carroll cannot possibly provide a full accounting of his work. Despite approaching the start of his fifth decade as a researcher, Ray has not slowed his pace one bit, and he continues to produce and inspire and mentor students and postdocs unabated. We fully expect to be called upon to put together "Volume 2," featuring still other areas Ray has already influenced and forthcoming contributions in areas that have yet to be defined.

Raleigh, NC
Boston, MA
Houston, TX
Raleigh, NC
December 2012

Marie Davidian
Xihong Lin
Jeffrey S. Morris
Leonard A. Stefanski

Contents

1	Measurement Error	1
	By John P. Buonaccorsi and Aurore Delaigle	
2	Transformation and Weighting	155
	By David Ruppert	
3	Epidemiology	195
	By Laurence Freedman, Mitchell H. Gail, and Dale L. Preston	
4	Nonparametric and Semiparametric Regression for Independent Data	293
	By Hua Liang	
5	Nonparametric and Semiparametric Regression for Dependent Data	371
	By Yehua Li and Naisyin Wang	
6	Robustness	447
	By Roger Koenker and Douglas Simpson	
7	Other Work	509
	By George Casella	
	Bibliography of Raymond J. Carroll	561

Biography of Raymond J. Carroll

Raymond J. Carroll is Distinguished Professor of Statistics, Nutrition, and Toxicology at Texas A&M University, where he has been on the faculty since 1987. He was the first statistician ever given a Method to Extend Research In Time (MERIT) Award from the National Cancer Institute (NCI) of the National Institutes of Health (NIH), receiving this honor for his seminal contributions to statistical methodology and the impact of that methodology on public health. He is the principal investigator of an NCI-funded Bioinformatics training program and is the founding director of the Texas A&M Center for Statistical Bioinformatics. He is also the Director for the Texas A&M Institute of Applied Mathematics and Computational Science (<http://iamcs.tamu.edu>).

Raymond Carroll was born April 21, 1949 in Yokohama, Japan, into an Irish Catholic military family, and he is the eldest of five siblings. His father, who spent the Second World War in India and China, was transferred successively from Yokohama to Nagoya, Japan, Washington DC, Wichita Falls, Texas, Ramstein, Germany, Wichita Falls, Omaha, Nebraska, and Seoul, Korea, and finally retired from his last assignment in Wichita Falls. He is married to Marcia Ory. A memorial tree with a plaque honoring the memory of his parents Regina and Norman is situated in the heart of the central campus a few feet southwest of “Sully,” a bronze statue of the first president of Texas A&M University. Three other memorial trees are adjacent, two honoring the memories of his father-in-law, mother-in-law, and brother-in-law, and the other honoring the memory of Don Risner, a good friend and fishing guide from North Texas. Raymond attended high schools in Germany, Texas, and Nebraska. He graduated from the University of Texas at Austin in 1971 with a BA in mathematics and was especially influenced by courses in analysis and measure theory given by E. W. Cheney and G. W. Stewart, respectively. He received his PhD in Statistics from Purdue in 1974 under the direction of Shanti Gupta, with wonderful advice from Leon Gleser. He has held positions at the University of North Carolina at Chapel Hill and the University of Pennsylvania. He has published over 350 papers and given over 300 invited talks. The peripatetic nature of his childhood has made him an avid traveler, a characteristic not shared by his siblings. Since his first invitation to Australia in 1987, he has visited that country over 20 times, and he

has visited Germany, the site of two of his sabbaticals, nearly yearly since 1980. He is in addition a bad golfer who takes mulligans liberally and a mediocre although enthusiastic fly fisherman.

Dr. Carroll is one of the world's foremost experts on problems of measurement error, data transformation, and nonconstant variation, and more generally on statistical regression modeling. His work has found application in a broad variety of fields, including marine biology, laboratory assay methods, econometrics, epidemiology, molecular biology, and many others. He has served as Editor of *Biometrics*, a journal of the International Biometric Society, and as Editor of the *Journal of the American Statistical Association (JASA)* Theory and Methods section. He has won many honors in the profession, including the two major research awards. The first is the 1988 Committee of Presidents of Statistical Societies (COPSS) Presidents' Award, given annually by five major statistical societies to the outstanding statistician under the age of 40. Secondly, he gave the COPSS Fisher Lecture at the 2002 Joint Statistical Meetings, an award given by these statistical societies in honor of a senior statistician "whose research has influenced the theory and practice of statistics."

Carroll's work is characterized by a combination of deep theoretical advances, innovative methodological development, and close contact with science. His first seminal contribution to statistical methodology was to create methods for the analysis of data with nonconstant variation; these methods being the transform-both-sides method for nonlinear regression (together with David Ruppert) and the variance function estimation approach (with Marie Davidian), both still in wide use. This work developed from two projects, one on marine fisheries where he worked with a team investigating how to model and manage the menhaden fishery in the Atlantic, and the other project involving immunoassays at Eli Lilly and Company. In the early 1990s, with the inspiration of his close friend Mitchell Gail, he developed a deep interest in epidemiologic case-control studies that led to his receiving the George W. Snedecor Award from COPSS in 1997 for work in this area (together with Bruce Lindsay and Katherine Roeder). The span of his scientific work is amazing, including among many others (a) modeling ozone exposure in Houston (the 1997 *JASA* Applications Editor's Invited Paper); (b) understanding the effects of diet on breast cancer; and (c) discovering interactions between genes and the environment (with Nilanjan Chatterjee and Yi-Hau Chen).

Carroll is no doubt most well known for his work in the area of nonlinear measurement error modeling, with applications to nutritional and radiation epidemiology. The body of seminal research is of such depth, and of such importance, that at the International Biometric Conference in 2000 in Berkeley, Scott Zeger described him as the "grandfather" of measurement error modeling. His 1995 book and 2006 second edition with David Ruppert, Len Stefanski, and Ciprian Crainiceanu is the standard reference in the field. This work began with his landmark 1984 paper in *Biometrika* on measurement error in the binary regression framework and has continued to the present. He was the first to suggest the use of likelihood methods in the nonlinear measurement error context. Along with Len Stefanski, he developed the theory for and coined the name for regression calibration, the most commonly

used method in nutritional epidemiology. His 1987 paper with Stefanski developed the method of conditional score function. His 1990 paper with Stefanski and his 1988 paper with Peter Hall on deconvolution established the theoretical basis showing how difficult it really is to understand latent variable distributions: this result provides the theoretical underpinnings for the semi-parametric approaches in measurement error models that have become increasingly popular. The deconvolution area has become of great importance and interest, and even 20 years later the papers have led others into the area. Carroll continues to produce important ideas, and his work continues to influence others, in such important problems as mixed models, segmented regression, instrumental variables, and nonparametric regression. More recently, he has written papers on reanalysis of important radiation epidemiology studies to account for measurement error, both in *Biometrics*.

Carroll's work on measurement error modeling is also one of the landmark works in nutritional epidemiology. He helped design the NCI-AARP Diet and Health Study, the first study to confirm a link between fat in diet and breast cancer. He was the senior author on the first major biomarker study (the OPEN Study) to understand how well common instruments such as the food frequency questionnaire actually measure diet. This study was funded because of the methodological developments done together in what is now a long collaboration with Laurence Freedman, Victor Kipnis, and Douglas Midthune suggesting that the heart of the problem of null studies was the instruments themselves.

Dr. Carroll has worked with many researchers from around the world, but no doubt his closest collaboration has been with David Ruppert, now of Cornell University. They were next door office neighbors at the University of North Carolina from 1977 to 1987, where they started their original collaboration, and they have written over 45 papers in addition to 4 books. Other colleagues with whom he has written 10 or more papers include Mitchell Gail, Victor Kipnis, and Douglas Midthune of the National Cancer Institute; Peter Hall of the University of Melbourne; Len Stefanski of North Carolina State University; Laurence Freedman of the Gertner Institute in Israel; Naisyin Wang of the University of Michigan; Joanne Lupton, Nancy Turner and Robb Chapkin, nutritionists at Texas A&M; Xihong Lin of Harvard; and Bani Mallick of Texas A&M.

More recently, Dr. Carroll has developed a deep interest in basic molecular cell biology and how it relates to nutrition and colon carcinogenesis. His research grants include as co-investigator Dr. Joanne Lupton (the endowed Professor of Human Nutrition at Texas A&M) and Dr. Nancy Turner. This work includes papers both in biology journals and in *JASA* and *Biostatistics*, with many more papers under development. Carroll is involved to the point of generating his own biological hypotheses, suggesting new ways of measurement, and providing support so that novel measurements can be undertaken to understand molecular pathways. More recently, this close work with biologists and electrical engineers has led to the establishment of an NCI-funded training program in Biostatistics and Bioinformatics, for which Dr. Carroll has been the principal investigator since 2001, and the program has recently been renewed until 2016. The program is unique because it aims to train

statisticians and electrical engineers in biology and includes mentors from biological fields.

Dr. Carroll is an inspirational teacher and a major innovator for the Department's teaching program. In the 1990s he introduced the use of the computer and class projects into STAT 302, an undergraduate course aimed at life science students. Similarly, since 2000, in STAT 651 he was the first non-distance education expert to create a distance course, something now routine in the department. Dr. Carroll has won a College of Science Teaching Award, and he has graduated 35 PhD students, many of whom are leading figures in academia and industry. He has also been the mentor to many faculty members around the USA, including many who are now full professors, and he is legendary for his willingness to give advice and technical assistance.

PhD Students of Raymond J. Carroll

Name	Institution	Year
Gordon Johnston	University of North Carolina at Chapel Hill	1979
Paul Gallo	University of North Carolina at Chapel Hill	1981
David Giltinan	University of North Carolina at Chapel Hill	1983
Len Stefanski	University of North Carolina at Chapel Hill	1983
Doug Simpson	University of North Carolina at Chapel Hill	1985
Marie Davidian	University of North Carolina at Chapel Hill	1986
Stena Kettl	University of North Carolina at Chapel Hill	1987
Yin Yin	University of North Carolina at Chapel Hill	1988
Lie Ju Hwang	Texas A&M University	1990
Jungsywan Sepanski	Texas A&M University	1991
Rick Landin	Texas A&M University	1992
C. Y. Wang	Texas A&M University	1993
Ron Knickerbocker	Texas A&M University	1993
Bobby Gutierrez	Texas A&M University	1995
Stephen Eckert	Texas A&M University	1995
Jeff Maca	Texas A&M University	1997
Christian Galindo	Texas A&M University	1998
Steve Iturria	Texas A&M University	1998
Jeffrey S. Morris	Texas A&M University	2000
Hua Liang	Texas A&M University	2001
Inyoung Kim	Texas A&M University	2002
Chan Hee Jo	Texas A&M University	2003
Tanya Apanasovich	Texas A&M University	2004
Gosia Leyk	Texas A&M University	2004
Christie Spinka	Texas A&M University	2004
Veera Baladandayuthapani	Texas A&M University	2005
Yehua Li	Texas A&M University	2006
Iryna Lobach	Texas A&M University	2006
Bo Li	Texas A&M University	2006

Name	Institution	Year
Lian Liu	Texas A&M University	2007
Arnab Maity	Texas A&M University	2008
Seokho Lee	Texas A&M University	2009
Andrew Redd	Texas A&M University	2010
Jiawei Wei	Texas A&M University	2010
Saijuan Zhang	Texas A&M University	2010
Trijya Singh	Texas A&M University	2011
Xiaolei Xun	Texas A&M University	2012

Contributors

John P. Buonaccorsi

University of Massachusetts, Amherst, MA, USA

George Casella

University of Florida, Gainesville, FL, USA

(deceased)

Aurore Delaigle

University of Melbourne, Melbourne, Australia

Laurence (Larry) Freedman

Gertner Institute for Epidemiology, Tel Hashomer, Israel

Mitchell H. Gail

National Cancer Institute, Bethesda, MD, USA

Roger Koenker

University of Illinois, Urbana-Champaign, IL, USA

Yehua Li

Iowa State University, Ames, IA, USA

Hua Liang

The George Washington University, Washington, DC, USA

Dale L. Preston

Hirosoft International, Eureka, CA, USA

David Ruppert

Cornell University, Ithaca, NY, USA

Douglas Simpson

University of Illinois, Urbana-Champaign, IL, USA

Naisyin Wang

University of Michigan, Ann Arbor, MI, USA