Quantile Regression

Quantile regression is gradually emerging as a unified statistical methodology for estimating models of conditional quantile functions. By complementing the exclusive focus of classical least-squares regression on the conditional mean, quantile regression offers a systematic strategy for examining how covariates influence the location, scale, and shape of the entire response distribution. This monograph is the first comprehensive treatment of the subject, encompassing models that are linear and nonlinear, parametric and nonparametric. The author has devoted more than 25 years of research to this topic. The methods are illustrated with a variety of applications from economics, biology, ecology, and finance. The treatment will find its core audiences in econometrics, statistics, and biostatistics.

Roger Koenker is McKinley Professor of Economics and Professor of Statistics at the University of Illinois at Urbana-Champaign. From 1976 to 1983 he was a member of the technical staff at Bell Laboratories. He has held visiting positions at The University of Pennsylvania; Charles University, Prague; Nuffield College, Oxford; University College London; and Australian National University. He is a Fellow of the Econometric Society.

Econometric Society Monographs No. 38

Editors:
Andrew Chesher, University College London
Matthew Jackson, California Institute of Technology

The Econometric Society is an international society for the advancement of economic theory in relation to statistics and mathematics. The Econometric Society Monograph Series is designed to promote the publication of original research contributions of high quality in mathematical economics and theoretical and applied econometrics.

Other titles in the series:
G. S. Maddala *Limited dependent and qualitative variables in econometrics*, 0 521 33825 5 – ESM 3
Gerard Debreu *Mathematical economics: Twenty papers of Gerard Debreu*, 0 521 33561 2 – ESM 4
Jean-Michel Grandmont *Money and value: A reconsideration of classical and neoclassical monetary economics*, 0 521 31364 3 – ESM 5
Franklin M. Fisher *Disequilibrium foundations of equilibrium economics*, 0 521 37856 7 – ESM 6
Andreu Mas-Colell *The theory of general economic equilibrium: A differentiable approach*, 0 521 26514 2, 0 521 38870 8 – ESM 9
Truman F. Bewley, Editor *Advances in econometrics – Fifth World Congress (Volume I)*, 0 521 46726 8 – ESM 13
Truman F. Bewley, Editor *Advances in econometrics – Fifth World Congress (Volume II)*, 0 521 46725 X – ESM 14
Herv́e Moulin *Axioms of cooperative decision making*, 0 521 36055 2, 0 521 42458 5 – ESM 15
L. G. Godfrey *Misspecification tests in econometrics: The Lagrange multiplier principle and other approaches*, 0 521 42459 3 – ESM 16
Tony Lancaster *The econometric analysis of transition data*, 0 521 43789 X – ESM 17
Alvin E. Roth and Marilda A. Oliviera Sotomayor, Editors *Two-sided matching: A study in game-theoretic modeling and analysis*, 0 521 43788 1 – ESM 18
Wolfgang Härdle, *Applied nonparametric regression*, 0 521 42950 1 – ESM 19
Jean-Jacques Laffont, Editor *Advances in economic theory – Sixth World Congress (Volume I)*, 0 521 48459 6 – ESM 20
Jean-Jacques Laffont, Editor *Advances in economic theory – Sixth World Congress (Volume II)*, 0 521 48460 X – ESM 21
Halbert White *Estimation, inference and specification*, 0 521 25280 6, 0 521 57446 3 – ESM 22
Christopher Sims, Editor *Advances in econometrics – Sixth World Congress (Volume I)*, 0 521 56610 X – ESM 23
Christopher Sims, Editor *Advances in econometrics – Sixth World Congress (Volume II)*, 0 521 56609 6 – ESM 24
Roger Guesnerie *A contribution to the pure theory of taxation*, 0 521 23689 4, 0 521 62956 X – ESM 25
David M. Kreps and Kenneth F. Wallis, Editors *Advances in economics and econometrics – Seventh World Congress (Volume I)*, 0 521 58011 0, 0 521 58983 5 – ESM 26
David M. Kreps and Kenneth F. Wallis, Editors *Advances in economics and econometrics – Seventh World Congress (Volume II)*, 0 521 58012 9, 0 521 58982 7 – ESM 27
David M. Kreps and Kenneth F. Wallis, Editors *Advances in economics and econometrics – Seventh World Congress (Volume III)*, 0 521 58013 7, 0 521 58981 9 – ESM 28
Donald P. Jacobs, Ehud Kalai, and Morton I. Kamien, Editors *Frontiers of research in economic theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, 0 521 63222 6, 0 521 63538 1 – ESM 29
A. Colin Cameron and Pravin K. Trivedi *Regression analysis of count data*, 0 521 63201 3, 0 521 63567 5 – ESM 30
Steinar Strøm, Editor *Econometrics and economic theory in the 20th century: The Ragnar Frisch Centennial Symposium*, 0 521 63323 0, 0 521 63365 6 – ESM 31
Eric Ghysels, Norman R. Swanson, and Mark Watson, Editors *Essays in econometrics: Collected papers of Clive W. J. Granger (Volume I)*, 0 521 77297 4, 0 521 77496 9, 0 521 79697 0 (2 volume paperback set) – ESM 32
Eric Ghysels, Norman R. Swanson, and Mark Watson, Editors *Essays in econometrics: Collected papers of Clive W. J. Granger (Volume II)*, 0 521 79207 X, 0 521 79649 0, 0 521 79697 0 (2 volume paperback set) – ESM 33
Cheng Hsiao, *Analysis of panel data*, second edition, 0 521 81855 9, 0 521 52271 4 – ESM 34
Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, Editors *Advances in economics and econometrics – Eighth World Congress (Volume I)*, 0 521 81872 8, 0 521 52411 3 – ESM 35
Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, Editors *Advances in economics and econometrics – Eighth World Congress (Volume II)*, 0 521 81873 7, 0 521 52412 1 – ESM 36
Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, Editors *Advances in economics and econometrics – Eighth World Congress (Volume III)*, 0 521 81874 5, 0 521 52413 X – ESM 37

Quantile Regression

Roger Koenker
University of Illinois

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521845731

First published 2005

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Koenker, Roger, 1947–
Quantile regression / Roger Koenker.
p. cm. – (Econometric Society monographs ; no. 38)
Includes bibliographical references and index.
ISBN 0-521-84573-4 (hardcover) – ISBN 0-521-60827-9 (pbk.)
1. Regression analysis. 2. Mathematical statistics. I. Title. II. Series.

QA278.2.K64 2005
519.5′36 – dc22 2004027656

ISBN-13 978-0-521-84573-1 hardback
ISBN-10 0-521-84573-4 hardback

ISBN-13 978-0-521-60827-5 paperback
ISBN-10 0-521-60827-9 paperback

To Emma, in memoriam

Contents

Preface

Francis Galton in a famous passage defending the "charms of statistics" against its many detractors, chided his statistical colleagues

> [who] limited their inquiries to Averages, and do not seem to revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of a native of one of our flat English counties, whose retrospect of Switzerland was that, if the mountains could be thrown into its lakes, two nuisances would be got rid of at once (*Natural Inheritance,* p. 62).

It is the fundamental task of statistics to bring order out of the diversity – at times the apparent chaos – of scientific observation. And this task is often very effectively accomplished by exploring how *averages* of certain variables depend on the values of other "conditioning" variables. The method of least squares, which pervades statistics, is admirably suited for this purpose. And yet, like Galton, one may question whether the exclusive focus on conditional mean relations among variables ignores some "charm of variety" in matters statistical.

As a resident of one of the flattest American counties, my recollections of Switzerland and its attractive nuisances are quite different from the retrospect described by Galton. Not only the Swiss landscape, but also many of its distinguished statisticians have in recent years made us more aware of the charms and perils of the diversity of observations and the consequences of too blindly limiting our inquiry to averages.

Quantile regression offers the opportunity for a more complete view of the statistical landscape and the relationships among stochastic variables. The simple expedient of replacing the familiar notions of sorting and ranking observations in the most elementary one-sample context by *optimization* enables us to extend these ideas to a much broader class of statistical models. Just as minimizing sums of squares permits us to estimate a wide variety of models for conditional mean functions, minimizing a simple asymmetric version of absolute errors yields estimates for conditional quantile functions. For linear parametric models, computation is greatly facilitated by the reformulation of the optimization problem as a parametric linear program. Formal duality

results for linear programs yield a new approach to rank statistics and rank-based inference for linear models.

I hope that this book can provide a comprehensive introduction to quantile regression methods and that it will serve to stimulate others to explore and further develop these ideas in their own research. Because ultimately the test of any statistical method must be its success in applications, I have sought to illustrate the application of quantile regression methods throughout the book wherever possible. Formal mathematical development, which plays an indispensable role in clarifying precise conditions under which statistical methods can be expected to perform reliably and efficiently, are generally downplayed, but Chapter 4 is devoted to an exposition of some of the basic asymptotic theory of quantile regression, and other chapters include technical material that provides further mathematical details.

Statistical software for quantile regression is now widely available in many well-known statistical packages. Fellow `R` users will undoubtedly recognize by the graphics that I am an `R`-ophile. I have devoted considerable research energy over the years to the development of software for quantile regression, first while I was at Bell Laboratories in the `S` language of John Chambers, later in `S`'s commercial manifestation Splus, and in recent years in its splendid open-source embodiment `R`.

I am extremely grateful to many colleagues who have, over the years, collaborated on various aspects of the work described here. Gib Bassett, first and foremost, whose Ph.D. thesis on l_1-regression served as a springboard for much of our subsequent work on this subject, has been a continuing source of insight and enthusiastic support. Jana Jurečková, who took an early interest in this line of research, has made an enormous contribution to the subject, especially in developing the close connection between quantile regression ideas and rank statistics in work with Cornelius Gutenbrunner. Recently, I have had the pleasure of working with Zhijie Xiao and Ivan Mizera on time-series and multivariate smoothing problems, respectively. Independent work by David Ruppert, Ray Carroll, Alan Welsh, Tertius Dewet, Jim Powell, Gary Chamberlain, Xuming He, Keith Knight, Probal Chaudhuri, Hira Koul, Marc Hallin, Brian Cade, Moshe Buchinsky, Berndt Fitzenberger, Victor Chernozhukov, and Andrew Chesher, among others, has also played a crucial role in the development of these ideas.

Continuing collaboration over the past decade with a number of Ph.D. students including José Machado, Pin Ng, Quanshui Zhao, Yannis Bilias, Beum-Jo Park, M. N. Hasan, Daniel Morillo, Ted Juhl, Olga Geling, Gregory Kordas, Lingjie Ma, Ying Wei, Roberto Perrelli, and Carlos Lamarche has been especially rewarding.

Some of the material was originally prepared for short courses offered at the 1997 Brazilian Statistical Association meeting at Campos do Jordão, the 2001 South African Statistical Association Annual Conference in Goudini Spa, in the spring of 2003 at University College London, and in the fall of 2003 at

the NAKE Workshop in Groningen. I would like to express my appreciation to organizers and participants for their comments and suggestions.

I would like to specially thank Steve Portnoy, with whom I have collaborated over many years on many aspects of quantile regression. He has been a great source of encouragement and inspiration from our earliest discussions in the mid-1970s. Originally, we had planned to write this book together, but in the end Steve's impatience with revisiting the scenes of old research and his eagerness to get on with the new proved even greater than my own, and I have had to fend for myself; readers are the poorer for it.

I would also like to express my appreciation to various institutions that have provided hospitable environments for this research: Bell Laboratories, Australian National University, Charles University, University College London, and my academic home for most of my career, the University of Illinois. Research support by the National Science Foundation over an extended period has contributed significantly and is deeply appreciated.

And finally, my most heartfelt thanks – to my wife Diane and daughter Hannah.

Urbana, July 2, 2004