

## 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  
 Hervé 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

Cambridge University Press  
0521845734 - Quantile Regression  
Roger Koenker  
Frontmatter  
[More information](#)

---

# Quantile Regression

Roger Koenker  
*University of Illinois*



Cambridge University Press  
0521845734 - Quantile Regression  
Roger Koenker  
Frontmatter  
[More information](#)

---

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](http://www.cambridge.org)  
Information on this title: [www.cambridge.org/9780521845731](http://www.cambridge.org/9780521845731)

© Roger Koenker 2005

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

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

Cambridge University Press has no responsibility for  
the persistence or accuracy of URLs for external or  
third-party Internet Web sites referred to in this book  
and does not guarantee that any content on such  
Web sites is, or will remain, accurate or appropriate.

Cambridge University Press  
0521845734 - Quantile Regression  
Roger Koenker  
Frontmatter  
[More information](#)

---

*To Emma, in memoriam*

## Contents

<i>Preface</i>	<i>page</i>	<i>xiii</i>
1 Introduction		1
1.1 Means and Ends		1
1.2 The First Regression: A Historical Prelude		2
1.3 Quantiles, Ranks, and Optimization		5
1.4 Preview of Quantile Regression		9
1.5 Three Examples		13
1.5.1 Salaries versus Experience		13
1.5.2 Student Course Evaluations and Class Size		17
1.5.3 Infant Birth Weight		20
1.6 Conclusion		25
2 Fundamentals of Quantile Regression		26
2.1 Quantile Treatment Effects		26
2.2 How Does Quantile Regression Work?		32
2.2.1 Regression Quantiles Interpolate $p$ Observations		33
2.2.2 The Subgradient Condition		34
2.2.3 Equivariance		38
2.2.4 Censoring		40
2.3 Robustness		42
2.3.1 The Influence Function		42
2.3.2 The Breakdown Point		45
2.4 Interpreting Quantile Regression Models		47
2.4.1 Some Examples		48
2.5 Caution: Quantile Crossing		55
2.6 A Random Coefficient Interpretation		59
2.7 Inequality Measures and Their Decomposition		62
2.8 Expectiles and Other Variations		63
2.9 Interpreting Misspecified Quantile Regressions		65
2.10 Problems		66

viii	<b>Contents</b>	
3	Inference for Quantile Regression	68
3.1	The Finite-Sample Distribution of Regression Quantiles	68
3.2	A Heuristic Introduction to Quantile Regression	
	Asymptotics	71
3.2.1	Confidence Intervals for the Sample Quantiles	72
3.2.2	Quantile Regression Asymptotics with IID Errors	73
3.2.3	Quantile Regression Asymptotics in Non-IID Settings	74
3.3	Wald Tests	75
3.3.1	Two-Sample Tests of Location Shift	75
3.3.2	General Linear Hypotheses	76
3.4	Estimation of Asymptotic Covariance Matrices	77
3.4.1	Scalar Sparsity Estimation	77
3.4.2	Covariance Matrix Estimation in Non-IID Settings	79
3.5	Rank-Based Inference	81
3.5.1	Rank Tests for Two-Sample Location Shift	81
3.5.2	Linear Rank Statistics	84
3.5.3	Asymptotics of Linear Rank Statistics	85
3.5.4	Rank Tests Based on Regression Rankscores	87
3.5.5	Confidence Intervals Based on Regression Rankscores	91
3.6	Quantile Likelihood Ratio Tests	92
3.7	Inference on the Quantile Regression Process	95
3.7.1	Wald Processes	97
3.7.2	Quantile Likelihood Ratio Processes	98
3.7.3	The Regression Rankscore Process Revisited	98
3.8	Tests of the Location-Scale Hypothesis	98
3.9	Resampling Methods and the Bootstrap	105
3.9.1	Bootstrap Refinements, Smoothing, and Subsampling	107
3.9.2	Resampling on the Subgradient Condition	108
3.10	Monte Carlo Comparison of Methods	110
3.10.1	Model 1: A Location-Shift Model	111
3.10.2	Model 2: A Location-Scale-Shift Model	112
3.11	Problems	113
4	Asymptotic Theory of Quantile Regression	116
4.1	Consistency	117
4.1.1	Univariate Sample Quantiles	117
4.1.2	Linear Quantile Regression	118
4.2	Rates of Convergence	120
4.3	Bahadur Representation	122
4.4	Nonlinear Quantile Regression	123
4.5	The Quantile Regression Rankscore Process	124
4.6	Quantile Regression Asymptotics under Dependent Conditions	126

Contents	ix
4.6.1 Autoregression	126
4.6.2 ARMA Models	128
4.6.3 ARCH-like Models	129
4.7 Extremal Quantile Regression	130
4.8 The Method of Quantiles	131
4.9 Model Selection, Penalties, and Large- $p$ Asymptotics	133
4.9.1 Model Selection	134
4.9.2 Penalty Methods	135
4.10 Asymptotics for Inference	138
4.10.1 Scalar Sparsity Estimation	139
4.10.2 Covariance Matrix Estimation	141
4.11 Resampling Schemes and the Bootstrap	141
4.12 Asymptotics for the Quantile Regression Process	142
4.12.1 The Durbin Problem	142
4.12.2 Khmaladization of the Parametric Empirical Process	144
4.12.3 The Parametric Quantile Process	145
4.12.4 The Parametric Quantile Regression Process	146
4.13 Problems	149
5 L-Statistics and Weighted Quantile Regression	151
5.1 L-Statistics for the Linear Model	151
5.1.1 Optimal L-Estimators of Location and Scale	153
5.1.2 L-Estimation for the Linear Model	155
5.2 Kernel Smoothing for Quantile Regression	158
5.2.1 Kernel Smoothing of the $\rho_\tau$ -Function	160
5.3 Weighted Quantile Regression	160
5.3.1 Weighted Linear Quantile Regression	160
5.3.2 Estimating Weights	161
5.4 Quantile Regression for Location–Scale Models	164
5.5 Weighted Sums of $\rho_\tau$ -Functions	168
5.6 Problems	170
6 Computational Aspects of Quantile Regression	173
6.1 Introduction to Linear Programming	173
6.1.1 Vertices	174
6.1.2 Directions of Descent	176
6.1.3 Conditions for Optimality	177
6.1.4 Complementary Slackness	178
6.1.5 Duality	180
6.2 Simplex Methods for Quantile Regression	181
6.3 Parametric Programming for Quantile Regression	185
6.3.1 Parametric Programming for Regression Rank Tests	188
6.4 Interior Point Methods for Canonical LPs	190
6.4.1 Newton to the Max: An Elementary Example	193
6.4.2 Interior Point Methods for Quantile Regression	199

x	<b>Contents</b>	
	6.4.3 Interior vs. Exterior: A Computational Comparison	202
	6.4.4 Computational Complexity	204
	6.5 Preprocessing for Quantile Regression	206
	6.5.1 “Selecting” Univariate Quantiles	207
	6.5.2 Implementation	207
	6.5.3 Confidence Bands	208
	6.5.4 Choosing $m$	209
	6.6 Nonlinear Quantile Regression	211
	6.7 Inequality Constraints	213
	6.8 Weighted Sums of $\rho_\tau$ -Functions	214
	6.9 Sparsity	216
	6.10 Conclusion	220
	6.11 Problems	220
7	<b>Nonparametric Quantile Regression</b>	222
	7.1 Locally Polynomial Quantile Regression	222
	7.1.1 Average Derivative Estimation	226
	7.1.2 Additive Models	228
	7.2 Penalty Methods for Univariate Smoothing	229
	7.2.1 Univariate Roughness Penalties	229
	7.2.2 Total Variation Roughness Penalties	230
	7.3 Penalty Methods for Bivariate Smoothing	235
	7.3.1 Bivariate Total Variation Roughness Penalties	235
	7.3.2 Total Variation Penalties for Triograms	236
	7.3.3 Penalized Triogram Estimation as a Linear Program	240
	7.3.4 On Triangulation	241
	7.3.5 On Sparsity	242
	7.3.6 Automatic $\lambda$ Selection	242
	7.3.7 Boundary and Qualitative Constraints	243
	7.3.8 A Model of Chicago Land Values	243
	7.3.9 Taut Strings and Edge Detection	246
	7.4 Additive Models and the Role of Sparsity	248
8	<b>Twilight Zone of Quantile Regression</b>	250
	8.1 Quantile Regression for Survival Data	250
	8.1.1 Quantile Functions or Hazard Functions?	252
	8.1.2 Censoring	253
	8.2 Discrete Response Models	255
	8.2.1 Binary Response	255
	8.2.2 Count Data	259
	8.3 Quantile Autoregression	260
	8.3.1 Quantile Autoregression and Comonotonicity	261
	8.4 Copula Functions and Nonlinear Quantile Regression	265
	8.4.1 Copula Functions	265

Contents	xi
8.5 High-Breakdown Alternatives to Quantile Regression	268
8.6 Multivariate Quantiles	272
8.6.1 The Oja Median and Its Extensions	273
8.6.2 Half-Space Depth and Directional Quantile Regression	275
8.7 Penalty Methods for Longitudinal Data	276
8.7.1 Classical Random Effects as Penalized Least Squares	276
8.7.2 Quantile Regression with Penalized Fixed Effects	278
8.8 Causal Effects and Structural Models	281
8.8.1 Structural Equation Models	281
8.8.2 Chesher's Causal Chain Model	283
8.8.3 Interpretation of Structural Quantile Effects	284
8.8.4 Estimation and Inference	285
8.9 Choquet Utility, Risk, and Pessimistic Portfolios	287
8.9.1 Choquet Expected Utility	287
8.9.2 Choquet Risk Assessment	289
8.9.3 Pessimistic Portfolios	291
9 Conclusion	293
A Quantile Regression in R: A Vignette	295
A.1 Introduction	295
A.2 What Is a Vignette?	296
A.3 Getting Started	296
A.4 Object Orientation	298
A.5 Formal Inference	299
A.6 More on Testing	305
A.7 Inference on the Quantile Regression Process	307
A.8 Nonlinear Quantile Regression	308
A.9 Nonparametric Quantile Regression	310
A.10 Conclusion	316
B Asymptotic Critical Values	317
<i>References</i>	319
<i>Name Index</i>	337
<i>Subject Index</i>	342

## 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 Biliias, 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