CLADAG 2021

BOOK OF ABSTRACTS AND SHORT PAPERS

13th Scientific Meeting of the Classification and Data Analysis Group

Firenze, September 9-11, 2021

edited by

Giovanni C. Porzio

Carla Rampichini

Chiara Bocci

FIRENZE UNIVERSITY PRESS

PROCEEDINGS E REPORT
ISSN 2704-601X (PRINT) - ISSN 2704-5846 (ONLINE)

SCIENTIFIC PROGRAM COMMITTEE

Giovanni C. Porzio (chair) (University of Cassino and Southern Lazio - Italy)

Silvia Bianconcini (University of Bologna - Italy)
Christophe Biernacki (University of Lille - France)
Paula Brito (University of Porto - Portugal)
Francesca Marta Lilja Di Lascio (Free University of Bozen-Bolzano - Italy)
Marco Di Marzio ("Gabriele d'Annunzio" University of Chieti-Pescara - Italy)
Alessio Farcomeni ("Tor Vergata" University of Rome - Italy)
Luca Frigau (University of Cagliari - Italy)
Luis Ángel García Escudero (University of Valladolid - Spain)
Bettina Grün (Vienna University of Economics and Business - Austria)
Salvatore Ingrassia (University of Catania - Italy)
Volodymyr Melnykov (University of Alabama - USA)
Brendan Murphy (University College Dublin -Ireland)
Maria Lucia Parrella (University of Salerno - Italy)
Carla Rampichini (University of Florence - Italy)
Monia Ranalli (Sapienza University of Rome - Italy)
J. Sunil Rao (University of Miami - USA)
Marco Riani (University of di Parma - Italy)
Nicola Salvati (University of Pisa - Italy)
Laura Maria Sangalli (Polytechnic University of Milan - Italy)
Bruno Scarpa (University of Padua - Italy)
Mariangela Sciandra (University of Palermo - Italy)
Luca Scrucca (University of Perugia - Italy)
Domenico Vistocco (Federico II University of Naples - Italy)
Mariangela Zenga (University of Milan-Bicocca - Italy)

LOCAL PROGRAM COMMITTEE

Carla Rampichini (chair) (University of Florence - Italy)

Chiara Bocci (University of Florence - Italy)
Anna Gottard (University of Florence - Italy)
Leonardo Grilli (University of Florence - Italy)
Monia Lupparelli (University of Florence - Italy)
Maria Francesca Marino (University of Florence - Italy)
Agnese Panzera (University of Florence - Italy)
Emilia Rocco (University of Florence - Italy)
Domenico Vistocco (Federico II University of Naples - Italy)

CLADAG 2021
BOOK OF ABSTRACTS AND SHORT PAPERS

13th Scientific Meeting of the Classification
and Data Analysis Group
Firenze, September 9-11, 2021

edited by
Giovanni C. Porzio
Carla Rampichini
Chiara Bocci

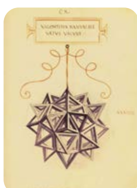
FIRENZE UNIVERSITY PRESS
2021

CLADAG 2021 BOOK OF ABSTRACTS AND SHORT PAPERS : 13th Scientific Meeting of the Classification and Data Analysis Group Firenze, September 9-11, 2021/ edited by Giovanni C. Porzio, Carla Rampichini, Chiara Bocci. — Firenze : Firenze University Press, 2021.
(Proceedings e report ; 128)

https://www.fupress.com/isbn/9788855183406

ISSN 2704-601X (print)
ISSN 2704-5846 (online)
ISBN 978-88-5518-340-6 (PDF)
ISBN 978-88-5518-341-3 (XML)
DOI 10.36253/978-88-5518-340-6

Graphic design: Alberto Pizarro Fernández, Lettera Meccanica SRLs
Front cover: Illustration of the statue by Giambologna, *Appennino* (1579-1580) by Anna Gottard



CLAssification and Data
Analysis Group (CLADAG)
of the Italian Statistical
Society (SIS)

Published by Firenze University Press
Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

This book is printed on acid-free paper
Printed in Italy

INDEX

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Giovanni C. Porzio, Carla Rampichini, Chiara Bocci (edited by), *CLADAG 2021 Book of abstracts and short papers. 13th Scientific Meeting of the Classification and Data Analysis Group Firenze, September 9-11, 2021*, © 2021 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN 978-88-5518-340-6 (PDF), DOI 10.36253/978-88-5518-340-6

Contributed Papers

UNCONDITIONAL M-QUANTILE REGRESSION

Luca Merlo[1], Lea Petrella[2] and Nikos Tzavidis[3]

[1] Department of Statistics, Sapienza University of Rome,
(e-mail: luca.merlo@uniroma1.it)

[2] MEMOTEF Department, Sapienza University of Rome,
(e-mail: lea.petrella@uniroma1.it)

[3] Department of Social Statistics and Demography and Southampton Statistical Sciences Research Institute, University of Southampton,
(e-mail: N.TZAVIDIS@soton.ac.uk)

ABSTRACT: In this paper we develop the unconditional M-quantile regression for modeling unconditional M-quantiles in the presence of covariates. Extending the paper by Firpo *et al.* (2009), we assess the impact of small changes in the explanatory variables on the M-quantile of the unconditional distribution of the dependent variable by running a mean regression of the recentered influence function of the unconditional M-quantile on the covariates. The proposed methodology is applied on the Survey of Household Income and Wealth (SHIW) 2016 conducted by the Bank of Italy.

KEYWORDS: Influence function, M-estimation, RIF regression, Robust method

1 Introduction

Quantile Regression (QR), as proposed by Koenker & Bassett Jr (1978), has proven to be a powerful tool to explore conditional distributions in many empirical applications. However, if one is interested in how the whole unconditional distribution of the dependent variable responds to changes in the covariates, using the well-known QR would yield misleading inferences (see Firpo *et al.* 2009 and Borah & Basu 2013). Motivated by this interest, Firpo *et al.* (2009) proposed the Unconditional Quantile Regression (UQR) approach for modeling unconditional quantiles of a dependent variable as a function of the explanatory variables. This method builds upon the concept of Recentered Influence Function (RIF) which originates from a widely used tool in robust statistics, namely the Influence Function (IF) discussed in Hampel *et al.* (2011). The RIF of a distributional statistic ν is obtained by adding back the statistic to the IF and it can be thought of as the contribution of an individual observation on ν. In the regression framework where covariates are available, Firpo *et al.* (2009) proposed to replace the dependent variable with the RIF to model the unconditional quantiles of the response and evaluate the effect of changes in the law of the covariates on unconditional quantiles. When the interest of the research is concentrated on the entire distribution of a response variable, in addition to the classical QR, a possible alternative is represented by the M-quantile regression (MQR) approach proposed by Breckling & Chambers (1988). This method provides a "quantile-like" generalization of the mean regression based on influence functions, combining in a common framework the robustness and efficiency properties of quantiles and expectiles (Newey & Powell 1987), respectively.

In this article, we extend the UQR of Firpo *et al.* (2009) to the M-quantile regression framework. We develop the Unconditional M-quantile Regression (UMQR) to model the M-quantiles of the unconditional distribution of the response variable. In order to analyze how the entire unconditional distribution of the outcome is affected by changes in the distribution of explanatory variables, we regress the RIF of the unconditional M-quantile on the covariates and denote such effect as Unconditional M-Quantile Partial Effect (UMQPE).

2 Methodology

Let Y denote a scalar random variable with absolutely continuous distribution function F_Y. The M-quantile of order $\tau \in (0,1)$ of Y is defined as the solution, $\theta_\tau \in \mathbb{R}$, of the following estimating equation:

$$\int \psi_\tau(y-\theta_\tau)dF_Y(y) = 0, \tag{1}$$

where $\psi_\tau(u) = \mid \tau - \mathbf{1}_{(u<0)} \mid \psi(u/\sigma_\tau)$, with ψ being the first derivative of a convex loss function ρ and σ_τ is a suitable scale parameter. In this work, we consider the well-known Huber influence function (Huber (1964)):

$$\psi(u) = u\mathbf{1}_{(|u|\leq c)} + c\,\mathrm{sign}(u)\mathbf{1}_{(|u|>c)}, \tag{2}$$

where c denotes a tuning constant bounded away from zero that can be used to trade robustness for efficiency in the model fit. In particular, M-quantiles nicely include quantiles when $c \to 0$, $\psi(u) = \mathrm{sign}(u)$, and expectiles when $c \to \infty$, $\psi(u) = u$.

To build the UMQR model, it follows from Firpo *et al.* (2009) and Hampel *et al.* (2011) that the RIF of the M-quantile θ_τ is defined as:

$$RIF(y;\theta_\tau) = \theta_\tau + IF(y;\theta_\tau) = \theta_\tau + \frac{\psi_\tau(y-\theta_\tau)}{\int \psi'_\tau(y-\theta_\tau)dF_Y(y)}, \tag{3}$$

where $IF(y;\theta_\tau)$ is the IF of θ_τ and $\psi'(u) = \mathbf{1}_{(|u|<c)}$ is the derivative of ψ in (2). In a regression framework when covariates $\mathbf{X} \subset \mathbb{R}^k$ are available, from (3) we define the UMQR model as follows:

$$\mathbb{E}[RIF(Y;\theta_\tau) \mid \mathbf{X} = \mathbf{x}] = \theta_\tau + \mathbb{E}\Big[\frac{\psi_\tau(y-\theta_\tau)}{\int \psi'_\tau(y-\theta_\tau)dF_Y(y)}\Big|\mathbf{X} = \mathbf{x}\Big]. \tag{4}$$

Our objective is to identify how small changes in the distribution of $\mathbf{X}$ affect the M-quantile of the unconditional distribution of Y. From (4) and Firpo *et al.* (2009), the unconditional effect of the τ-th M-quantile, that we denote Unconditional M-quantile Partial Effect, α_τ, is formally defined as:

$$\alpha_\tau = \int \frac{d\mathbb{E}[RIF(Y;\theta_\tau) \mid \mathbf{X} = \mathbf{x}]}{d\mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}) = \frac{1}{s_\tau} \int \frac{d\mathbb{E}[\psi_\tau(Y-\theta_\tau) \mid \mathbf{X} = \mathbf{x}]}{d\mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}), \tag{5}$$

where $F_{\mathbf{X}}$ is the distribution function of $\mathbf{X}$ and $s_\tau = \int \psi'_\tau(y-\theta_\tau)dF_Y(y)$. As suggested by Firpo *et al.* (2009), we can estimate α_τ in (5) via a mean regression of the $RIF(Y;\theta_\tau)$ as dependent variable onto $\mathbf{X}$ by using a two-step procedure. Specifically, an estimate $\widehat{\theta}_\tau$ of θ_τ is obtained by solving (1) via Iterative Reweighted Least Squares, substitute $\widehat{\theta}_\tau$ in (3) and then regress the $RIF(Y;\widehat{\theta}_\tau)$ on $\mathbf{X}$.

3 Application

We investigate the effect of economic and socio-demographic characteristics on italian households' log-consumption using data from the SHIW 2016. We fit the UMQR at different points of the unconditional distribution of the response and compare the results with standard conditional M-quantile regressions. The tuning constant c in (2) has been set to 1.345 and 100. In the second case, we obtain the Unconditional Expectile Regression (UER). The results in Table 1 highlight that the impact of income, gender, age and education is very different on the conditional and unconditional distributions of consumption, especially in the tails. This demonstrates the ability of the UMQR to extend mean regression for estimating the effect of covariates, not only at the center, but also at different parts of the unconditional distribution of interest.

References

BORAH, BIJAN J, & BASU, ANIRBAN. 2013. Highlighting differences between conditional and unconditional quantile regression approaches through an

Variable	MQR			UMQR			ER			UER		
τ	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
Log-Income	**0.570**	**0.595**	**0.442**	**0.447**	**0.391**	**0.429**	**0.483**	**0.413**	**0.263**	**0.450**	**0.413**	**0.436**
	(0.011)	(0.007)	(0.010)	(0.038)	(0.032)	(0.038)	(0.011)	(0.008)	(0.011)	(0.038)	(0.033)	(0.038)
Gender	−0.019	−0.011	**−0.043**	−0.011	**−0.024**	**−0.038**	−0.023	**−0.026**	**−0.046**	−0.010	**−0.026**	−0.035
	(0.016)	(0.009)	(0.014)	(0.018)	(0.012)	(0.018)	(0.016)	(0.011)	(0.016)	(0.017)	(0.012)	(0.018)
Age	−0.002	0.001	0.004	**−0.013**	**0.006**	**0.013**	−0.001	0.004	**0.008**	**−0.011**	0.004	**0.011**
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Marital status												
never married	**−0.062**	**−0.084**	**−0.164**	**−0.094**	**−0.141**	**−0.187**	**−0.095**	**−0.138**	**−0.201**	**−0.101**	**−0.138**	**−0.176**
	(0.020)	(0.012)	(0.018)	(0.025)	(0.017)	(0.022)	(0.020)	(0.014)	(0.020)	(0.024)	(0.017)	(0.022)
separated	**−0.066**	**−0.056**	**−0.127**	**−0.102**	**−0.151**	**−0.155**	**−0.111**	**−0.137**	**−0.207**	**−0.105**	**−0.137**	**−0.141**
	(0.025)	(0.015)	(0.022)	(0.034)	(0.024)	(0.030)	(0.025)	(0.017)	(0.026)	(0.033)	(0.024)	(0.030)
widowed	−0.040	**−0.063**	**−0.119**	**−0.116**	**−0.136**	**−0.111**	**−0.074**	**−0.123**	**−0.193**	**−0.110**	**−0.123**	**−0.107**
	(0.022)	(0.013)	(0.020)	(0.029)	(0.019)	(0.025)	(0.022)	(0.015)	(0.022)	(0.028)	(0.019)	(0.025)
Education level												
elementary school	**0.175**	**0.120**	**0.151**	**0.488**	**0.125**	−0.037	**0.188**	**0.161**	**0.187**	**0.446**	**0.161**	−0.000
	(0.039)	(0.023)	(0.035)	(0.069)	(0.024)	(0.022)	(0.039)	(0.027)	(0.040)	(0.066)	(0.027)	(0.022)
middle school	**0.240**	**0.203**	**0.316**	**0.645**	**0.269**	**0.060**	**0.281**	**0.294**	**0.398**	**0.590**	**0.294**	**0.094**
	(0.041)	(0.024)	(0.037)	(0.070)	(0.028)	(0.029)	(0.041)	(0.028)	(0.042)	(0.067)	(0.030)	(0.028)
high school	**0.248**	**0.235**	**0.383**	**0.652**	**0.355**	**0.147**	**0.313**	**0.363**	**0.500**	**0.598**	**0.363**	**0.168**
	(0.042)	(0.025)	(0.038)	(0.072)	(0.033)	(0.037)	(0.042)	(0.029)	(0.043)	(0.069)	(0.034)	(0.036)
university	**0.298**	**0.297**	**0.521**	**0.631**	**0.440**	**0.506**	**0.391**	**0.484**	**0.705**	**0.608**	**0.484**	**0.515**
	(0.045)	(0.027)	(0.040)	(0.076)	(0.040)	(0.053)	(0.045)	(0.031)	(0.046)	(0.073)	(0.042)	(0.052)
Employment status												
self-employed	**−0.087**	0.010	**0.083**	**−0.060**	0.021	**0.121**	**−0.058**	0.023	**0.081**	**−0.046**	0.023	**0.107**
	(0.024)	(0.014)	(0.022)	(0.021)	(0.019)	(0.038)	(0.024)	(0.017)	(0.025)	(0.020)	(0.018)	(0.037)
not-employed	0.008	**0.027**	0.035	−0.046	**0.037**	0.037	−0.002	0.014	0.017	**−0.052**	0.014	0.031
	(0.021)	(0.013)	(0.019)	(0.025)	(0.016)	(0.025)	(0.021)	(0.015)	(0.022)	(0.024)	(0.015)	(0.024)

Table 1. *M-quantile and Expectile regression results at* $\tau = (0.1, 0.5, 0.9)$. *Parameter estimates are displayed in boldface when significant at the 5% level.*

application to assess medication adherence. *Health Economics*, **22**(9), 1052–1070.

BRECKLING, JENS, & CHAMBERS, RAY. 1988. M-quantiles. *Biometrika*, **75**(4), 761–771.

FIRPO, SERGIO, FORTIN, NICOLE M, & LEMIEUX, THOMAS. 2009. Unconditional quantile regressions. *Econometrica: Journal of the Econometric Society*, **77**(3), 953–973.

HAMPEL, FRANK R, RONCHETTI, ELVEZIO M, ROUSSEEUW, PETER J, & STAHEL, WERNER A. 2011. *Robust statistics: the approach based on influence functions*. Vol. 196. John Wiley & Sons.

HUBER, PETER J. 1964. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, **35**(1), 73–101.

KOENKER, ROGER, & BASSETT JR, GILBERT. 1978. Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

NEWEY, WHITNEY K, & POWELL, JAMES L. 1987. Asymmetric least squares estimation and testing. *Econometrica*, 819–847.