MID-QUANTILE REGRESSION FOR DISCRETE PANEL DATA

Alessio Farcomeni[1], Alfonso Russo[1] and Marco Geraci[2]

[1] Department of Economics and Finance, University of Rome "Tor Vergata", (e-mail: alessio.farcomeni@uniroma2.it, alfonso.russo@uniroma2.it)

[2] Department of Methods and Models for Economics, Territory and Finance, Sapienza - University of Rome (e-mail: marco.geraci@uniroma1.it)

ABSTRACT: We propose a novel method for quantile regression for discrete longitudinal data. The approach is based on the notion of conditional mid-quantiles, which have good theoretical properties even in the presence of ties, and a Ridge-type penalised framework to accommodate dependent data. We illustrate the methods with a simulation study and an original application to the use of macroprudential policies in more than one hundred countries over a period of fifteen years.

KEYWORDS: mid-quantile regression, macroprudential policies, fixed effects, random effects, longitudinal data

1 Introduction

Quantile Regression (QR) involves modeling effects of predictors at specific quantiles of an endogenous variable. Most QR methodologies are restricted to continuous outcomes, with some notable exceptions (Machado & Santos Silva, 2005; Frumento & Salvati, 2021). Recently, Geraci & Farcomeni, 2022 proposed a method based on conditional mid-quantiles (see also Ma *et al.*, 2011). We extend their approach to the case of discrete panel data. Our approach can also be seen as an extension to discrete outcomes of the penalised framework for QR for continuous panel data (Koenker, 2004). We develop a collect of methods that are based on a two-step algorithm. At the first step, the conditional mid-quantile function is estimated through a semiparametric approach; at the second step we optimise a possibly penalised objective function to obtain parameter estimates. We illustrate the methods by means of a simulation study, and an original application to macroprudential policies in a panel of countries.

2 Penalized mid-quantile regression

Let y_{it}, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n$, denote a discrete/ordered outcome and $\tilde{x}_{it}$ an associated vector of covariates. Measurements are repeatedly taken $T_i \geq 1$ times for each unit, with $T_i > 1$ at least for one subject. We define the conditional mid-CDF of Y as $G_{Y|X}(y|x) = F_{Y|X}(y|x) - 0.5 \cdot \mathrm{m}_{Y|X}(y|x)$ and $G^C_{Y|X}(y|x)$ the continuous function that interpolates $G_{Y|X}(y|x)$, where $F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x)$ and $\mathrm{m}_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x)$. Let $p \in (0,1)$. The conditional mid-quantile function is the generalised inverse $H_{Y|X}(p) = G^{-1}_{Y|X}(y|x)$.

We assume a p-specific model that is linear on the scale of a link function $h(\cdot)$:

$$h\{\eta_{it}(p)\} = \alpha_i(p) + \tilde{x}_{it}^T \beta(p) = H_{h(Y)|X}(p) \tag{1}$$

Estimation, as in Geraci & Farcomeni, 2022, proceeds in two steps. In the first step, one obtains estimates of the conditional mid-CDF. Similarly to Peracchi, 2002, we define outcome variables $1\{y_{it} \leq c\}$ at appropriate cut-points. We then estimate logistic regression models with either (1) fixed subject-specific, (2) random subject-specific, or (3) homogeneous intercepts. For a fixed penalty $\lambda > 0$, our objective function for the second step is given by

$$\psi_n[\theta(p); p] = \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ p - \hat{G}^c_{Y|X}(\eta_{it}|\tilde{x}_{it}) \right\}^2 + \lambda \sum_{i=1}^{n} \alpha_i^2. \tag{2}$$

The optimum is available in closed form as a Ridge-type estimator. For selection of the penalty parameter we use an heuristic reasoning as in Ruppert *et al.*, 2003. As long as $\min_i T_i > 1$ it is possible also to set $\lambda = 0$; and it is also possible to set $\lambda \to \infty$, therefore obtaining homogenenous intercepts $\alpha_i = \alpha$. In summary we are proposing three possible routes for estimation of the conditional CDF and three possible routes for the second step. The case with homogeneous intercepts at the first step and $\lambda \to \infty$ recovers the methodology in Geraci & Farcomeni, 2022.

3 Simulation study

In Figure 1 we show mean squared error (500 replicates) for regression coefficient estimates for ten alternative model specifications, reported by quantiles $(0.2, 0.5, 0.8)$ and two sample sizes. The first nine model specifications involve our proposed class, where at the first step intercepts can be homogenous (*HMG*), treated as fixed (*FE*), or random (*RE*). Each specification is

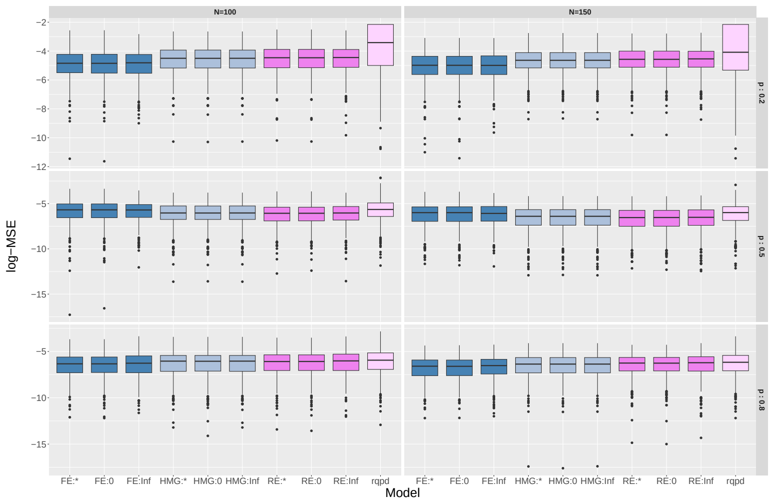

Figure 1. *Log Mean Squared Error of parameter estimates for a Poisson response with two continuous covariates.*

paired, at the second step, with three different choices for the penalty parameter $\lambda \to 0, \lambda \to \infty$, or $\lambda = \lambda^*$. Finally *rqpd* denotes the quantile regression procedure in Koenker, 2004. At the data generation stage we simulate Poisson responses with two Gaussian covariates. Several other settings are available in the accompanying paper. The general conclusions that can be drawn are that (i) our method outperforms *rqpd*, which does not take into account the discrete nature of the outcome, and (ii) the MSE decreases at the expected rate.

4 Real data example

Macroprudential policies (MP) (Galati & Moessner, 2013) are used by central banks to protect macroeconomic performance from the drawbacks of externalities, market failures, excessive procyclicality and other factors. They involve currency instruments, limits to bank exposure, and similar requirements. In this work our focus is on the determinants of the use of MP. Our endogenous variable is the number (up to twelve) of different MP used by a country in a given year. We collect data on a panel of $n = 115$ countries over $T = 18$ years starting from 2001. Predictors include World Bank label for the economy, debt to gdp ratio, unemployment rate, trade as % of GDP. All covariates are lagged by one year.

Results for optimal model specification selected through 10-fold cross val-

Table 1. *Macroprudential policy determinants in 115 countries from 2001 to 2017. Parameter estimates (95% CI in parenthesis) at different quantiles p.*

	$p = 0.2$	$p = 0.5$	$p = 0.8$	$p = 0.9$
Trade-to-GDP	$0.09(0.05, 0.12)$	$0.06(0.03, 0.09)$	$0.06(0.03, 0.09)$	$0.05(0.03, 0.08)$
Unempl.	$-0.03(-0.06, -0.01)$	$-0.03(-0.05, -0.00)$	$-0.02(-0.04, 0.00)$	$-0.02(-0.04, 0.00)$
Debt-to-GDP	$0.03(0.01, 0.05)$	$0.02(0.00, 0.04)$	$0.02(0.01, 0.04)$	$0.03(0.01, 0.04)$
High income	$0.31(0.24, 0.37)$	$0.38(0.32, 0.44)$	$0.31(0.26, 0.36)$	$0.30(0.25, 0.35)$
Up-Mid Income	$0.54(0.46, 0.61)$	$0.60(0.53, 0.66)$	$0.47(0.42, 0.53)$	$0.45(0.40, 0.50)$
Low-Mid Income	$0.29(0.23, 0.35)$	$0.34(0.29, 0.40)$	$0.28(0.23, 0.32)$	$0.26(0.22, 0.31)$
Time	$0.05(0.04, 0.05)$	$0.04(0.04, 0.05)$	$0.04(0.04, 0.04)$	$0.04(0.04, 0.04)$

idation are reported in Table 1. Consistently with the literature upper-middle income countries tend to use more MP. Effects are quantile-dependent, with high trade-to-GDP and debt-to-GDP prompting larger use at low quantiles.

References

FRUMENTO, P., & SALVATI, N. 2021. Parametric modeling of quantile regression coefficient functions with count data. *Statistical Methods & Applications*, **30**, 1237–1258.

GALATI, G., & MOESSNER, R. 2013. Macroprudential Policy - a literature review. *Journal of Economic Surveys*, **27**, 846–878.

GERACI, M, & FARCOMENI, A. 2022. Mid-quantile regression for discrete responses. *Statistical Methods in Medical Research*, **31**(5), 821–838.

KOENKER, R. 2004. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**(1), 74–89.

MA, Y., GENTON, M. G., & PARZEN, E. 2011. Asymptotic properties of sample quantiles of discrete distributions. *Annals of the Institute of Statistical Mathematics*, **63**(2), 227–243.

MACHADO, J. A. F., & SANTOS SILVA, J. M. C. 2005. Quantiles for counts. *Journal of the American Statistical Association*, **100**(472), 1226–1237.

PERACCHI, F. 2002. On estimating conditional quantiles and distribution functions. *Computational Statistics & Data Analysis*, **38**(4), 433–447.

RUPPERT, D., WAND, M. P., & CARROLL, R. J. 2003. *Semiparametric regression*. Cambridge university press.