Finite Mixtures of Quantile and M-quantile regression models

Marco Alfò[1] Nicola Salvati[2] M.G. Ranalli[3]

[1]Sapienza Università di Roma [2]Università di Pisa [3]Università di Perugia

Workshop on "Recent Advances in Quantile and M-quantile Regression"
Università di Pisa — July 15th, 2016

Essential References


Alfó, M., Salvati, N., Ranalli M.G. (2016)
Finite mixtures of quantile and M-quantile regression models.
Statistics and Computing


Tzavidis, N., Salvati, N., Schmid, T., Flouri, E., Midouhas, E. (2016)
Longitudinal analysis of the strengths and difficulties questionnaire scores of the Millennium Cohort Study children in England using M-quantile random-effects regression,
Journal of the Royal Statistical Society: Series A

The presentation at a glance

Data are seldom i.i.d. and without outliers!

- Dependent Observations (multilevel, longitudinal, panel data)
- Quantile and M-quantile regression models
- Introducing Finite Mixtures (nonparametric distribution for the random effects)
- Maximum Likelihood Estimation
- Multivariate extension

Outline

1. Introduction on Finite Mixtures
   - Dependent Observations
   - Finite mixtures of regression models
2. Finite Mixtures for Quantile and M-Quantile regression models
   - Likelihood Inference (focus on MQ)
3. Applications
   - Pain Labor Data & Treatment of lead-exposed children
   - The Millennium Cohort Study (Joint work with MF Marino & N Tzavidis)
4. Conclusions

Hierarchically structured data

Regression model for multilevel data

$$E(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_i) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{b}_i, \quad i = 1, \ldots, n, \; j = 1, \ldots, r_i$$

- y_{ij}, observed response variable
- $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ vector of explanatory variables; let $x_{ij1} \equiv 1$
- Linear Models (for ease of notation) $\rightarrow$ GLMs

Hierarchically structured data

Regression model for multilevel data

$$E(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_i) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{b}_i, \quad i = 1, \ldots, n, \; j = 1, \ldots, r_i$$

- $\boldsymbol{w}_{ij}$ is a subset of $\boldsymbol{x}_{ij}$ that contains those $p_1 \leqslant p$ variables whose effects are assumed to be individual-specific
- the effects $\boldsymbol{b}_i$ $i = 1, \ldots, n$, vary across individuals according to a distribution $h(\cdot)$

Likelihood

Local independence assumption

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} \left\{ \int_{\mathcal{B}} \prod_{j=1}^{r_i} f(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{b}_i) \mathrm{d}H(\boldsymbol{b}_i) \right\},$$

- $\mathbf{\Phi}$ global set of parameters,
- $f(\cdot)$ is the Gaussian density,
- $H(\cdot)$ is the random coefficient cdf and $\mathcal{B}$ the corresponding support
- In the general case, the integral defining the likelihood can not be analytically computed (GQ, aGQ, MCML, Composite Lik, etc.)

Nonparametric distribution for the random coefficients

- Leave $h(\cdot)$ unspecified
- Approximate $h(\cdot)$ by a discrete distribution on $G < n$ locations $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_G\}$, with associated probabilities defined by $\pi_k = \Pr(\boldsymbol{b}_i = \boldsymbol{b}_k)$, $i = 1, \ldots, n$ and $k = 1, \ldots, G$.

$$\boldsymbol{b}_i \sim \sum_{k=1}^{G} \pi_k \delta_{\boldsymbol{b}_k}$$

where δ_θ is a one-point distribution putting a unit mass at θ.

Comparing the Likelihoods

Nonparametric distribution for the random effects

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{G} \prod_{j} f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_k)\pi_k \right\} =: \prod_{i=1}^{n} \left\{ \sum_{k=1}^{G} \prod_{j} f_{ijk}\pi_k \right\}.$$

Parametric distribution for the random effects

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} \left\{ \int_{\mathcal{B}} \prod_{j=1} f(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{b}_i)\mathrm{d}H(\boldsymbol{b}_i) \right\},$$

Comparing the Likelihoods

Nonparametric distribution for the random effects

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{G} \prod_{j} f(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{b}_k)\pi_k \right\} =: \prod_{i=1}^{n} \left\{ \sum_{k=1}^{G} \prod_{j} f_{ijk}\pi_k \right\}.$$

- $\mathbf{\Phi} = \{\boldsymbol{\beta}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_G, \pi_1, \ldots, \pi_G\}$
- f_{ijk} is the distribution of the response variable for the j-th measurement in the i-th cluster when the k-th component of the finite mixture, $k = 1, \ldots, G$ is considered
- resembles the likelihood function for a finite mixture of Gaussian distributions

Regression model

- semi-parametric approximation to a fully parametric, possibly continuous, distribution for the random coefficients
- a model-based clustering approach, where the population of interest is assumed to be divided in G homogeneous sub-populations which differ for the values of the regression parameters

Considering the k-th component of the mixture,

$$E(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_k) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{b}_k.$$

Estimation of model parameters (1)

The score function can be written as the posterior expectation of the score function corresponding to a standard LM:

$$\mathcal{S}(\mathbf{\Phi}) = \frac{\partial \log[L(\mathbf{\Phi})]}{\partial \mathbf{\Phi}} = \sum_{i=1}^{n} \sum_{k=1}^{G} \tau_{ik} \sum_{j} \frac{\partial \log f_{ijk}}{\partial \mathbf{\Phi}},$$

where the weights

$$\tau_{ik} = \frac{\prod_j f_{ijk} \pi_k}{\sum_l \prod_j f_{ijl} \pi_l}$$

represent the posterior probabilities of component membership.

Estimation of model parameters (2)

- Likelihood equations that are essentially weighted sums of the likelihood equations for a standard LM, with weights τ_{ik}.
- The basic EM algorithm is defined by solving equations for a given set of the weights, and updating the weights as a function of the current parameter estimates.

Outline

Quantile and M-Quantile regression models for dependent observations

Linear Quantile Random Effect models (Geraci & Bottai, 2007, 2014; Liu & Bottai, 2009)

$$Q_q(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_{i,q}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_q + \boldsymbol{w}'_{ij}\boldsymbol{b}_{i,q}$$

Linear M-Quantile Random Effect models (Tzavidis et al., 2016)

$$MQ_q(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_{i,q}, \psi, c) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_q + \boldsymbol{w}'_{ij}\boldsymbol{b}_{i,q}$$

- Note that both fixed and random coefficients vary with $q \in (0, 1)$
- Random effects are normally distributed

Finite Mixtures of Q and MQ regression models

Approximate the distribution of the random coefficients through a discrete distribution defined on a finite, G-dimensional, set of locations. Then, conditional on k,

$$Q_q(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_{k,q}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_q + \boldsymbol{w}'_{ij}\boldsymbol{b}_{k,q}$$

$$MQ_q(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{b}_{k,q}, \psi, c) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_q + \boldsymbol{w}'_{ij}\boldsymbol{b}_{k,q}$$

for $k = 1, \ldots, G$.

- Each component of the mixture is characterised by a different (sub-) vector of regression coefficients, $\boldsymbol{b}_{k,q}$, $k = 1, \ldots, G$
- Note that the distribution of $\boldsymbol{b}_{k,q}$ may vary with quantiles

Estimation of model parameters (focus on MQ)

$$L(\boldsymbol{\Phi}_q) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{G} \prod_{j} f_q(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{b}_{k,q}) \pi_{k,q} \right\}.$$

- $\boldsymbol{\Phi}_q = \left\{ \boldsymbol{\beta}_q, \boldsymbol{b}_{1,q}, \ldots, \boldsymbol{b}_{G,q}, \sigma_q, \pi_{1,q}, \ldots, \pi_{G,q} \right\}$
- $f_q(\cdot)$ is the ALID (Asymmetric Least Informative Density, Bianchi et al., 2015):

$$f_q(\cdot) = \frac{1}{B_q(\sigma_q, c)} \exp\{-\rho_q(\cdot)\}$$

  - $B_q(\sigma_q, c)$ is a normalising constant that ensures the density integrates to one
  - $\rho_q(\cdot)$ is the Huber loss function.

Missing data approach

$$z_{ik,q} = \begin{cases} 1 & \text{if unit } i \text{ is in component } k \text{ of the mixture} \\ 0 & \text{otherwise} \end{cases}$$

- $P(z_{ik,q} = 1) = \pi_{k,q} = P(\boldsymbol{b}_{i,q} = \boldsymbol{b}_{k,q})$
- $\boldsymbol{z}_{i,q} = (z_{i1,q}, ..., z_{iG,q})'$, $i = 1, ..., n$, are considered as missing data

Complete data log-likelihood

Should we have observed, for each i, $(\boldsymbol{y}_i, \boldsymbol{z}_{i,q})$, the log-likelihood for the *complete* data would have been:

$$\ell_c(\boldsymbol{\Phi}_q) = \sum_{i=1}^{n} \sum_{k=1}^{G} z_{ik,q} \left\{ \log \left[f_q(\boldsymbol{y}_i \mid \boldsymbol{\beta}_q, \boldsymbol{b}_{k,q}, \sigma_q) \right] + \log(\pi_{k,q}) \right\}$$

Missing data approach

$$z_{ik,q} = \begin{cases} 1 & \text{if unit } i \text{ is in component } k \text{ of the mixture} \\ 0 & \text{otherwise} \end{cases}$$

- $P(z_{ik,q} = 1) = \pi_{k,q} = P(\boldsymbol{b}_{i,q} = \boldsymbol{b}_{k,q})$
- $\boldsymbol{z}_{i,q} = (z_{i1,q}, ..., z_{iG,q})'$, $i = 1, ..., n$, are considered as missing data

Complete data log-likelihood

Should we have observed, for each i, $(\boldsymbol{y}_i, \boldsymbol{z}_{i,q})$, the log-likelihood for the *complete* data would have been:

$$\ell_c(\boldsymbol{\Phi}_q) = \sum_{i=1}^{n} \sum_{k=1}^{G} z_{ik,q} \left\{ \log \left[f_q(\boldsymbol{y}_i \mid \boldsymbol{\beta}_q, \boldsymbol{b}_{k,q}, \sigma_q) \right] + \log(\pi_{k,q}) \right\}$$

Maximum Likelihood via the EM algorithm – E-step

Expected value of $\ell_c(\boldsymbol{\Phi}_q)$ over $\boldsymbol{z}_{i,q}$, conditional on the observed data and the current parameter estimates:

$$\begin{aligned} Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q^{(t)}) &= \mathrm{E}_{\widehat{\boldsymbol{\Phi}}_q^{(t)}} \left[\ell_c(\boldsymbol{\Phi}_q) \mid \boldsymbol{y}_i \right] \\ &= \sum_{i=1}^{n} \sum_{k=1}^{G} \tau_{ik,q}^{(t+1)} \left\{ \log \left[f_q(\boldsymbol{y}_i \mid \boldsymbol{\beta}_q, \boldsymbol{b}_{k,q}, \sigma_q) \right] + \log(\pi_{k,q}) \right\}. \end{aligned}$$

That is, the unobservable indicators are replaced by their conditional expectation, which, at iteration $(t+1)$ are given by

$$\tau_{ik,q}^{(t+1)} = \frac{\pi_{k,q}^{(t)} f_{ik,q}(\widehat{\boldsymbol{\Phi}}_q^{(t)})}{\sum_l \pi_{l,q}^{(t)} f_{il,q}(\widehat{\boldsymbol{\Phi}}_q^{(t)})}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, G.$$

Maximum Likelihood via the EM algorithm – E-step

Expected value of $\ell_c(\boldsymbol{\Phi}_q)$ over $\boldsymbol{z}_{i,q}$, conditional on the observed data and the current parameter estimates:

$$\begin{aligned} Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q^{(t)}) &= \mathrm{E}_{\widehat{\boldsymbol{\Phi}}_q^{(t)}} \left[\ell_c(\boldsymbol{\Phi}_q) \mid \boldsymbol{y}_i\right] \\ &= \sum_{i=1}^{n}\sum_{k=1}^{G} \tau_{ik,q}^{(t+1)} \left\{\log\left[f_q(\boldsymbol{y}_i \mid \boldsymbol{\beta}_q, \boldsymbol{b}_{k,q}, \sigma_q)\right] + \log(\pi_{k,q})\right\}. \end{aligned}$$

That is, the unobservable indicators are replaced by their conditional expectation, which, at iteration $(t+1)$ are given by

$$\tau_{ik,q}^{(t+1)} = \frac{\pi_{k,q}^{(t)} f_{ik,q}(\widehat{\boldsymbol{\Phi}}_q^{(t)})}{\sum_l \pi_{l,q}^{(t)} f_{il,q}(\widehat{\boldsymbol{\Phi}}_q^{(t)})}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, G.$$

Maximum Likelihood via the EM algorithm – M-step

Maximise the function $Q(\cdot)$ w.r.t. $\boldsymbol{\Phi}_q$ to update parameter estimates.
Then $\widehat{\boldsymbol{\Phi}}_q^{(t+1)}$ are defined to be the solutions to the following score equation:

$$\frac{\partial Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q^{(t)})}{\partial \boldsymbol{\Phi}_q} = \mathbf{0},$$

which are equivalent to the score equations for the observed data, $\mathcal{S}(\boldsymbol{\Phi}_q) = \mathbf{0}$.

Standard Errors

Oakes (1999)'s identity

$$\begin{aligned} \boldsymbol{I}(\widehat{\boldsymbol{\Phi}}_q) &= -\left\{ \left. \frac{\partial^2 Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q)}{\partial \boldsymbol{\Phi}_q \partial \boldsymbol{\Phi}_q'} \right|_{\widehat{\boldsymbol{\Phi}}_q = \boldsymbol{\Phi}_q} + \left. \frac{\partial^2 Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q)}{\partial \widehat{\boldsymbol{\Phi}}_q \partial \widehat{\boldsymbol{\Phi}}_q'} \right|_{\widehat{\boldsymbol{\Phi}}_q = \boldsymbol{\Phi}_q} \right\} \\ &= \text{A} + \text{B} \end{aligned}$$

A Cond. exp. of the complete data Hessian given the obs. data (EM)

B First derivative of the cond. exp. of the complete data Score given the obs. data (`numDeriv` in `R`)

Sandwich $\widehat{\text{Cov}}\left(\widehat{\boldsymbol{\Phi}}_q\right) = \boldsymbol{I}(\widehat{\boldsymbol{\Phi}}_q)^{-1} \boldsymbol{V}(\widehat{\boldsymbol{\Phi}}_q) \boldsymbol{I}(\widehat{\boldsymbol{\Phi}}_q)^{-1}$, where $\boldsymbol{V}(\widehat{\boldsymbol{\Phi}}_q) = \sum_{i=1}^{n} \mathcal{S}_i(\widehat{\boldsymbol{\Phi}}_q) \mathcal{S}_i(\widehat{\boldsymbol{\Phi}}_q)'$.

Standard Errors

Oakes (1999)'s identity

$$\begin{aligned} \boldsymbol{I}(\widehat{\boldsymbol{\Phi}}_q) &= -\left\{ \left. \frac{\partial^2 Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q)}{\partial \boldsymbol{\Phi}_q \partial \boldsymbol{\Phi}_q'} \right|_{\widehat{\boldsymbol{\Phi}}_q = \boldsymbol{\Phi}_q} + \left. \frac{\partial^2 Q(\boldsymbol{\Phi}_q \mid \widehat{\boldsymbol{\Phi}}_q)}{\partial \widehat{\boldsymbol{\Phi}}_q \partial \widehat{\boldsymbol{\Phi}}_q'} \right|_{\widehat{\boldsymbol{\Phi}}_q = \boldsymbol{\Phi}_q} \right\} \\ &= \quad \text{A} \quad + \quad \text{B} \end{aligned}$$

A Cond. exp. of the complete data Hessian given the obs. data (EM)

B First derivative of the cond. exp. of the complete data Score given the obs. data (`numDeriv` in `R`)

Sandwich $\widehat{\mathrm{Cov}}\left(\widehat{\boldsymbol{\Phi}}_q\right) = \boldsymbol{I}(\widehat{\boldsymbol{\Phi}}_q)^{-1} \boldsymbol{V}(\widehat{\boldsymbol{\Phi}}_q) \boldsymbol{I}(\widehat{\boldsymbol{\Phi}}_q)^{-1}$, where $\boldsymbol{V}(\widehat{\boldsymbol{\Phi}}_q) = \sum_{i=1}^{n} \mathcal{S}_i(\widehat{\boldsymbol{\Phi}}_q) \mathcal{S}_i(\widehat{\boldsymbol{\Phi}}_q)'$.

Applications

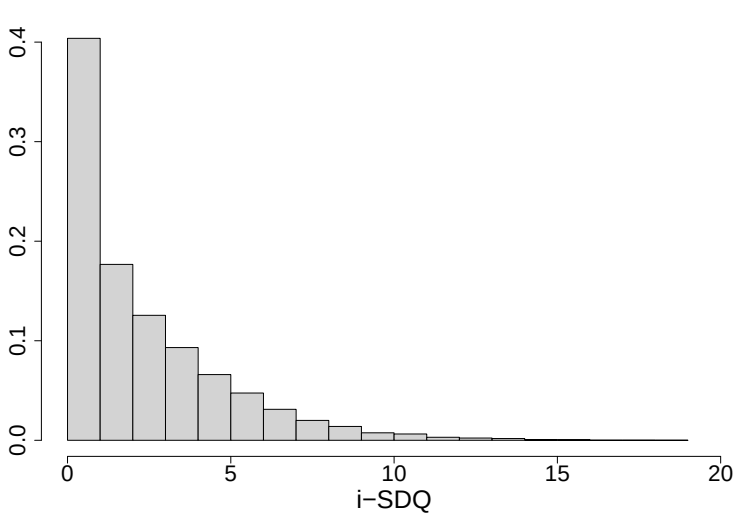
- Univariate response (Alfó, Salvati, Ranalli, Stat. Comp., 2016)
  - Pain Labor Data
  - Treatment of lead-exposed children
- Multivariate response (Joint work with M.F. Marino & N. Tzavidis)
  - The Millennium Cohort Study

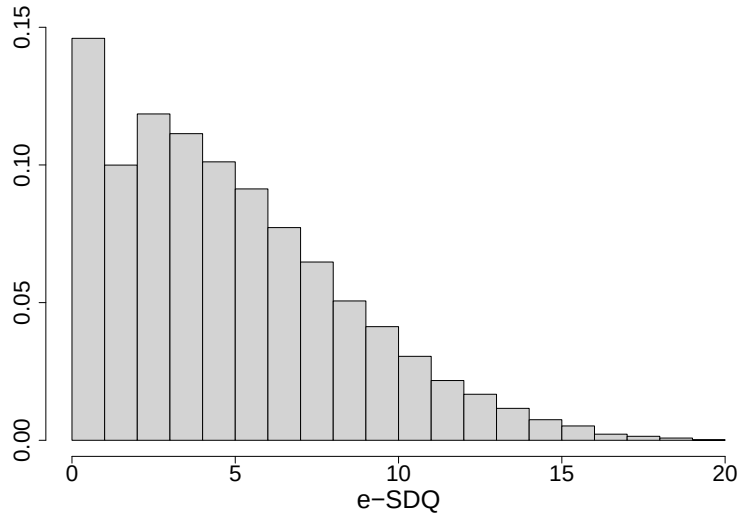
The Millennium Cohort Study

- Longitudinal study on children's emotional/behavioural problems measured via the Strengths and Difficulties Questionnaire (SDQ)
- $n = 9021$ children born in the UK between Sept. 2000 and Sept 2001
- First information collected when children were around 9 months old. Waves 2, 3, 4 took place around ages 2, 5, and 7

Outcome variables

- internalizing SDQ - i-SDQ (emotional problems): total score on 5 emotional symptom items + 5 peer problem items $(0-20)$
- externalising SDQ - e-SDQ (behavioural problems): total score on 5 conduct problem items + 5 hyperactivity items $(0-20)$





Multivariate Extension

- y_{ijh}, $h = 1, 2$ observed outcomes
- The joint conditional distribution from unit i is

$$f_q(\boldsymbol{y}_i \mid \boldsymbol{\beta}_q, \boldsymbol{b}_{i,q}, \boldsymbol{\sigma}_q) = \prod_{h=1}^{H} \prod_{j} f_q\left(y_{ijh} \mid \boldsymbol{\beta}_{h,q}, \boldsymbol{b}_{ih,q}, \sigma_{h,q}\right).$$

- Conditional independence assumption

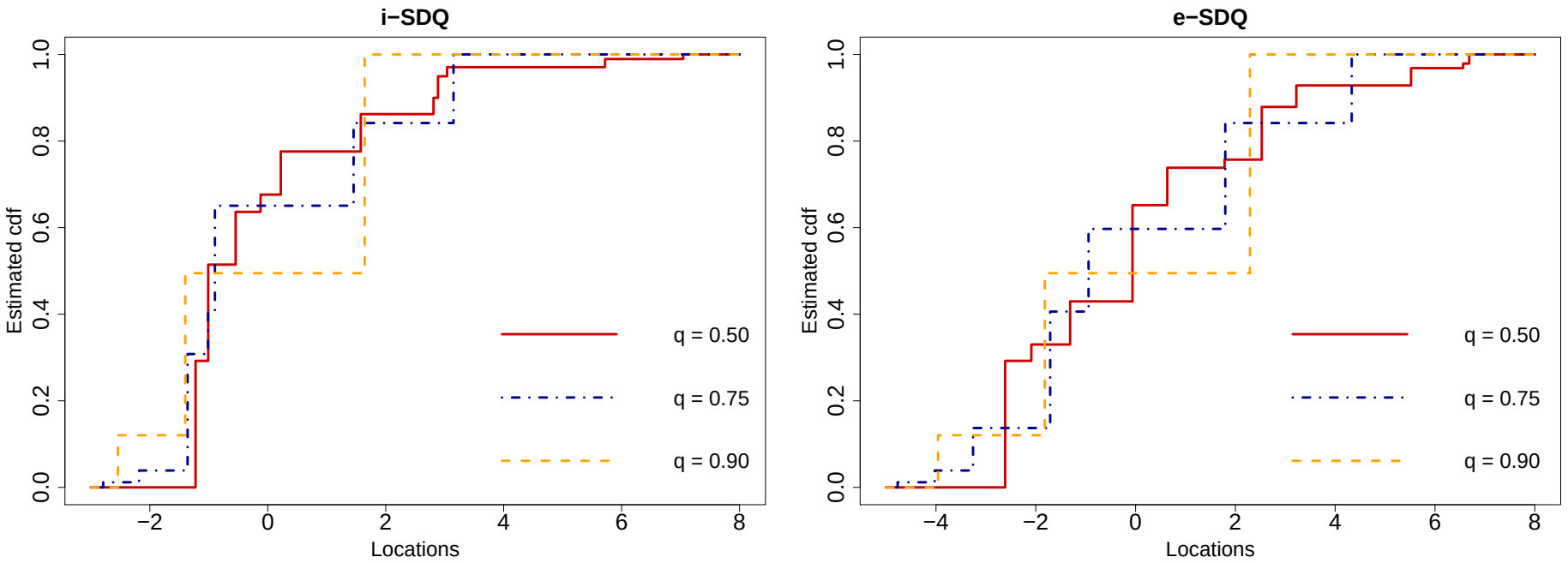
Covariates

- ALE_{11} : number of potentially Adverse Life Events $(0-11)$
- SED_4 : family poverty score measured on the SED scale $(0-4)$
- KESSM: maternal depression score measured on the Kessler scale $(0-24)$
- IMD: neighborhood deprivation rank measured by the Index of Multiple Deprivation with lower values corresponding to higher deprivation $(1-10)$
- Age: child's age
- Maternal education: no qualification (bsl.), degree, GCSE
- Ethnicity : non-white (bsl.), white
- Gender: female (bsl.), male
- Statification: advantaged (bsl.), ethnic, disadvantaged

Modeling details

- Focus on more severe emotional and behavioural problems, i.e. $q = \{0.50, 0.75, 0.90\}$
- Discrete random intercepts to account for dependence
- Age is centered around the mean and a squared effect is also considered
- ALE_{11}, SED_4, KESSM, and IMD are centered around their individual means to account for between/within individual effects
- BIC is used to select the optimal model ($G = 1, \ldots, 15$)

Discrete distributions of random effects



- Higher dispersion for e-SDQ intercepts
- The probability of higher components increases with q
- Random intercept distribution is quite far from symmetry and unimodality

Model for the M-median

	i-SDQ		e-SDQ	
	Est	se	Est	se
Age	-0.02	0.04	**-0.45**	0.05
Age^2	**0.07**	0.01	**0.21**	0.02
ALE_{11} mean	0.09	0.23	**0.19**	0.04
ALE_{11}	**0.06**	0.02	0.09	0.06
SED_4 mean	**0.12**	0.05	0.17	0.14
SED_4	-0.04	0.06	-0.01	0.07
Kessm mean	**0.17**	0.08	**0.23**	0.09
Kessm	**0.08**	0.01	**0.11**	0.02
Degree	-0.66	0.74	-1.17	0.44
Gcse	-0.41	0.34	-0.50	0.27
White	**-0.31**	0.11	0.17	0.16
Male	0.05	0.12	**0.75**	0.16
IMD mean	-0.02	0.04	-0.04	0.04
IMD	-0.00	0.03	-0.03	0.04
Ethnic st.	0.18	0.10	-0.05	0.22
Disadv st.	0.07	0.39	0.11	0.32
σ_u	1.72		2.52	

- Both i-SDQ and e-SDQ reduce as the time passes by until children are 5 years old and start increase afterwards
- Adverse life events (ALE_{11}) and maternal depression (KESSM) are positively associated with both responses
- Family poverty (SED_4) seems to affect i-SDQ only
- White children have lower i-SDQ
- Males have higher e-SDQ

Model for M-$q = 0.75$

	i-SDQ		e-SDQ	
	Est	se	Est	se
Age	-0.01	0.01	**-0.47**	0.01
Age^2	**0.08**	0.01	**0.24**	0.01
ALE_{11} mean	**0.19**	0.04	**0.32**	0.06
ALE_{11}	**0.08**	0.02	**0.10**	0.03
SED_4 mean	**0.12**	0.05	**0.23**	0.07
SED_4	-0.03	0.04	0.00	0.05
Kessm mean	**0.24**	0.01	**0.26**	0.02
Kessm	**0.10**	0.01	**0.13**	0.01
Degree	**-0.78**	0.12	**-1.40**	0.18
Gcse	**-0.48**	0.11	**-0.60**	0.15
White	**-0.34**	0.12	0.42	0.22
Male	**0.17**	0.05	**0.97**	0.10
IMD mean	**-0.05**	0.02	**-0.05**	0.02
IMD	-0.01	0.02	-0.03	0.03
Ethnic st.	0.22	0.13	-0.05	0.25
Disadv st.	0.06	0.07	0.18	0.12
σ_u	1.73		2.54	

- ALE_{11}, SED_4, and Kessm positively affect both responses and their impact is higher wrt $q = 0.50$
- Males have more severe internalising and externalising problems that females
- Children living in less deprived areas (higher IMD) have lower i-SDQ and e-SDQ

Model for M-$q = 0.90$

	i-SDQ		e-SDQ	
	Est	se	Est	se
Age	**0.04**	0.01	**-0.46**	0.02
Age^2	**0.09**	0.01	**0.25**	0.01
ALE_{11} mean	**0.37**	0.06	**0.51**	0.08
ALE_{11}	**0.10**	0.03	**0.10**	0.04
SED_4 mean	**0.21**	0.08	**0.34**	0.10
SED_4	-0.05	0.06	**0.01**	0.07
Kessm mean	**0.35**	0.02	**0.36**	0.03
Kessm	**0.13**	0.02	**0.16**	0.02
Degree	**-1.05**	0.14	**-1.65**	0.21
Gcse	**-0.63**	0.13	**-0.75**	0.19
White	**-0.42**	0.13	0.37	0.24
Male	**0.35**	0.09	**1.25**	0.14
IMD mean	**-0.09**	0.02	**-0.07**	0.03
IMD	-0.00	0.04	-0.03	0.04
Ethnic st.	0.14	0.16	-0.18	0.25
Disadv st.	-0.02	0.11	0.25	0.18
σ_u	1.70		2.40	

- The effect of ALE_{11}, SED_4, maternal depression (KESSM), and neighbourhood deprivation (IMD) becomes much stronger for high SDQ scores
- Severe problems are less likely with higher mother's educational levels
- The effect of race and gender becomes more evident for higher percentiles

Conclusions

- We have developed Q and MQ regression models that can deal with dependent observations: the dependence within observations from the same individual is modelled via individual-specific discrete random parameters
- By suitably setting the tuning constant c to a large value, we get Finite Mixtures of Expectile regression models
- Nonparametric distribution of the random effects is more in the spirit of Q and MQ models
- It is possible to carry out a ML inference and obtain analytical SEs
- It can be extended to handle Multivariate outcomes

Future developments

- Consider time-varying random parameters to model sources of unobserved heterogeneity that evolve over time, e.g. via Latent Markov Models (Farcomeni, 2012)
- Extension to zero-inflated data
- Extension to count data
- Application in the small area estimation setting (focus is on prediction, rather than estimation)