

# Quantile Regression for Longitudinal Data

Roger Koenker

University of Illinois, Urbana-Champaign

University of Copenhagen 18-20 May 2016

# Classical Linear Fixed/Random Effects Model

Consider the model,

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \alpha_i + u_{ij} \quad j = 1, \dots, m_i, \quad i = 1, \dots, n,$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}.$$

The matrix  $\mathbf{Z}$  represents an incidence matrix that identifies the  $n$  distinct individuals in the sample. If  $\mathbf{u}$  and  $\boldsymbol{\alpha}$  are independent Gaussian vectors with  $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$  and  $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ . Observing that  $\mathbf{v} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}$  has covariance matrix  $\text{E}\mathbf{v}\mathbf{v}^T = \mathbf{R} + \mathbf{Z}\mathbf{Q}\mathbf{Z}^T$ , we can immediately deduce that the minimum variance unbiased estimator of  $\boldsymbol{\beta}$  is,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T (\mathbf{R} + \mathbf{Z}\mathbf{Q}\mathbf{Z}^T)^{-1} \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{R} + \mathbf{Z}\mathbf{Q}\mathbf{Z}^T)^{-1} \mathbf{y}.$$

## A Penalty Interpretation of $\hat{\beta}$

**Proposition.**  $\hat{\beta}$  solves  $\min_{(\alpha, \beta)} \|y - X\beta - Z\alpha\|_{R^{-1}}^2 + \|\alpha\|_{Q^{-1}}^2$ , where  $\|x\|_A^2 = x^\top Ax$ .

### Proof.

Differentiating we obtain the normal equations,

$$X^\top R^{-1} X \hat{\beta} + X^\top R^{-1} Z \hat{\alpha} = X^\top R^{-1} y$$

$$Z^\top R^{-1} X \hat{\beta} + (Z^\top R^{-1} Z + Q^{-1}) \hat{\alpha} = Z^\top R^{-1} y$$

Solving, we have  $\hat{\beta} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} y$  where

$$\Omega^{-1} = R^{-1} - R^{-1} Z (Z^\top R^{-1} Z + Q^{-1})^{-1} Z^\top R^{-1}.$$

But  $\Omega = R + ZQZ^\top$ , see e.g. Rao(1973, p 33.). □

This result has a long history: Henderson(1950), Goldberger(1962), Lindley and Smith (1972), etc.

## Quantile Regression with Fixed Effects

Suppose that the conditional quantile functions of the response of the  $j$ th observation on the  $i$ th individual  $y_{ij}$  takes the form:

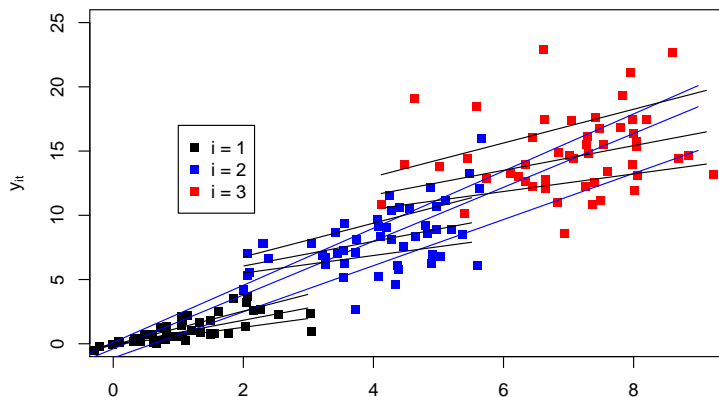
$$Q_{y_{ij}}(\tau|x_{ij}) = \alpha_i + x_{ij}^\top \beta(\tau) \quad j = 1, \dots, m_i, \quad i = 1, \dots, n.$$

In this formulation the  $\alpha$ 's have a pure location shift effect on the conditional quantiles of the response. The effects of the covariates,  $x_{ij}$  are permitted to depend upon the quantile,  $\tau$ , of interest, but the  $\alpha$ 's do not. To estimate the model for several quantiles simultaneously, we propose solving,

$$\min_{(\alpha, \beta)} \sum_{k=1}^q \sum_{j=1}^n \sum_{i=1}^{m_i} w_k \rho_{\tau_k}(y_{ij} - \alpha_i - x_{ij}^\top \beta(\tau_k))$$

Note that the usual between/within transformations are not permitted.

# Penalized Quantile Regression with Fixed Effects



Time invariant, individual specific intercepts are quantile independent;  
slopes are quantile dependent.

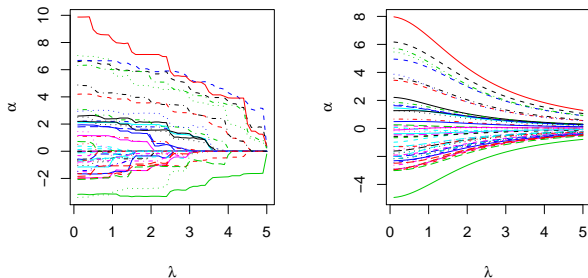
## Penalized Quantile Regression with Fixed Effects

When  $n$  is large relative to the  $m_i$ 's shrinkage may be advantageous in controlling the variability introduced by the large number of estimated  $\alpha$  parameters. We will consider estimators solving the penalized version,

$$\min_{(\alpha, \beta)} \sum_{k=1}^q \sum_{j=1}^n \sum_{i=1}^{m_i} w_k \rho_{\tau_k}(y_{ij} - \alpha_i - x_{ij}^T \beta(\tau_k)) + \lambda \sum_{i=1}^n |\alpha_i|.$$

For  $\lambda \rightarrow 0$  we obtain the fixed effects estimator described above, while as  $\lambda \rightarrow \infty$  the  $\hat{\alpha}_i \rightarrow 0$  for all  $i = 1, 2, \dots, n$  and we obtain an estimate of the model purged of the fixed effects. In moderately large samples this requires sparse linear algebra. Example R code is available from my webpages.

# Shrinkage of the Fixed Effects



Shrinkage of the fixed effect parameter estimates,  $\hat{\alpha}_i$ . The left panel illustrates an example of the  $\ell_1$  shrinkage effect. The right panel illustrates an example of the  $\ell_2$  shrinkage effect.

# Dynamic Panel Models and IV Estimation

Galvao (2010) considers dynamic panel models of the form:

$$Q_{y_{it}}(\tau|y_{i,t-1}, x_{it}) = \alpha_i + \gamma(\tau)y_{i,t-1} + x_{it}^\top \beta(\tau) \quad t = 1, \dots, T_i, \quad i = 1, \dots, n.$$

In “short” panels estimation suffers from the same bias problems as seen in least squares estimators Nickel (1981) Hsiao and Anderson (1981); using the IV estimation approach of Chernozhukov and Hansen (2004) this bias can be reduced.



## Correlated Random Effects

Abrevaya and Dahl (JBES, 2008) adapt the Chamberlain (1982) correlated random effects model and estimate a model of birthweight like that of Koenker and Hallock (2001).

The R package `rqp` written with Stefan Bache implements both this method and the penalized fixed effect approach. Available from R-Forge with the command:

```
install.packages("rqp", repos="http://R-Forge.R-project.org")
```

This is a challenging, but very important, problem and hopefully there will be new and better approaches in the near future.

## Recent Developments

There is an extensive recent literature including:

- Arellano and Bonhomme, (2016) Nonlinear Panel Data Estimation via Quantile Regression, technical report.
- Graham, B., J. Hahn, A. Poirier, and J. Powell (2015): Quantile Regression with Panel Data, technical report.
- Galvao, A. F., K. Kato and G. Montes-Rojas (2012): Asymptotics for quantile regression models with individual effects, *Journal of Econometrics*, 170, 7691.
- Chernozhukov, V., I. Fernandez-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015) Nonparametric Identification in Panels using Quantiles, *Journal of Econometrics*, forthcoming.
- Canay, I. A. (2011) A Simple Approach to Quantile Regression for Panel Data, *Econometrics Journal*, 14 (3), 368386.
- Arellano, M., and M. Weidner (2015): Instrumental Variable Quantile Regressions in Large Panels with Fixed Effects, UCL Technical report.
- Lamarche, C. (2010): Robust Penalized Quantile Regression for Panel Data, *Journal of Econometrics*, 157, 396408.
- Chetverikov, Denis, Bradley Larsen, and Christopher Palmer, (2016), IV Quantile Regression for group-level treatments, with an application to the effects of trade on the distribution of wages, *Econometrica*, 84, 809–834.