Quantile Regression for Longitudinal Data

Roger Koenker

CEMMAP and University of Illinois, Urbana-Champaign

LSE: 17 May 2011

Classical Linear Fixed/Random Effects Model

Consider the model,

$$y_{ij} = x_{ij}^\top \beta + \alpha_i + u_{ij} \quad j = 1, ...m_i, \quad i = 1, ..., n,$$

or

$$y = X\beta + Z\alpha + u.$$

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

$$\hat{\beta} = (X^\top (R + ZQZ^\top)^{-1} X)^{-1} X^\top (R + ZQZ^\top)^{-1} y.$$

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

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

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 jth observation on the ith individual y_{ij} takes the form:

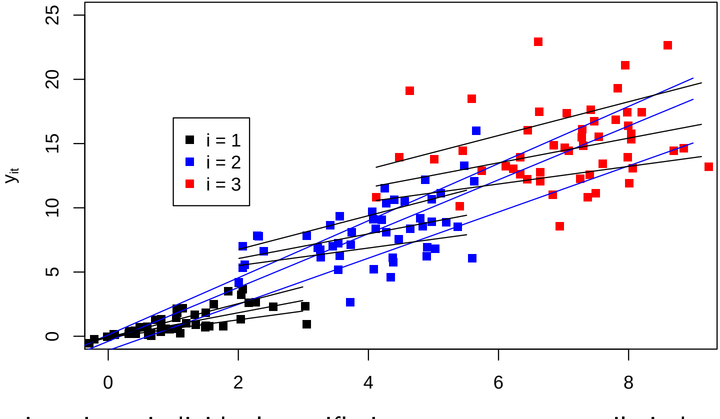
$$Q_{y_{ij}}(\tau|x_{ij}) = \alpha_i + x_{ij}^\top \beta(\tau) \quad j = 1, ...m_i, \quad i = 1, ..., n.$$

In this formulation the α'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, τ, of interest, but the α'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.

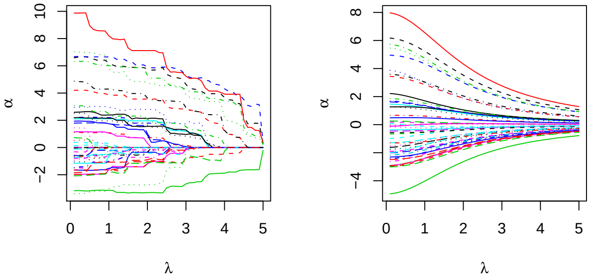
Penalized Quantile Regression with Fixed Effects

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 α parameters. We will consider estimators solving the penalized version,

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

For $\lambda \to 0$ we obtain the fixed effects estimator described above, while as $\lambda \to \infty$ the $\hat{\alpha}_i \to 0$ for all $i = 1, 2, ..., n$ 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 ℓ_1 shrinkage effect. The right panel illustrates an example of the ℓ_2 shrinkage effect.

Dynamic Panel Models and IV Estimation

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

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

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 `rqpd` implements both this method and the penalized fixed effect approach. Available from R-Forge with the command:

```
install.packages("rqpd", 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.