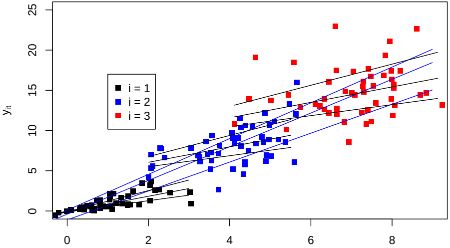
Quantile Regression for Panel/Longitudinal Data

Roger Koenker

University of Illinois, Urbana-Champaign

University of Minho 12-14 June 2017



Classical Linear Fixed/Random Effects Model

Consider the model,

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

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), penalty as Gaussian prior on α.

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:

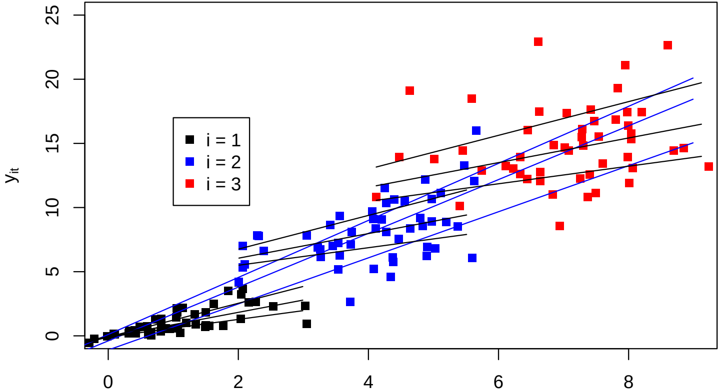
$$Q_{y_{ij}}(\tau|x_{ij}) = \alpha_i + x_{ij}^\top \beta(\tau) \quad j = 1, \ldots m_i, \quad i = 1, \ldots, n.$$

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 helpful here due to the nonlinearity of the estimator. Differences in quantiles are NOT quantiles of differences.

Penalized Quantile Regression with Fixed Effects



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

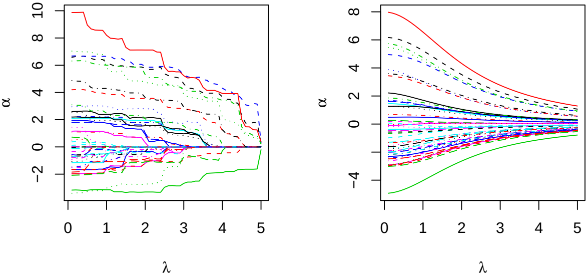
Penalized Quantile Regression with Fixed Effects

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 $i = 1, 2, \ldots, n$ 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. Penalty does not *have* to be interpreted as a prior, but it may be so interpreted.

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 (lasso) shrinkage effect. The right panel illustrates an example of the ℓ_2 (ridge) shrinkage effect.

Selection of the Penalty Parameter

There are many ways to skin the λ selection cat:

- Stein unbiased risk criterion,
- Cross validation, and its offspring,
- AIC, SIC and related effective dimension methods

Desirable to balance variability of the (in)fidelity term and the variability of the penalty term. See Lamarche (2010) and the vast, related literature on lasso λ selection.

A Half-Baked Empirical Bayes Proposal

Shrinking all the α_i's toward zero seems a bit extreme, a more plausible idea would be to shrink them toward a set of common group values, as in recent work by Manresa and Bonhomme (2015), who use k-means clustering. An alternative approach would be to employ the Kiefer-Wolfowitz (1956) NPMLE for mixture models:

1. Estimate initial unrestricted α's
2. Treating the $\hat{\alpha}_i$'s as a sample from a Gaussian mixture model, estimate the mixing distribution, F, using the Kiefer-Wolfowitz NPMLE, produces discrete mixing distribution,
3. Using the mixture density as a (prior) penalty function reestimate the α's shrinking them toward their nearest group center
4. Iterate

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 seen in least squares estimators as noted by Nickel (1981) Hsiao and Anderson (1981) and others. Using the QRIV approach of Chernozhukov and Hansen (2004), Galvao shows that this bias can be reduced.

Correlated Random Effects

Abrevaya and Dahl (2008) and Bache, Dahl and Kristensen (2013) adapt the Chamberlain (1982) correlated random effects model and estimate a model of birthweights for panel data.

The R package `rqpd` (K and Bache) 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 there has been considerable recent attention devoted to it by prominent researchers.

Recent Developments

There is an extensive recent literature including:

- Arellano and Bonhomme, (2016) Nonlinear Panel Data Estimation via Quantile Regression, Econometrics Journal.
- Graham, B., J. Hahn, A. Poirier, and J. Powell (2015): Quantile Regression with Panel Data, NBER Working Paper 21034.
- 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. Fernández-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.
- 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.

Arellano-Bonhomme Model

Arellano-Bonhomme view the individual specific effects as latent covariates,

$$Q_{Y_{it}}(\tau|x_{it}, \eta_i) = x_{it}^\top \beta(\tau) + \eta_i \gamma(\tau)$$

or in random coefficient form,

$$y_{it} = x_{it}^\top \beta(u_{it}) + \eta_i \gamma(u_{it}) \equiv w_{it}(\eta_i)^\top \theta(u_{it}),$$

for iid $u_{it} \sim U[0,1]$, so $\theta = (\beta, \gamma)$ viewed as random coefficients are comonotonic. In the spirit of Chamberlain (1982) the η_i's are assumed to arise from (what else??) another quantile regression model,

$$Q_{\eta_i}(\tau|x_i) = z_i^\top \delta(\tau)$$

where $z_i = \varphi(x_i)$, some basis expansion of the x_i.

Arellano-Bonhomme Estimation

The parameters, $\theta(\tau), \delta(\tau))$ can obviously be estimated on a grid of τ values on $(0,1)$, except (oops!) we don't observe the η_i's.
A "Stochastic EM" algorithm is proposed:

1. Initialize (θ, δ) on $\tau \in (0,1)$ grid,
2. Generate T_i values for each of the η_i's from the estimated density,

$$\tilde{f}(\eta | y, x, \theta, \delta) = \frac{\prod_{t=1}^{T_i} f_{Y_t|X_t,\eta}(y_t | x_t, \eta, \theta) f_{\eta|X}(\eta | x, \delta)}{\int \prod_{t=1}^{T_i} f_{Y_t|X_t,\eta}(y_t | x_t, \tilde{\eta}, \theta) f_{\eta|X}(\tilde{\eta} | x, \delta) d\tilde{\eta}}$$

3. Re-estimate (θ, δ) after plugging in the generated η's,
4. Iterate

The density estimation step is facilitated by the recent addition of the density prediction option in the `quantreg` package. The algorithm combines features of the traditional EM algorithm, MCMC, and a large helping of standard quantile regression optimization.

Conclusions

Quantile regression methods for panel data is still a very active research area and it would be premature to draw any definitive conclusions, nevertheless:

- Shrinkage methods for fixed effect models still have some appeal and offer interesting avenues for development,
- Under special circumstances the simple Hausman-Taylor approach of Chetverikov, Larson and Palmer (2016) can be quite effective and expedient,
- In my view, the most promising general approach is that of Arellano and Bonhomme because
  - Conditioning on latent effects avoids conceptual difficulties with various other random effect formulations,
  - Conditional density estimation opens the way to various other local likelihood methods for other problems such as sample selection and errors in variables that are currently under development.