QUANTILE REGRESSION FOR LONGITUDINAL DATA

ROGER KOENKER

Abstract. The penalized least squares interpretation of the classical random effects estimator suggests a possible way forward for quantile regression models with a large number of fixed effects. Sparse linear algebra and interior point methods for solving large linear programs are essential practical tools.

1. Introduction

The almost exclusive focus on least squares estimators under Gaussian conditions for longitudinal data analysis can be taken as a challenge. Can a more flexible, more robust approach to longitudinal data analysis be forged outside the Gaussian random effects framework? I will argue that quantile regression might play a constructive role in such a development.

Recent contributions to the literature on linear and nonlinear mixed models have emphasized the strong link with penalty methods for nonparametric function estimation. Shrinkage of highly overparameterized models toward simpler, plausible models suggested by prior smoothness considerations or exchangeability of nominal effects share many common features. The construction of infant and adolescent growth charts provides a motivating application in which both ordinal and nominal factors appear. Both Cox and Jones have recently suggested in the discussion of Cole (1988) that quantile regression methods may offer advantages over conventional parametric approaches to the analysis of growth charts. Computational methods that exploit the inherently sparse nature of the linear algebra for interior point solution of the resulting linear programming problems play an essential role.

2. Models and Methods

Consider the classical linear random effects model,

\[ y_{ij} = x_{ij}' \beta + \alpha_i + u_{ij} \quad j = 1, \ldots, m_i, \quad i = 1, \ldots, n, \]

which we will write in matrix form as,

\[ y = X\beta + Z\alpha + u. \]

---

Version: October 24, 2003. The author would like to express his appreciation to Steve Portnoy, Xuming He, and Carlos Lamarche for discussions related to this work.
Suppose \( u \) and \( \alpha \) 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
\[
Evv^\top = R + ZZ^\top,
\]
we can immediately deduce that the minimum variance unbiased estimator of \( \beta \) is,
\[
\hat{\beta} = (X^\top (R + ZZ^\top)^{-1} X)^{-1} X^\top (R + ZZ^\top)^{-1} y.
\]
This estimator is certainly not very appealing from a robustness standpoint, but the optimization problem that gives rise to \( \hat{\beta} \) is suggestive of a larger class of possible candidate estimators.

**Proposition.** \( \hat{\beta} \) solves \( \min_{(\alpha, \beta)} \|y - X\beta - Z\alpha\|_{R^{-1}}^2 + \|\alpha\|_Q^2 \), 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 \( \Omega = R + ZZ^\top \), see e.g. Rao(1973, p 33.). \( \blacksquare \)

This result has a long history. Robinson (1991) attributes the normal equations above to Henderson (1950). Goldberger (1962) introduced the terminology “best linear unbiased predictor”, subsequently rendered as BLUP, to describe the estimator \( \hat{\beta} \) and its associated “estimator” \( \hat{\alpha} \) of the random effects. The implicit estimation of the random effects may appear strange, but viewing the random effects estimator as a penalized least squares estimator opens the door to the consideration of alternative penalties. In the Bayesian paradigm the penalty formulation is natural, as emphasized by Lindley and Smith (1972), and alternatives to the Gaussian penalty \( \|\alpha\|_Q^{-2} \), simply reflects differences in prior beliefs about the \( \alpha \)’s. By shrinking the unconstrained \( \hat{\alpha} \)’s toward a common value we achieve not only improved performance of the individual fixed-effect estimates, but also for the estimate of \( \beta \).

### 2.1. Quantile Regression with Fixed Effects.

Contemplating the extension of the model (2.1) to models for conditional quantile functions we must first confront the question: What role should the \( \alpha \)’s play? If the number of observations \( m_i \) were large for each cross-sectional unit then we might hope to estimate a distributional shift \( \alpha_i(\tau) \) for each individual. However, in most applications the \( m_i \) are relatively modest and then it is quite unrealistic to attempt to estimate a \( \tau \)-dependent individual effect. At best we may be able to estimate an individual specific location-shift effect, and even this may strain credulity.

We will consider the model
\[
Q_{q_0}(\tau|x) = \alpha_i + x_{ij}^\top \beta(\tau) \quad j = 1, ..., m_i, \quad i = 1, ..., n.
\]

2 Quantile Regression for Longitudinal Data
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 (2.2) for several quantiles we propose solving,

$$
(2.3) \quad \min_{(\alpha, \beta)} \sum_{k=1}^{q} \sum_{j=1}^{n} \sum_{i=1}^{m_k} w_k \rho_{r_k}(y_{ij} - \alpha_i - x_{ij}^\top \beta(\tau))
$$

where $\rho_{r_k}(u) = u(\tau - I(u < 0))$, as in Koenker and Bassett (1978). The weights $w_k$ control the relative influence of the $q$ quantiles on the estimation of the $\alpha_i$ parameters. Koenker (1984) considered a reversed situation in which only an intercept parameter was permitted to depend upon $\tau$ and the slope parameters associated with the included covariates were constrained to be identical for several $\tau$’s. The weighted sum of the quantile regression objective functions acts somewhat like an L-estimator with discrete weights.

Solving the problem (2.3) may appear somewhat quixotic when the dimensions $n$, $m$ and $q$ are large. In least squares applications the usual strategy would be to transform $y$ and $X$ to deviations from individual means, and then compute $\hat{\beta}$ from the transformed data. For quantile regression this composition of projections isn’t available and we are required to deal directly with the full problem. Fortunately, in typical applications the problem is quite sparse, that is the design matrix of the full problem is mostly zeros. Storing the dense version of this matrix with all the zeros treated as double precision floats may well be infeasible, but standard sparse matrix storage schemes that store only the non-zero elements and their indexing locations are quite feasible. Interior point methods for solving (2.3) proceed iteratively by solving a sequence of diagonally weighted least squares steps using a Cholesky factorization. The sparsity of the design is typically preserved quite well in this factorization, as noted by Saunders (1994), and the computational effort is roughly proportional to the number of non-zero elements. Implementations of this approach for the public domain dialect R, Ihaka and Gentleman (1996), of Chambers (1998) S language are discussed in Koenker and Ng (2003) and are available on CRAN at www.r-project.org.

2.2. Penalized Quantile Regression with Fixed Effects. We have seen that the optimal estimator for the Gaussian prototype model (2.1) involves shrinking the $\hat{\alpha}$’s toward a common value. When the $x_{ij}$ contain an intercept, as we will henceforth assume, this common value can be taken to the the conditional mean of the response at a point determined by the centering of the other covariates. In the quantile regression version of the model (2.2) this would be some corresponding conditional quantile of the response, although this would require further conditions including symmetry of the $\tau_k$’s and the $w_k$’s to be specified.

Particularly when $n$ is large relative to $\sum m_i$ shrinkage is advantageous in controlling the variability introduced by the large number of estimated $\alpha$ parameters. For the
quantile loss function, \( p \) it is convenient to use the \( \ell_1 \) penalty,

\[
P(\alpha) = \sum_{i=1}^{n} |\alpha_i|
\]

in place of the conventional Gaussian penalty. This choice maintains the linear programming form of the problem and also preserves the sparsity of the resulting design matrix. Several authors, notably Tibshirani (1996) and Donoho, Chen, and Saunders (1998), have pointed out that \( \ell_1 \) shrinkage offers some advantages over more conventional Gaussian \( \ell_2 \) penalties.

We will consider estimators solving,

\[
(2.4) \quad \min_{(\alpha, \beta)} \sum_{k=1}^{q} \sum_{j=1}^{n} \sum_{i=1}^{m} w_k \rho_{\tau_k}(\alpha_i - x_{ij}^T \beta(\tau_k)) + \lambda \sum_{i=1}^{n} |\alpha_i|.
\]

For \( \lambda \to 0 \) we obtain the fixed effects estimator described above, while for \( \lambda \to \infty \) 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. Note that since \( x_{ij} \) is assumed to contain an intercept, in either case we will also have \( q \) \( \tau \)-specific estimates of the intercept. If you consider the special case that \( m_i \equiv m \) for all \( i \), we can write the design matrix for a single quantile as,

\[
[X: I_n \otimes e_m]
\]

where \( X = (x_{ij}) \) is \( nm \) by \( p \), and \( e_m \) denotes an \( m \)-vector of ones. The design matrix for \( q > 1 \) may be written as,

\[
[W \otimes X: w \otimes (I_n \otimes e_m)].
\]

Appending the the penalty term we have the augmented design matrix,

\[
\begin{bmatrix}
W \otimes X & w \otimes (I_n \otimes e_m) \\
0 & \lambda I_n
\end{bmatrix}
\]

which has dimension \( qnm + n \) by \( qp + n \). The corresponding response vector is \( \tilde{y} = ((w \otimes y)^T 0_n^T)^T \). These dimensions may seem even more daunting than before, but again the sparsity of the design matrix comes to the rescue, and we will see that quite large problems of this type can be handled successfully on rather modest machines.

3. \textsc{asymptopia}

To explore the asymptotic behavior of the penalized quantile regression estimator solving (2.3) we will consider a balanced design for which (2.2) holds with \( m_i = m \) for all \( i = 1, \ldots, n \). Then since \( Z = I_n \otimes e_m \), we have \( Z^T Z = mI_n \). We will consider
settings in which both $n$ and $m$ tend to infinity. For convenience of exposition we will also assume that $\alpha_i \neq 0$ for $i = 1, \ldots, n$. Let

$$
V_{mn}(\delta) = \sum_{k=1}^{q} \sum_{j=1}^{m} \sum_{i=1}^{n} w_k [\rho_{ij} (y_{ij} - \xi_{ij}(\tau_k)) - z_{ij}^T \delta_0 / \sqrt{m} - x_{ij}^T \delta_k / \sqrt{mn}]
$$

$$
- \rho_{ij} (y_{ij} - \xi_{ij}(\tau_k)] + \lambda_m \sum_{i=1}^{n} |\alpha_i - \delta_{0i} / \sqrt{m}| - |\alpha_i|,
$$

where $\xi_{ij}(\tau_k) = \alpha_i + x_{ij}^T \beta(\tau_k)$. Note that

$$
\hat{\delta} = \begin{pmatrix}
\hat{\delta}_0 \\
\hat{\delta}_1 \\
\vdots \\
\hat{\delta}_q
\end{pmatrix} = \begin{pmatrix}
\sqrt{m}(\hat{\alpha} - \alpha) \\
\sqrt{mn}(\hat{\beta}(\tau_1) - \beta(\tau_1)) \\
\vdots \\
\sqrt{mn}(\hat{\beta}(\tau_q) - \beta(\tau_q))
\end{pmatrix}
$$

minimizes the function $V_{mn}$. We will impose the following regularity conditions:

A1. The conditional distribution function of $y_{ij}$ given $x_{ij}, F_{ij}$, has continuous density, $0 \leq f_{ij} < \infty$ at $\xi_{ij}(\tau_k)$ for $k = 1, \ldots, q$, $j = 1, \ldots, m$, $i = 1, \ldots, n$.

A2. Let $\Omega$ denote the $q$ by $q$ matrix with typical element $\tau_k \land \tau_l - \tau_k \tau_l$ and $\Phi_k = \text{diag}(f_{ij}(\xi_{ij}(\tau_k)))$. The limiting forms of the following matrices are positive definite:

$$
D_0 = \lim_{m \to \infty} m^{-1} \begin{pmatrix}
w^T \Omega w Z^T Z \\
W \Omega W \otimes X^T Z / \sqrt{n} \\
W \Omega W \otimes X^T X / \sqrt{n}
\end{pmatrix},
$$

$$
D_1 = \lim_{m \to \infty} m^{-1} \begin{pmatrix}
\sum w_k Z^T \Phi_k Z & w_1 X^T \Phi_1 Z / \sqrt{n} & \cdots & w_q X^T \Phi_q Z / \sqrt{n} \\
w_1 Z^T \Phi_1 X / \sqrt{n} & w_1 X^T \Phi_1 X / mn & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
w_q Z^T \Phi_q X / \sqrt{n} & 0 & \cdots & w_q X^T \Phi_q X / mn
\end{pmatrix}.
$$

A3. $\max_{1 \leq i \leq n} \max_{1 \leq j \leq m} ||x_{ij}|| / \sqrt{mn} \to 0$.

**Theorem 1.** Under conditions A1-3, provided that $\lambda_m / \sqrt{m} \to \lambda_0$,

$$
\hat{\delta} \sim \text{argmin} V_0(\delta)
$$

where

$$
V_0(\delta) = -\delta^T B + \frac{1}{2} \delta^T D_1 \delta + \lambda_0 \delta^T s
$$

where $B$ is a zero mean Gaussian vector with covariance matrix $D_0$, and $s = (s_0^T 0_{pq})^T$ and $s_0 = (\text{sgn}(\alpha_i))$. 
Proof: The function $V_{mn}$ can be decomposed into three parts using the identity of Knight (1998),

$$\rho_r(u - v) - \rho_r(u) = -v \psi_r + \int_0^v (I(u \leq s) - I(u \leq 0)) ds$$

where $\psi_r(u) = \tau - I(u < 0)$ denotes the quantile influence function. We will write,

$$V_{mn}(\delta) = V_{mn}^{(1)}(\delta) + V_{mn}^{(2)}(\delta) + V_{mn}^{(3)}(\delta),$$

where $v_{ijk} = z_{ij}^\top \delta_0 + x_{ij} \delta_k / \sqrt{n}$ and,

$$V_{mn}^{(1)}(\delta) = -m^{-1/2} \sum_k \sum_j \sum_i w_k \psi_{\tau_k} (y_{ij} - \xi_{ij}(\tau_k)) v_{ijk}$$

$$V_{mn}^{(2)}(\delta) = -m^{-1/2} \sum_k \sum_j \sum_i w_k \int_0^{v_{ijk}} (F_{ij}(\xi_{ij}(\tau_k) + t/\sqrt{m}) - F_{ij}(\xi_{ij}(\tau_k))) dt$$

$$V_{mn}^{(3)}(\delta) = \lambda_m \sum_i |\alpha_i - \delta_{0i}/\sqrt{m}| - |\alpha_i|.$$ 

The first term is asymptotically Gaussian. Let $\Psi_k = \text{diag}(\psi_{\tau_k}(y_{ij} - \xi_{ij}(\tau_k)))$ and note that $\Psi_k e_{mn} e_{mn}^\top \Psi_{t_i} = (\tau_k ^\top \tau_i - \tau_k \tau_i) I_{mn}$. Conditions A2-3 imply a Lindeberg condition and we have,

$$V_{mn}^{(1)}(\delta) = -m^{-1/2} \sum_k w_k (Z^\top \Psi_k \delta_0 + X^\top \Psi_k \delta_k)$$

$$\sim -B \delta.$$ 

The second term is asymptotically quadratic in $\delta$,

$$V_{mn}^{(2)}(\delta) = m^{-1} \sum_k \sum_j \sum_i w_k \int_0^{v_{ijk}} \sqrt{m} (F_{ij}(\xi_{ij}(\tau_k) + t/\sqrt{m}) - F_{ij}(\xi_{ij}(\tau_k))) dt$$

$$= m^{-1} \sum_k \sum_j \sum_i w_k \int_0^{v_{ijk}} f_{ij}(\xi_{ij}(\tau_k)) dt + o(1)$$

$$= \frac{1}{2m} \sum_k \sum_j \sum_i w_k f_{ij}(\xi_{ij}(\tau_k)) (z_{ij}^\top \delta_0 + x_{ij} \delta_k / \sqrt{m})^2 + o(1)$$

$$= \frac{1}{2m} \sum_k w_k (\delta_0^\top Z^\top \Phi_k Z \delta_0 + 2 \delta_0^\top Z^\top \Phi_k X \delta_k / \sqrt{n} + \delta_k^\top X^\top \Phi_k X \delta_k / n) + o(1)$$

$$\to \frac{1}{2} \delta^\top D_1 \delta.$$
Finally,

\[ V_{mn}^{(3)}(\delta) = \frac{\lambda_m}{\sqrt{m}} \sum_{i=1}^{n} \delta_i \text{sgn}(\alpha_i) \]

\[ \rightarrow \lambda_0 \delta_0^T s. \]

Convexity of the objective function, \( V_{mn} \), and the uniqueness of the minimum of \( V_0 \) yields uniformity in \( \delta \) completing the argument as described by Knight and Fu (2000).

**Corollary 1.** Under the conditions of the preceding theorem,

\[ \hat{\delta} \sim \mathcal{N}(\lambda_0 D_1^{-1} s, D_1^{-1} D_0 D_1^{-1}). \]

The corollary gives an explicit expression of the bias-variance tradeoff at least for the case that both \( m \) and \( n \) are large. It would be clearly desirable to explore what happens when \( m \) is fixed and only \( n \) tends to infinity. The assumption that the \( \alpha_i \)'s are non-zero can be relaxed as in Knight and Fu (2000) and simply produces another term in \( V_0(\delta) \).

The conditions A1 and A3 are now quite standard in the quantile regression literature. Condition A2 is not, but if one supposes for the moment that the model is of the pure location shift form (2.1), then \( D_1 \) simplifies somewhat and A2 reduces to a condition on the matrices \( X^T X/(mn) \) and \( Z^T Z/m \). We have seen that the latter is equal to \( I_n \), and the former condition is again familiar. If \( Z^T X = 0 \) so that there is no “between” variability in the \( x \)'s, then the expressions simplify considerably, but this case is quite atypical, and generally we would expect that there would be some potential improvement in the estimation of \( \beta \)'s due to the shrinkage of the \( \alpha \)'s. These expectations are investigated further in the next section through a small simulation experiment.

4. **Monte-Carlo**

In this section a very brief glimpse into the finite sample behavior of the penalized quantile regression estimator is offered. I begin by contrasting the shrinkage effect of \( \ell_1 \) and \( \ell_2 \) penalty methods. Consider a simple example with \( n = 50 \) and \( m = 5 \) and response generated by the model,

\[ y_{ij} = \alpha_i + u_{ij} \]

with \( \alpha_i \)'s iid from the \( \chi^2_3 \) distribution, and \( u_{ij} \) iid also from \( \chi^2_3 \). In the left panel of Figure 1 we illustrate the estimated, \( \hat{\alpha}_i \)'s as a function of the regularization parameter \( \lambda \). Here we have used the estimator (2.3) with weights \( w = (.25, .50, .25) \) on the three quartiles. The \( x_{ij} \)'s were generated as Gaussian according to (4.3) below. In the right panel we illustrate the corresponding shrinkage effects for the \( \ell_2 \) Gaussian penalty method. The \( \ell_1 \) shrinkage method is more tolerant of large discrepancies; note that the gradient condition condition involves only the signs of the estimated effects, not
their magnitudes, so highly unusual \( \alpha_i \)'s can be substantially shrunken toward zero without the extreme prejudice implied by the \( \ell_2 \) criterion.

Two simple versions of our basic model are considered in the simulation experiments. In the first, reported in Table 1, the scalar covariate, \( x_{ij} \), exerts a pure location shift effect. In the second, reported in Table 2, \( x_{ij} \) has a both a location and scale shift effect. In the former case the response, \( y_{ij} \), is generated by the model,

\[
y_{ij} = \alpha_i + x_{ij}\beta + u_{ij}
\]

while in the latter case,

\[
y_{ij} = \alpha_i + x_{ij}\beta + (1 + x_{ij}\gamma)u_{ij}.
\]

Without loss of generality we will take \( \beta = 0 \). Interest will focus on the effect of the covariate, \( x_{ij} \), at the median. Sample sizes are fixed, with \( n = 50 \), and \( m = 5 \) for both versions of the model. In the first version of the model the covariate effect is clearly zero, in the second version of the model it depends on the choice of the quantile of interest and the form of the error distribution. In all cases the reported entries are based on 400 replications of the simulations.

A critical aspect governing the performance of penalty methods in these settings is the "between" versus "within" variability of the covariate. A convenient way to
summarize this is the interclass correlation coefficient. If we generate \( x_{ij} \)'s as
\[
(4.3) \quad x_{ij} = \gamma_i + v_{ij}
\]
with \( \gamma_i \) and \( v_{ij} \) independent and identically distributed over \( i \) and \( i,j \) respectively, then the interclass correlation coefficient,
\[
\rho_x = \frac{\sigma^2_\gamma}{(\sigma^2_\gamma + \sigma^2_u)},
\]
see e.g. Scheffé (1959, p. 223) is just the ordinary correlation coefficient between any two \( x_{ij} \) and \( x_{ik} \) observations with \( j \neq k \). We will take \( \rho_x = .75 \) in our simulations.

We consider three variants of model 1. In all three variants the \( x_{ij} \)'s are generated from (4.3) with both \( \gamma_i \)’s and \( v_{ij} \)’s as Gaussian with respective variances, 3 and 1. The response \( y \) is then generated from (4.1). In the first variant both the \( \alpha_i \)'s and \( u_{ij} \)'s are standard Gaussian, in the second variant both are Student \( t \) on three degrees of freedom, and in the third variant both are central \( \chi^2_3 \). So the interclass correlation coefficient of the response is 0.50 for all three variants.

Six estimators are considered: three from the least squares family, three from the quantile regression family. The ordinary least squares estimator (LS) simply ignores the \( \alpha_i \) effects entirely, maximally shrinking all of these estimates to zero. The penalized least squares estimator (PLS) is the classical random effects estimator for the model (2.1) using the (known) optimal variance ratio. The least squares fixed effects, or “within” estimator (LSFE) simply implements the unpenalized least squares estimator of the model (2.1). Correspondingly, the ordinary quantile regression estimator (QR) fully shrinks the \( \hat{\alpha}_i \)'s to zero, the fixed effects estimator (QRFE) shrinks them not at all, and the penalized quantile regression estimator (PQR) shrinks them with \( \lambda \) chosen to be the ratio of scale parameters \( \sigma_u/\sigma_\alpha \).

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>PLS</th>
<th>LSFE</th>
<th>QR</th>
<th>PQR</th>
<th>QRFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{N} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-0.0025</td>
<td>-0.0009</td>
<td>0.0005</td>
<td>-0.0023</td>
<td>-0.0022</td>
<td>0.0008</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0733</td>
<td>0.0529</td>
<td>0.0662</td>
<td>0.0813</td>
<td>0.0619</td>
<td>0.0716</td>
</tr>
<tr>
<td>( t_3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-0.0006</td>
<td>0.0037</td>
<td>0.0076</td>
<td>0.0057</td>
<td>0.0099</td>
<td>0.0077</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1648</td>
<td>0.1024</td>
<td>0.1168</td>
<td>0.1009</td>
<td>0.0715</td>
<td>0.0738</td>
</tr>
<tr>
<td>( \chi^2_3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.0016</td>
<td>0.0074</td>
<td>0.0125</td>
<td>0.0086</td>
<td>0.0001</td>
<td>0.0013</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1768</td>
<td>0.1426</td>
<td>0.1650</td>
<td>0.1853</td>
<td>0.1438</td>
<td>0.1410</td>
</tr>
</tbody>
</table>

**Table 4.1.**

Table 1 reports the results of the location shift simulations. Bias is small in all cases. In the Gaussian setting we see roughly the anticipated efficiency loss due to estimating the median rather than the mean. The gain from penalization, while
not overwhelming, is certainly worthwhile. In the \( t_3 \) setting the penalized quantile regression estimators do considerably better than their least squares competitors. The \( \chi^2 \) case is somewhat anomalous, since the penalized quantile regression estimator does slightly worse than the unpenalized fixed effects procedure, but both are competitive with the least squares results.

<table>
<thead>
<tr>
<th></th>
<th>LS</th>
<th>PLS</th>
<th>LSFE</th>
<th>QR</th>
<th>PQR</th>
<th>QRFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-0.0049</td>
<td>-0.0032</td>
<td>-0.0009</td>
<td>-0.0039</td>
<td>-0.0024</td>
<td>0.0025</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0444</td>
<td>0.0477</td>
<td>0.0775</td>
<td>0.0439</td>
<td>0.0451</td>
<td>0.0656</td>
</tr>
<tr>
<td>( t_3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.0009</td>
<td>0.0007</td>
<td>0.0006</td>
<td>0.0059</td>
<td>0.0066</td>
<td>0.0050</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0718</td>
<td>0.0767</td>
<td>0.1231</td>
<td>0.0536</td>
<td>0.0586</td>
<td>0.0777</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.0781</td>
<td>0.0805</td>
<td>0.0836</td>
<td>0.0136</td>
<td>0.0117</td>
<td>0.0001</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1394</td>
<td>0.1606</td>
<td>0.2257</td>
<td>0.0985</td>
<td>0.1063</td>
<td>0.1325</td>
</tr>
</tbody>
</table>

Table 4.2.

In the location-scale version of the model we adopt the same three distributions for generating the \( \alpha_i \)'s and the \( u_{ij} \)'s. In the location-scale model it is important that the resulting linear quantile functions do not cross, an eventuality we avoid by now taking the \( x_{ij} \)'s as \( \chi^2 \) instead of Gaussian. In the Gaussian and \( t_3 \) cases, since we are focusing on the estimation of the median effect, by symmetry the effect of the covariate \( x_{ij} \) on median response is still zero. However, in the \( \chi^2 \) case the median effect is,

\[
\beta(1/2) = \beta + \gamma Q_u(1/2),
\]

which in our case with \( \beta = 0 \) and \( \gamma = 1/10 \), is .236.

In Table 2 we report the results of the location-scale model simulations. Again, we see that the quantile regression estimators perform quite well in the \( t_3 \) case, but they now are also competitive even in the Gaussian case, a finding that may be attributed to the effect of the heteroscedasticity in this formulation of the model. It is also apparent that the imposing more aggressive shrinkage is helpful in these cases. The comparison in the \( \chi^2 \) case is somewhat difficult, since the procedures are inherently estimating different functions. The quantile regression methods are all intended to estimate the conditional median function and do reasonably well in the sense that the bias is still very modest. The least squares estimators are targeting the conditional mean function, which is now nonlinear in \( x_{ij} \), so we have evaluated both bias and root mean square error as if the least squares methods were also estimating the conditional median function. This obviously puts the least squares methods at some disadvantage.
5. EXTENSIONS

Almost everything remains to be investigated. As in all problems of regularization there are serious issues about the choice of the regularization parameter, $\lambda$. There are many variants of the model that would extend the oneway layout structure, including the incorporation of ordinal factors and nonparametric smoothing components. The analysis of the performance of the methods for fixed $m_i$ sample sizes is also a critical direction for future research. Applications to reference growth curves would appear to be a natural laboratory for further development of quantile regression models for longitudinal data.

REFERENCES


UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN