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 exchangability 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,

$$(2.1) \qquad y_{ij} = x_{ij}^\top \beta + \alpha_i + u_{ij} \quad j = 1, ...m_i, \quad i = 1, ..., n,$$

which we will write in matrix form as,

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

1

Suppose 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,

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

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 $\min_{(\alpha,\beta)} \|y - X\beta - Z\alpha\|^2_{R^{-1}} + \|\alpha\|^2_{Q^{-1}}$, 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. 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 leat 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 $\|\alpha\|^2_{Q^{-1}}$, simply reflects differences in prior beliefs about the α'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 β.

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 α'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 τ-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

$$(2.2) \qquad Q_{y_{ij}}(\tau | x_{ij}) = \alpha_i + x_{ij}^\top \beta(\tau) \quad j = 1, ... m_i, \quad i = 1, ..., n.$$

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 (2.2) for several quantiles we propose solving,

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

where $\rho_{\tau_k}(u) = u(\tau - I(u < 0))$, as in Koenker and Bassett (1978). The weights w_k control the relative influence of the q quantiles on the estimation of the α_i parameters. Koenker (1984) considered a reversed situation in which only an intercept parameter was permitted to depend upon τ and the slope parameters associated with the included covariates were constrained to be identical for several τ'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 τ_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 α parameters. For the

quantile loss function, ρ_τ it is convenient to use the ℓ_1 penalty,

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

in place of the conventional Gaussian penalty. This choice maintains the linear pro-
gramming form of the problem and also preserves the sparsity of the resulting design
matrix. Several authors, notably Tibshirani (1996) and Donoho, Chen, and Saun-
ders (1998), have pointed out that ℓ_1 shrinkage offers some advantages over more
conventional Gaussian ℓ_2 penalties.

We will consider estimators solving,

$$(2.4) \qquad \min_{(\alpha,\beta)} \sum_{k=1}^{q} \sum_{j=1}^{n} \sum_{i=1}^{m_i} w_k \rho_{\tau_k}(y_{ij} - \alpha_i - x_{ij}^\top \beta(\tau_k)) + \lambda \sum_{i=1}^{n} |\alpha_i|.$$

For $\lambda \to 0$ we obtain the fixed effects estimator described above, while for $\lambda \to \infty$
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. Note that since x_{ij} is assumed to contain an intercept, in either case we
will also have q τ-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,

$$[X \vdots I_n \otimes e_m]$$

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,

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

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
$\tilde{y} = ((w \otimes y)^\top 0_n^\top)^\top$. 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. 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 $i = 1, ..., n$. Then since $Z = I_n \otimes e_m$, we have $Z^\top Z = mI_n$. 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 $i = 1, ..., n$. Let

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

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

where $\xi_{ij}(\tau_k) = \alpha_i + x_{ij}^\top \beta(\tau_k)$. 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 $k = 1, ..., q$, $j = 1, ..., m$, $i = 1, ..., n$.

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

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

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

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

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

$$\hat{\delta} \rightsquigarrow argmin V_0(\delta)$$

where

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

where B is a zero mean Gaussian vector with covariance matrix D_0, and $s = (s_0^\top 0_{pq}^\top)^\top$ and $s_0 = (sgn(\alpha_i))$.

Proof: The function V_{mn} can be decomposed into three parts using the identity of Knight (1998),

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau + \int_0^v (I(u \le s) - I(u \le 0))ds$$

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

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

where $v_{ijk} = z_{ij}^\top \delta_0 + x_{ij}^\top \delta_k / \sqrt{n}$ 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 $\Psi_k e_{mn} e_{mn}^\top \Psi_l = (\tau_k \wedge \tau_l - \tau_k \tau_l) I_{mn}$. Conditions A2-3 imply a Lindeberg condition and we have,

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

$$\rightsquigarrow -B\delta.$$

The second term is asymptotically quadratic in δ,

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

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

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

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

$$\rightarrow \frac{1}{2}\delta^\top D_1 \delta.$$

Finally,

$$V_{mn}^{(3)}(\delta) \;=\; \frac{\lambda_m}{\sqrt{m}} \sum_{i=1}^{n} \delta_{0i} \mathrm{sgn}(\alpha_i)$$
$$\to\; \lambda_0 \delta_0^\top s.$$

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

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

$$\hat{\delta} \rightsquigarrow \mathcal{N}(\lambda_0 D_1^{-1} s,\, D_1^{-1} D_0 D_1^{-1}).$$

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 α_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 $X^\top X/(mn)$ and $Z^\top Z/m$. We have seen that the latter is equal to I_n, and the former condition is again familiar. If $Z^\top X = 0$ 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 β's due to the shrinkage of the α'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 ℓ_1 and ℓ_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 α_i's iid from the χ_3^2 distribution, and u_{it} iid also from χ_3^2. In the left panel of Figure 1 we illustrate the estimated, $\hat{\alpha}_i$'s as a function of the regularization parameter λ. 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 ℓ_2 Gaussian penalty method. The ℓ_1 shrinkage method is more tolerant of large discrepancies; note that the gradient condition condition involves only the signs of the estimated effects, not
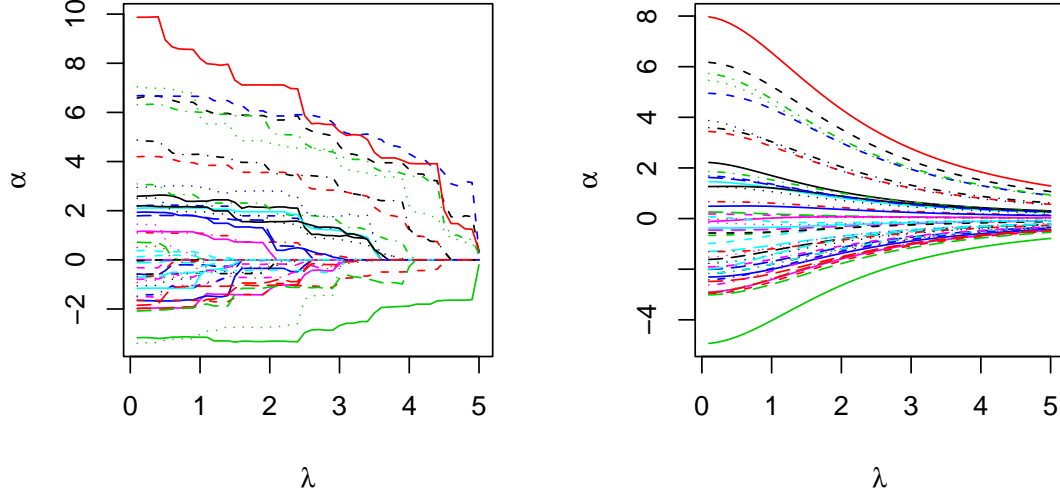
FIGURE 4.1. Shrinkage of the fixed effect parameter estimates, $\hat{\alpha}_i$. The left panel illustrates an example of the ℓ_1 shrinkage effect. The right panel illustrates an example of the ℓ_2 shrinkage effect.

their magnitudes, so highly unusual α_i's can be substantially shrunken toward zero without the extreme prejudice implied by the ℓ_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,

$$(4.1) \qquad\qquad y_{ij} = \alpha_i + x_{ij}\beta + u_{ij}$$

while in the latter case,

$$(4.2) \qquad\qquad y_{ij} = \alpha_i + x_{ij}\beta + (1 + x_{ij}\gamma)u_{ij}.$$

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) $$x_{ij} = \gamma_i + v_{ij}$$

with γ_i and v_{ij} independent and identically distributed over i and i, j respectively, then the interclass correlation coefficient,

$$\rho_x = \sigma_\gamma^2 / (\sigma_\gamma^2 + \sigma_u^2),$$

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 γ_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 α_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 χ_3^2. 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 α_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 λ chosen to be the ratio of scale parameters σ_u / σ_α.

	LS	PLS	LSFE	QR	PQR	QRFE
$\mathcal{N}$						
Bias	-0.0025	-0.0009	0.0005	-0.0023	-0.0022	0.0008
RMSE	0.0733	0.0529	0.0662	0.0813	0.0619	0.0716
t_3						
Bias	-0.0006	0.0037	0.0076	0.0057	0.0099	0.0077
RMSE	0.1648	0.1024	0.1168	0.1009	0.0715	0.0738
χ_3^2						
Bias	0.0016	0.0074	0.0125	0.0086	0.0001	0.0013
RMSE	0.1768	0.1426	0.1650	0.1853	0.1438	0.1410

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 χ_3^2 case is somewhat anomolous, since the penalized quantile regression esimator does slightly worse than the unpenalized fixed effects procedure, but both are competitive with the least squares results.

	LS	PLS	LSFE	QR	PQR	QRFE
$\mathcal{N}$						
Bias	-0.0049	-0.0032	-0.0009	-0.0039	-0.0024	0.0025
RMSE	0.0444	0.0477	0.0775	0.0439	0.0451	0.0656
t_3						
Bias	0.0009	0.0007	0.0006	0.0059	0.0066	0.0050
RMSE	0.0718	0.0767	0.1231	0.0536	0.0586	0.0777
χ_3^2						
Bias	0.0781	0.0805	0.0836	0.0136	0.0117	0.0001
RMSE	0.1394	0.1606	0.2257	0.0985	0.1063	0.1325

TABLE 4.2.

In the location-scale version of the model we adopt the same three distributions for generating the α_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 χ_3^2 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 χ_3^2 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 agressive shrinkage is helpful in these cases. The comparison in the χ_3^2 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, λ. 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

Chambers, J. M. (1998): *Programming with Data: A Guide to the S Language*. Springer.

Cole, T. J. (1988): "Fitting smoothed centile curves to reference data," *Journal of the Royal Statistical Society, Series A*, 151, 385–406.

Donoho, D., S. Chen, and M. Saunders (1998): "Atomic decomposition by basis pursuit," *SIAM J. of Scientific Computing*, 20, 33–61.

Goldberger, A. (1962): "Best linear unbiased prediction in the generalized linear regression model," *Journal of the American Statistical Association*, 57, 369–375.

Henderson, C. (1950): "Estimation of Genetic Parameters (Abstract)," *Annals of Mathematical Statistics*, 21, 309–310.

Ihaka, R., and R. Gentleman (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5, 299–314.

Knight, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *The Annals of Statistics*, 26, 755–770.

Knight, K., and W. Fu (2000): "Asymptotics for Lasso-type Estimators," *The Annals of Statistics*, 28, 1356–1378.

Koenker, R. (1984): "A Note on L-estimators for Linear Models," *Stat and Prob Letters*, 2, 323–325.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and P. Ng (2003): "SparseM: A Sparse Linear Algebra Package for R," *Journal of Statistical Software*, 8.

Lindley, D., and A. Smith (1972): "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society (B)*, 34, 1–41.

Rao, C. (1973): *Linear Statistical Inference and its Applications*. Wiley.

Robinson, G. (1991): "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–31.

Saunders, M. A. (1994): "Major Cholesky would feel proud," *ORSA J. on Computing*, 6, 23–27.

Scheffé, H. (1959): *Analysis of Variance*. Wiley.

Tibshirani, R. (1996): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

University of Illinois at Urbana-Champaign