An Interior Point Algorithm for
Nonlinear Quantile Regression

Roger Koenker[*]

and

Beum J. Park[*]

March 1994

Abstract

A new algorithm for computing quantile regression estimates for problems in which the response function is nonlinear in parameters is described. The nonlinear l_1 estimation problem is a special (median) case. The algorithm is closely related to recent developments on interior point methods for solving linear programs. Performance of the algorithm on a variety of test problems including the censored linear quantile regression problem of Powell (1986) is reported.

Keywords: Quantile Regression, Nonlinear Regression, Linear Programming, Interior Point Algorithms, Nonlinear Programming

Correspondence: Roger Koenker

1.   Introduction

About a century ago Edgeworth observed that methods of estimation based upon minimizing sums of *absolute* residuals could be far superior to least-squares methods under non-Gaussian error conditions. Laplace had drawn similar conclusions a century earlier. But computation of l_1-estimators, even for simple linear regression, remained a major impediment to applications until the emergence of the simplex algorithm for linear programming in the 1940's. Papers by Charnes, Cooper and Ferguson (1955) and Wagner (1959) provided a foundation for modern algorithms for linear l_1-regression by Barrodale and Roberts (1973), Bartels and Conn (1980) and others. These algorithms are readily extended to linear quantile regression, introduced in Koenker and Bassett (1978), of which l_1-regression is an important (median) special case. See Koenker and d'Orey (1987) for a description of a simplex-based, modified Barrodale-Roberts algorithm for linear quantile regression.

The current state of algorithms for *nonlinear* quantile regression is far less satisfactory. Certainly nothing comparable to the venerable Gauss-Newton algorithm for nonlinear least squares problems has emerged. Despite a flurry of interest by prominent numerical analysts in the 1970's and early 1980's, see, e.g., Osborne and Watson (1971), Murray and Overton (1981) and Bartels and Conn (1982), occasional applications of nonlinear quantile regression have relied on the Nelder and Mead (1965) algorithm and other generic optimization methods. An excellent statement of the current state-of-the-art is provided in the thesis of Busovaca (1985).

In contrast, the statistical theory of nonlinear quantile regression has developed rapidly in recent years. Powell (1986) has emphasized its value in the analysis of censored and truncated responses. Asymptotic theory for the case of serially independent errors has been developed by Oberhofer(1982), Dupačová(1987), Powell(1991), and Jurečková and Procházka(1993).

Theoretical developments by Weiss (1991) and White (1991) have stressed applications to time-series analysis. Applications of Horowitz and Neumann (1987), Chamberlain (1990), and others have demonstrated its value in applied econometrics. Jurečková and Procházka(1993) describe an interesting application in pharmacology

In this paper we will describe a new approach to the computation of nonlinear quantile regression estimators based on recent interior point methods for solving linear programs. In the next section we review interior point methods for strictly linear problems. Section 3 describes our approach to nonlinear problems, and Section 4 describes our computational experience.

2. Interior Point Methods for Linear Programs

In this section we provide a brief discussion of interior point methods for solving strictly linear programs including the linear quantile regression problem. Our exposition will follow closely that of Vanderbei, Meketon, and Freedman (1986) and Meketon (1986). For linear (in parameters) quantile regression the interior point method may be interpreted as iteratively reweighted least squares (IRLS). However, in sharp contrast to other attempts to compute l_1 regression estimates by IRLS, the interior point approach can be shown to converge to the correct answer. See Bassett and Koenker(1992) for a critique of some recent l_1 IRLS algorithms which fail to possess this property. We should emphasize that *in our experience* interior point algorithms for linear quantile regression do *not* appear to be competitive in efficiency terms with existing simplex method algorithms. However, unlike simplex based methods they do appear to offer a natural extension to nonlinear problems. Thus a clear understanding of the linear case is an essential first step in our exposition.

2.1. A Canonical LP

Consider the equality constrained linear program

$$\min \{ \mathbf{c}'\omega \,|\, \omega \in \Omega \equiv \{ \omega \in \mathbf{R}_+^n, \ \mathbf{A}\omega = \mathbf{b} \} \} \tag{2.1}$$

where $\mathbf{R}_+^n$ denotes the positive orthant of $\mathbf{R}^n$, A is an $m \times n$ matrix, and $\mathbf{b}$ is an m-vector. Given a feasible point in the interior of the constraint set, $\omega \in$ int (Ω), interior point methods proceed in two steps. First we transform coordinates to reposition ω so it is *centered* relative to the set Ω. Then a (projected) gradient step is taken toward the boundary of Ω. Repeating this process brings us arbitrarily close to a solution, and a stopping criterion is eventually invoked.

To flesh out this brief description, let $\mathbf{D}$ be a diagonal matrix with the elements ω on the diagonal and consider the transformation

$$\omega \to \mathbf{D}^{-1}\omega$$

We have $\mathbf{D}^{-1}\omega = \mathbf{1}_n$, an n-vector of ones, so the transformation $\mathbf{D}$ has the effect of centering ω relative to the orthant boundaries of Ω. Correspondingly, we may define $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}$ and $\tilde{\mathbf{c}} = \mathbf{D}\mathbf{c}$. In the transformed coordinates we wish to move in the gradient direction $-\tilde{\mathbf{c}}$, but to preserve feasibility we should instead project $\tilde{\mathbf{c}}$ onto the null space of $\tilde{\mathbf{A}}$ to insure that the equality constraints are satisfied.

Let $\hat{\mathbf{c}}$ denote this projection, i.e.,

$$\hat{\mathbf{c}} = (\mathbf{I}_n - \tilde{\mathbf{A}}'(\tilde{\mathbf{A}}'\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}})\tilde{\mathbf{c}} \tag{2.2}$$

Clearly, $\hat{\mathbf{c}}$ is a direction of descent; and we now move toward the boundary of Ω in this direction. Let

$$\hat{\alpha} = \max_{i=1,\ldots,n} \{ \mathbf{e}_i'\hat{\mathbf{c}} \}$$

where $\mathbf{e}_i$ is the ith unit basis vector for $\mathbf{R}^n$. For some fixed $\eta \in (0, 1)$, consider

$$\omega - (\eta/\alpha)\mathbf{D}\hat{\mathbf{c}} \to \omega$$

which defines a sequence of iterations $\omega_{k+1} = T(\omega_k)$. Since at each iteration

$$\mathbf{c}'\omega_{k+1} = \mathbf{c}'\omega_k - (\eta/\alpha)\mathbf{c}'\mathbf{D}\hat{\mathbf{c}} = \mathbf{c}'\omega_k - (\eta/\alpha)\|\hat{\mathbf{c}}\|^2, \tag{2.3}$$

we expect to see an improvement in the objective function at each iteration as long as $\alpha > 0$. The parameter η, which, following Meketon, we take as .97 in our implementation of the algorithm, insures that the updated ω is feasible. This would be true, of course, as long as $\eta \in (0, 1)$.

Proposition. If $\hat{\mathbf{c}} \leq 0$ the problem (2.1) is unbounded, unless $\hat{\mathbf{c}} = 0$ in which case every $\omega \in \Omega$ is optimal. Otherwise, the problem is bounded and the sequence $\{\mathbf{c}'\omega_k\}$ is strictly decreasing, hence convergent.

Proof. Since the proof of this proposition, found in Vanderbei, Meketon and Freedman (1986), is both elementary and revealing we repeat it here for the sake of completeness. If $\hat{\mathbf{c}} = 0$, there exists a vector z such that $\tilde{\mathbf{c}} = \tilde{\mathbf{A}}'z$, hence $\mathbf{D}\tilde{\mathbf{c}} = \mathbf{D}\tilde{\mathbf{A}}'z$ and since $\omega \in int(\Omega)$ it follows that $\mathbf{c} = Az$. But then for any $\omega \in \Omega$,

$$c'\omega = z'\mathbf{A}\omega = z'\mathbf{b}$$

which is independent of ω, establishing that $\mathbf{c}'\omega$ is constant on all of Ω. Next consider $\hat{\mathbf{c}} \leq 0$. Note that

$$\mathbf{c}'\omega_1 \equiv \mathbf{c}'\omega - \gamma(\omega)\tilde{\mathbf{c}}'\hat{\mathbf{c}} = \mathbf{c}'\omega - \gamma(\omega)\|\hat{\mathbf{c}}\|^2 \tag{2.4}$$

where $\gamma(\omega) = \eta/\alpha$, The dependence on ω is obviously through α. Since $\hat{\mathbf{c}} \leq 0$,

$$\omega_\rho = \omega - \rho\mathbf{D}\hat{\mathbf{c}}$$

is feasible for any $\rho > 0$ and

$$\mathbf{c}'\omega_\rho = \mathbf{c}'\omega - \rho\|\hat{\mathbf{c}}\|^2$$

implies that $\mathbf{c}'\omega_\rho \to -\infty$ as $\rho \to \infty$. Finally, if $\hat{\mathbf{c}} \geq 0$, then $\gamma(\omega) > 0$, so $\mathbf{c}'\omega_1 < \mathbf{c}'\omega$ follows from (2.4), establishing that the step is a direction of descent. ∎

2.2. Linear l_1-regression

In the linear model

$$y_i = \mathbf{x}_i'\beta + u_i \qquad i = 1, \cdots, n,$$

as noted in the introduction, the l_1-estimator of β which minimizes

$$R(\mathbf{b}) = \sum_{i=1}^{n} |y_i - \mathbf{x}_i'\mathbf{b}|.$$

may be formulated as a linear program. The primal form of the l_1 linear program may be written as

$$\min\{\mathbf{1}_n'\mathbf{u}^+ + \mathbf{1}_n'\mathbf{u}^- \mid (\mathbf{b}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbf{R}^p \times \mathbf{R}_+^{2n}, \mathbf{Xb} + \mathbf{u}^+ - \mathbf{u}^- = \mathbf{y}, \}$$

where $\mathbf{y}$ is the n-vector of responses, $\mathbf{X}$ is the $n \times p$ design matrix and $\mathbf{1}_n$ is an n-vector of 1's.

Having distinguished the positive and negative parts of the residual vector, we are simply minimizing a linear function subject to linear constraints. The dual problem may be written as

$$\max \{\mathbf{y}'\mathbf{d} \mid \mathbf{d} \in \Omega = \{ \mathbf{d} \in [-1, 1]^n, X'\mathbf{d} = 0 \} \}.$$

The dual variables, $\mathbf{d}$, may be viewed as Lagrange multipliers on the constraints, i.e., marginal costs of relaxing the constraints. If u_i is nonzero, then $d_i = sgn(u_i)$; otherwise, when $u_i=0$, $d_i \in (-1, 1)$. By complementary slackness there will be, barring degeneracy in the primal problem, exactly p of the u_i's equal zero at an optimum and consequently p of the d_i not equal to ± 1. To solve the dual problem we proceed as before, except that the centering is slightly altered to accommodate the altered form of Ω. For any initial feasible point $\mathbf{d}$, e.g., $\mathbf{d} = 0$, following Meketon (1986), set

$$\mathbf{D} = \text{diag} (\min \{1+d_i, 1-d_i\}).$$

In the transformed coordinates $\mathbf{D}^{-1}\mathbf{d}$ the projected gradient is

$$\mathbf{D}\hat{\mathbf{u}} = (\mathbf{I}_n - \mathbf{DX}(\mathbf{X}'\mathbf{D}^2\mathbf{X})^{-1}\mathbf{X}'\mathbf{D})\mathbf{D}y = \mathbf{D}(\mathbf{y} - \mathbf{Xb})$$

where $\mathbf{b} = (\mathbf{X}'\mathbf{D}^2\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}^2\mathbf{y}$. Note that, as in the former case, the transformation has the effect

of centering the point $\mathbf{d}$ in the feasible set Ω. Now let

$$\alpha = \max_{i=1,\ldots,n} \{\max \{\frac{\mathbf{e}_i{}'\mathbf{D}^2\hat{\mathbf{u}}}{1 + d_i} , \frac{-\mathbf{e}_i{}'\mathbf{D}^2\hat{u}}{1 - d_i} \} \}$$

and again for $\eta \in (0, 1)$ to assure dual feasibility we take the step

$$\mathbf{d} \leftarrow \mathbf{d} + (\eta/\alpha)\mathbf{D}^2\hat{\mathbf{u}}.$$

Note the change in sign since we are now maximizing. The iteration sequence $\mathbf{d}_{k+1} = T(\mathbf{d}_k)$ in the dual vector implicitly defines a corresponding primal sequence with

$$\mathbf{b}_k = (\mathbf{X}'\mathbf{D}_k^2\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_k^2\mathbf{y}.$$

As Meketon notes, the duality theory yields a natural stopping criterion. Since

$$\mathbf{y}'\mathbf{d}_k \leq \sum |y_i - \mathbf{x}_i{}'\mathbf{b}_k|$$

with optimality if and only if equality holds, it is reasonable to stop iterating when the difference between the dual and primal values is less than a specified tolerance. It is well known that the l_1 solution need not be unique, so it is worthwhile to recall that degeneracy of the dual solution implies multiple optimal solutions to the primal and vice versa. Both problems are most easily resolved by small random perturbations of the design matrix and response vector.

2.3 Linear Quantile Regression

If we replace the (symmetric) l_1-criterion with an asymmetric linear criterion so we minimize

$$R_\theta(\mathbf{b}) = \sum_{i=1}^{n}\rho_\theta(y_i - \mathbf{x}_i\mathbf{b})$$

$\rho_\theta(u) = u(\theta - I(u < 0))$, we obtain the regression quantiles of Koenker and Bassett (1978). The dual problem is now,

$$\max \{\mathbf{y}'\mathbf{d} | \mathbf{d} \in \Omega = \{\mathbf{d} \in [\theta{-}1, \theta]^n, \ \mathbf{X}'\mathbf{d} = 0 \} \}$$

This leads to an algorithm identical to the l_1 special case except that now

$$\mathbf{D} = \text{diag } (\min(\theta - d_i, \ 1 - \theta + d_i))$$

and

$$\alpha = \max_{i = 1, \cdots, n} (\max\left[\frac{\mathbf{e}_i' \mathbf{D}^2 \hat{\mathbf{u}}}{\theta - d_i}, \ \frac{-\mathbf{e}_i' \mathbf{D}^2 \hat{\mathbf{u}}}{1 - \theta + d_i} \right]).$$

The dual vector, $\mathbf{d}$, plays an important *statistical* role in the theory of linear quantile regression. Gutenbrunner and Jurečková (1992), generalizing the rank score process of Hájek and Šidák (1967), introduce the regression rank score process

$$\hat{\mathbf{a}}(\theta) = \underset{\mathbf{a} \in [0, \ 1]^n}{\text{argmax}} \{\mathbf{y}' \mathbf{a} \,|\, \mathbf{X}' \mathbf{a} = (1 - \theta) \mathbf{1}_n \}$$

which is obviously just a translation of the dual vector $\mathbf{d}$ introduced above, $\hat{\mathbf{a}}(\theta) = \hat{\mathbf{d}}(\theta) + (\theta - 1) \mathbf{1}_n$. Hájek and Šidák's rank score process for the location model, $\mathbf{X} \equiv \mathbf{1}_n$, takes the simple form,

$$\hat{a}_i(\theta) = \begin{cases} 1 & \text{if } \theta \le (R_i - 1)/n \\ R_i - \theta n & \text{otherwise} \\ 0 & \text{if } \theta > R_i/n \end{cases}$$

where R_i is the rank of y_i among $y_1, \ \cdots, \ y_n$.

Gutenbrunner and Jurečková (1992) suggest an elegant new approach to estimation in the linear model based on the regression rank score process. They construct weights

$$w_i = \int_0^1 \hat{a}_i(\theta) d\phi(\theta)$$

for appropriately chosen score function ϕ and consider weighted least squares estimators of the form

$$\hat{\beta} = (\mathbf{X}' W \mathbf{X})^{-1} \mathbf{X}' W \mathbf{y}$$

where $W = \text{diag } (w_i)$. For the simple "trimming-ϕ." $\phi_\alpha(\theta) \equiv (1 - 2\alpha)^{-1} I(\alpha \le \theta \le 1 - \alpha)$ we have, integrating by parts, $w_i = \hat{a}_i(1 - \alpha) - \hat{a}_i(\alpha)$ which takes the value 1 if observation y_i is above the αth regression quantile plane and below the $(1 - \alpha)^{\text{th}}$ plane, and takes the value 0 if y_i

is

above both, or below both. If y_i lies *on* either the αth or $(1 - \alpha)^{th}$ regression quantile plane, w_i takes a value strictly between 0 and 1 . This resolves the ambiguity implicit in early proposals by Koenker and Bassett (1978) and Ruppert and Carroll (1980) for similar trimmed least squares estimators. The regression rank score process also provides a natural means of generalizing linear rank tests to the linear regression model. See Gutenbrunner, Jurečková, Koenker and Portnoy (1993) for details.

3.   Nonlinear Quantile Regression

To extend these ideas to the case of nonlinear response functions we begin by considering the nonlinear l_1 problem

$$\min_{t \in \mathbf{R}^p} \ \sum |f_i(\mathbf{t})| \tag{3.1}$$

where, for example,

$$f_i(\mathbf{t}) = y_i - f_0(x_i, \mathbf{t}).$$

As noted by El Attar, *et al* (1979) a necessary condition for t^* to solve (3.1) is that there exists a vector $\mathbf{d} \in [-1, 1]^n$ such that

$$\mathbf{J}(\mathbf{t}^*)'\mathbf{d} = 0 \tag{3.2}$$

$$\mathbf{f}(\mathbf{t}^*)'\mathbf{d} = \sum |f_i(\mathbf{t}^*)| \tag{3.3}$$

where $\mathbf{f}(\mathbf{t}) = (f_i(\mathbf{t}))$ and $\mathbf{J}(\mathbf{t}) = (\partial f_i(\mathbf{t})/\partial t_j)$

Thus, as proposed by Osborne and Watson (1971), one approach to solving (3.1) is to solve a succession of linearized l_1 problems minimizing

$$\sum |f_i(\mathbf{t}) - \mathbf{J}_i(\mathbf{t})'\delta|,$$

choosing a step length, λ, at each iteration, by line search in the resulting directions δ. The difficulty, as we see it, with this approach is twofold. First, it requires us to fully solve an l_1

problem at each iteration. Secondly, and perhaps more significantly, the resulting search directions may actually be inferior to directions determined by incomplete solutions to the sequence of linearized problems. Indeed the information contained in the initial steps of the dual iteration seems better suited to solving the nonlinear problem since they embody the relevant local information at the current value of $\mathbf{t}$. Further iteration is only effective when the problem "nearly linear."

Let t be the value of the parameter at the current iteration, and consider the dual problem

$$\max \{\mathbf{f}'\mathbf{d} \mid \mathbf{d} \in [-1,\ 1]^n,\ \ \mathbf{J}'\mathbf{d} = 0\}. \tag{3.4}$$

If the model were linear so

$$\mathbf{f(s)} = \mathbf{f(t)} - K(\mathbf{s-t})$$

for some fixed matrix K, then a solution could be found by applying Meketon's algorithm to find $\mathbf{d}^*$ to solve (3.4), computing

$$\delta^* = (K'\mathbf{D}^2 K)^{-1} K'\mathbf{D}^2 \mathbf{f}.$$

where $\mathbf{D} = \text{diag} \ (\min \ \{1-d_i^*,\ 1+d_i^*\ \})$ and setting $\mathbf{t}^* = \mathbf{t} + \delta^*$. When $\mathbf{f}$ is nonlinear there is no longer a compelling argument for fully solving (3.4) at each iteration, indeed, in our experience only a few iterations to refine the dual vector is preferable. In the version of the algorithm we have implemented to conduct the tests reported in the next section we typically take two dual steps between successive updates of $\mathbf{f}$ and $\mathbf{J}$. A detailed description of the algorithm is now provided.

3.1. Dual Step

For any feasible $\mathbf{d}$ in the interior of the constraint set of (3.4) we refine $\mathbf{d}$, following Meketon, as follows. Let

$$\mathbf{D} = \text{diag} \ (\min \ \{1-d_i, \ 1+d_i \ \})$$

$$\mathbf{s} = \mathbf{D}^2 (I - \mathbf{J}(\mathbf{J}'\mathbf{D}^2\mathbf{J})^{-1}\mathbf{J}'\mathbf{D}^2)\mathbf{f}.$$

$$\mathbf{d} \leftarrow \mathbf{d} + (\eta/\alpha)\mathbf{s}$$

where

$$\alpha = \max \ \{ \ \max \ \{s_i/(1 - d_i), \ -s_i/(1 + d_i) \ \} \ \}$$

and $\eta \in (0, 1)$ is the constant chosen to insure feasibility. Following Meketon, we use $\eta = .97$.

Updating $\mathbf{D}, \mathbf{s}$, and the new $\mathbf{d}$ continues the iteration. To this point the algorithm is exactly as in

Section 2.2, but now $\mathbf{f}$ and $\mathbf{J}$ depend upon the current value of $\mathbf{t}$ so the process is embedded in a

sequence of primal iterations in which we update $\mathbf{f}$ and $\mathbf{J}$ as follows.

3.2. Primal Step

The dual step yields the primal direction

$$\delta = (\mathbf{J}'\mathbf{D}^2\mathbf{J})^{-1}\mathbf{J}'\mathbf{D}^2\mathbf{f}$$

which we explore by line search. Our current implementation uses the S implementation of

Gay(1984) in the S function `nlminb`. Updating we have

$$\mathbf{t} \leftarrow \mathbf{t} + \lambda^*\delta$$

where $\lambda^* = \underset{\lambda}{\text{argmin}} \ \sum|f_i(\mathbf{t} + \lambda\delta)|$, and we then update $\mathbf{f}$ and $\mathbf{J}$. However before returning to the

dual step we must adjust the current $\mathbf{d}$ to ensure that it is feasible for the new value of $\mathbf{J}$. This is

accomplished, somewhat naively, by projecting the current $\mathbf{d}$ onto the null space of the new $\mathbf{J}$,

i.e. $\hat{\mathbf{d}} = (\mathbf{I}_n - \mathbf{J}(\mathbf{J}'\mathbf{J})^{-1}\mathbf{J}')\mathbf{d}$ and then shrinking it to insure that it lies in $[-1, 1]^n$, so

$$\mathbf{d} \leftarrow \hat{\mathbf{d}}/(\underset{i}{\max} \ \{ \ |\hat{d}_i| \ \} + \varepsilon)$$

for some tolerance parameter $\varepsilon > 0$. Obviously, when the problem is linear, so $\mathbf{J}$ is fixed, this

"adjustment" is nugatory since $\mathbf{d}$ is already in the null space of $\mathbf{J}$, and the algorithm is essentially

like Meketon's.

3.3. Stopping

The algorithm currently terminates when the new iterate fails to improve the objective function by a specified tolerance. In the implementation we use 10^{-7}. Exploration of alternative stopping rules is a topic for future research. Obviously, the tolerance here must exceed the smallest safely detectable value of $|(x - y)/x|$ as in `R1MACH(4)` of the PORT3 library, Fox (1984).

3.4 Related Literature

Gill, Murray, Saunders, Tomlin, and Wright (1986) and Bayer and Lagarias (1991) have recently pointed out the close relationship of "projected Newton barrier" methods (see Fiacco and McCormick (1965)) and interior point methods. Algorithms closely related to the one described above could presumably be formulated employing logarithmic barrier functions in the dual vector d.

3.5. Quantile Regression

As in the case of the linear problem the generalization of the l_1 problem to other quantiles is straightforward involving only a modification of the constraint set $[-1, 1]^n$ to $[\theta-1, \theta]^n$ for some $\theta \in (0, 1)$. θ appears only in the recentering of the d vector and the computation of α in the dual step described in Section 2.3 above. Obviously, the shrinkage of the dual vector to ensure dual feasibility described in Section 3.2 must also be appropriately modified.

4. Numerical Experience

In this section we describe our computational experience with a variety of test problems. To facilitate comparison with existing results in the literature we have chosen problems from

Busovaca(1985) and Wormersley(1986). We focus exclusively on the l_1 case since there are no comparable results in the literature for other quantiles. The problems used are described in detail in Appendix A. We have attempted to investigate all of the test problems reported in the published literature, however in a few cases we were unable find a complete description of the problem. The problem taken from Wormersley is included to explore the important special case of piecewise linear response functions which arise in Powell's(1986) formulation of the quantile regression problem for censored data.

All of the reported tests were carried out in S on a DEC station 5000. To implement a simple version of the Osborne and Watson(1971) algorithm in S we employed the S function `l1fit` which does l_1 regression using the Barrodale and Roberts(1973) algorithm. The S function `lsfit` carries out the corresponding weighted least squares computations for the interior point algorithm. The algorithm and the test problems are available on the internet; sending the email message `send index for S` to `statlib@stat.cmu.edu` will generate a reply which explains how to use this valuable resource.

A summary of our experience on the test problems appears in Tables 5.1 and 5.2.. For Wormersley's(1986) censored regression problem (Problem 1) our version of the interior point algorithm converges to Wormersley's reported solution. However, it should be noted that the solution to this problem is notoriously nonunique. Busovaca's algorithm, which requires second derivitives cannot be employed on Problem 1 due to the fact that the Hessian of the response function is identically zero almost everywhere. The remaining problems are all taken from Busovaca, and generally our interior point solutions correspond closely to his. In Problems 7 and 13 there are small discrepancies favoring Busovaca; in Problem 9 there is a larger discrepancy favoring the interior point method. Results for our implementation of the Osborne

and Watson algorithm are somewhat less satisfactory. It fails completely on Problems 11 and 12, performs poorly in Problems 1, 4b, and 13, but does slightly better than the interior point method on Problem 5. All three algorithms fail for Problem 4a which is highly degenerate at the specified initial point. At an alternative starting point, labeled Problem 4b, the interior point algorithm performs well. Finally, we observe that two Meketon steps per iteration generally perform well, however in the Watson problem two steps leads to an unsatisfactory solution, while one step per iteration performs nicely. Unfortunately, we do not have a good explanation for this.

While our primary objective at this (early) stage of research is getting the correct answer, a few remarks might be made on efficiency. In Table 5.1 we report the number of iterations required for each algorithm for all 14 test problems. Explicit timings are not meaningful since the three algorithms were not coded comparably. We may observe that the computational effort per iteration is approximately the same for our interior point algorithm and the Busovaca algorithm, while the Osborne and Watson algorithm requires the full solution to a linear l_1 problem at each iteration and therefore requires substantially more effort per iteration. Comparing iteration counts for our interior point method versus Busovaca we see that of the 11 problems for which both have solutions 4 favor the interior point methods, 6 favor Busovaca with one tie. In one case, Problem 7, the interior point algorithm stops due to exceeding the maximal iteration count.

5.   Some Concluding Remarks

We have described a simple approach to computing quantile regression estimates for problems with nonlinear response functions. The approach is based on recent developments on inte-

rior point methods for linear programming, but may be viewed as a variant of well-known iteratively reweighted least squares. While the algorithm seems to perform well on a variety of test problems, there is considerable room for improvement. Handling rank deficiency in the model Jacobian is critical. Alternative stopping criteria also should be explored.

References

Bard, Y., 1970, Comparison of gradient methods for the solution of nonlinear parameter estimation problems. *SIAM Journal of Numerical Analysis 7*, 157-186.

Barrodale, I and F.D.K. Roberts, 1974, Solution of an overdetermined system of equations in the l_1 norm, *Communications of the ACM, 17*, 319-320.

Bartels, R.H. and A.R. Conn, 1980, Linearly constrained discrete l1 problems, *ACM Transactions on Mathematical Software 6,* 594-608.

Bartels, R.H. and A.R. Conn, 1982, An approach to nonlinear l1 data fitting, in: A. Dold and B. Eckman, eds., *Lecture Notes in Mathematics and Numerical Analysis* (Springer-Verlag, New York), 48-58.

Bassett, G. and Koenker, R. 1992, A note on recent proposals for computing l_1 estimates, *Computational Statistics and Data Analysis 14,* 207-211.

Bayer, D.A. and J.C. Lagarias, 1991, Karmarkar's linear programming algorithm and Newton's method, *Mathematical Programming 50,* 291-330.

Beale, E.M.L., 1958, On an iterative method of finding a local minimum of a function of more than one variable. Technical Report No. 25, Statistical Techniques Research Group, Princeton University.

Becker, R.A., J.M. Chambers and A.R. Wilks, 1988, *The New S Language* (Wadsworth, Pacific Grove).

Biggs, M.C., 1971, Minimization algorithms making use of non-quadratic properties of the objective function. *Journal of Institutional Mathematical Applications 8*, 315-327.

Brown, K.M. and J.E. Dennis, 1971, New computational algorithms for minimizing a sum of squares of nonlinear functions. Report No. 71-6, Yale University, Department of Computer Science.

Busovaca, S., 1985, Handling degeneracy in a nonlinear l_1 algorithm, Technical Report CS-85-34, Dept. of Computer Science, Univ. of Waterloo.

Chamberlain, G., 1990, Quantile regression, censoring and the structure of wages, Invited paper presented at the 6th World Congress of the Econometric Society.

Charnes, A., W.W. Cooper and R.O. Ferguson, 1955, Optimal Estimation of executive compensation by linear programming, *Management Science 1,* 138-151.

Colville, A.R., 1968, A comparative study of nonlinear programming codes. Report 320-2949, IBM New York Scientific Center.

Dupačová, J., 1987, Asymptotic properties of restricted L_1-estimates of regression, in: Y.Dodge, eds., *Statistical Data Analysis Based on the L_1-norm and Related Methods,* (North-Holland, Amsterdam).

El-Attar, R.A., M. Vidyasagar and S.R.K. Dutta, 1979, An algorithm for l_1-norm minimization with application to nonlinear l_1-approximation, *SIAM Journal of Numerical Analysis 16*, 70-86.

Fiacco, A.V. and G.P. McCormick, 1965, *Nonlinear programming: Sequential unconstrained*

minimization techniques (John Wiley, New York).

Fox, P.A., 1984, *The PORT Mathematical Subroutine Library* (Bell Laboratories, Murray Hill, NJ).

Gay, D. M. (1984). A trust region approach to linearly constrained optimization. in *Numerical Analysis. Proceedings, Dundee 1983,* F. A Lootsma (ed.), Springer, Berlin, pp. 171-189.

Gill, P.E., W. Murray, M.A. Saunders, J.A. Tomlin and M.M. Wright, 1986, On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method, *Mathematical Programming 36,* 183-209.

Horowitz, J. and G. Neumann, 1987, Semiparametric estimation of employment duration models, *Econometric Reviews 6,* 5-46.

Jurečková, J. and B.Procházka, 1993, Regression quantiles and trimmed least squares estimators in nonlinear regression models, to appear in *Journal of Nonparametric Statistics.*

Koenker, R. and V. d'Orey, 1987, Computing Regression Quantiles, *Applied Statistics 36* , 383-390.

Koenker, R.W. and G.W. Bassett, 1978, Regression quantiles. *Econometrica 46,* 33-50.

Kowalik, J.S. and M.R. Osborne, 1968, *Methods for Unconstrained Optimization Problems.* Elsevier (North-Holland, New York).

Meketon, M., 1986, Least absolute value regression, AT&T Bell Laboratories Technical Report, Holmdel, NJ.

Murray, W and M. Overton, 1981, A projected Lagrangian algorithm for nonlinear l1 optimization, *SIAM Journal of Scientific and Statistical Computing 1,* 207-224.

Nelder, J. and R. Mead, 1965, A simplex method for function minimization, *Computer Journal 7,* 308-313.

Oberhofer, W., 1982, The consistency of nonlinear regression minimizing the L_1-norm, *Annals of Statistics 10,* 316-319.

Osborne, M.R., 1972, Some aspects of nonlinear least squares calculations, in: E.A. Lootsma, ed., *Numerical Methods for Nonlinear Optimization*, (Academic Press, New York) 171-189.

Osborne, M.R. and G.A. Watson, 1971, On an algorithm for discrete nonlinear l1 approximation, *Computer Journal 14,* 184-188.

Powell, J., 1991, Estimation of monotone regression models under quantile restrictions, in: W.A. Barnett, J. Powell, and G. Tauchen eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (Cambridge University Press, New York).

Powell, J., 1986, Censored regression quantiles, *Journal of Econometrics 32,* 143-155.

Powell, M.J.D., 1962, An iterative method for finding stationary values of a function of several variables. *Computer Journal 5*, 147-151.

Rosenbrock, H.H., 1960, An automatic method for finding the greatest or least value of a function. *Computer Journal 3*, 175-184.

Vanderbei, R.J., M.J. Meketon and B.A. Freedman, 1986, A modification of Karmarkar's linear programming algorithm, *Algorithmica 1,* 395-407.

Wagner, H.M., 1959, Linear programming techniques for regression analysis, *Journal of American Statistical Association 54,* 206-212.

Weiss, A., 1991, Estimating nonlinear dynamic models using least absolute error estimation, *Econometric Theory 7,* 46-68.

White H., 1991, Nonparametric estimation of conditional quantiles using neural networks, unpublished.

Wormersley, R.S., 1986, Censored discrete linear l_1 approximation. *SIAM Journal of Scientific & Statistical Computing 7*, 105-122.

Table 5.1
Algorithmic Performance on Several Test Problems

Example	Starting Point	Interior Point Algorithm		Osborne-Watson Algorithm		Busovaca Algorithm	
		Optimal Objective	Number of Iterations	Optimal Objective	Number of Iterations	Optimal Objective	Number of Iterations
1.Wormersley	(0,0)	3.032542	3	5.234825	3	NA	
2.Bard	(1,1,1)	0.1243383	7	0.1243383	5	0.1243406	13
3.Beale	(1,0.1)	0.0	6	0.1547611e-06	6	0.3695488e-05	8
4.a.Biggs	(1,1,1,1,1,1)	F		F		F	
b.Biggs	(1,8,2,2,2,2)	0.0	10	0.7289559	45	NA	
5.Brown&Dennis	(25,5,-5,-1)	905.528	30	903.2406	113	903.2343	2
6.El-Attar 5.1	(1,2)	0.4704242	8	0.4704267	6	0.4704247	8
7.El-Attar 5.2	(1,1,1)	7.914516	100	7.904731	22	7.894227	5
8.Madsen	(3,1)	1.0	14	1.000010	11	1.000002	13
9.Osborne 1	(0.5,1.5,-1,0.01,0.02)	0.0293912	11	0.0293914	10	0.8203727	55
10.Osborne 2	(1.3,0.65,0.65,0.7,0.6, 3,5,7,2,4.5,5.5)	2.570152	3	F		F	
11.Powell	(3,-1,0,1)	0.25423e-07	14	F		0.29039e-08	3
12.Rosenbrock	(-1.2,1)	0.0	13	F		0.506642e-06	51
13.Watson	(1,1,1,1)	0.623611	12	1.278432	5	0.6018584	24
14.Wood	(0,0,0,0)	0.0	7	0.000003	10	0.0	25

See Appendix A for a detailed description of the test problems. F indicates the algorithm failed to meet convergence criteria for the problem. NA indicates results are not available for this entry. All interior point results are based on two Meketon iterations per step with the exception of Problem 13 (Watson) where we use one.

Table 5.2
Optimal Points for Several Test Problems

Example	Interior Point Algorithm	Osborne-Watson Algorithm	Busovaca Algorithm
1.Wormersley	-6.7268, 4.58618	-7.29804, 4.74178	NA
2.Bard	0.10094, 1.52516, 1.97211	0.10094, 1.52516, 1.97211	0.10094, 1.52513, 1.97214
3.Beale	3, 0.5	3, 0.5	2.99999, 0.49999
4a.Biggs	F	F	F
b.Biggs	1, 10, 1, 5, 4, 3	1.82143, 81.94978, 2.27882	NA
5.Brown&Dennis	-9.77778, 11.67668, -0.48447 0.30207	-10.0227, 11.91354, -0.44026 0.55823	-10.2236, 11.90843, -0.45804 0.58032
6.El-Attar 5.1	2.84250, 1.92018	2.84250, 1.92018	2.84250, 1.92018
7.El-Attar 5.2	0.52829, -0.00212, 0.02391	0.53148, -0.00004, 0.02751	0.53606, 0.0, 0.00319
8.Madsen	0.0, 0.00035	0.0, 0.0022	0.0, -0.00205
9.Osborne 1	0.37706, 2.19246, -1.72552 0.01332, 0.02129	0.37706, 2.19246, -1.72552 0.01332, 0.02129	1.06716, 1.80257, -1.80731 0.00345, 0.00109
10.Osborne 2	1.108554 0.1561289 0.4774578 0.5395307 0.3549851 2.91116 1.749787 4.85514 2.344482 4.570197 5.635008	F	F
11.Powell	1.453e-05, -1.4532e-05 2.3251e-05, 2.3251e-05	F	0.5588e-08, -0.3725e-09 0.1250e-08, 0.1716e-08
12.Rosenbrock	1.0, 1.0	F	0.99999, 0.99999
13.Watson	-0.4225, 1.1747 -0.4564, 0.38409	-0.23584, 1.03241 -0.22747, 0.41384	-0.44271, 1.19321 -0.47676, 0.38449
14.Wood	1.0, 1.0, 1.0, 1.0	1.0, 1.0, 1.0, 1.0	1.0, 1.0, 1.0, 1.0

```
"nlrq"<-
function(model, t, k = 2, theta = .5, big=1e+20, nit.max = 100,
        eps = 1e-07, beta = 0.97)
{
#function to compute nonlinear rq estimate
#       t is the initial value of the unknown parameter
#       model is a user-provided function which returns components
#                f=(f_i (x_i , t)
#                J=(grad f_i )
#       theta is the desired quantile
#       k is the number of Meketon steps per iteration
#       eps and eta are tolerance parameters
#
#function returns
#       coef is the value of the parameter at the solution
#       obj is the value of the objective function at the solution
#       nit is the number of "Meketon steps" taken
        m <- model(t)
        n <- length(m$f)
        w <- rep(0, n)
        snew <- sum(rho.rq(m$f,theta))
        sold <- big
        nit <- 0
        while(sold - snew > eps & nit < nit.max) {
                z <- mek.rq(m$J, m$f, k, w, theta=theta, int = F,
                                  eps = eps, beta = beta)
                step <- z$coef
                t0 <- t
                l <- nlminb(start = 1, objective = model.step.rq, lower = 0,
                        upper = 1, model = model, t0 = t, theta=theta,
                        step = step)$parameters
                t <- t0 + l * step
                m <- model(t)
                sold <- snew
                snew <- sum(rho.rq(m$f,theta))
                w <- lsfit(m$J, z$w, int = F)$resid
                w1 <- max(pmax(w,0))
                if(w1>theta)
                        w <- w*theta/(w1 + eps)
                w0 <- max(pmax(-w,0))
                if(w0>1-theta)
                        w <- w*(1-theta)/(w0 + eps)
                print(c(t, l, sum(rho.rq(m$f,theta))))
                nit <- nit+1
        }
        return(coef=t,obj=snew,nit=nit)

}
"rho.rq"<-
function(u,theta){theta*pmax(u,0)+(theta - 1)*pmin(u,0)}
"mek.rq"<-
function(x, y, kmax = 1000, w, theta=.5, int = T, big=1e+20,
        eps = 1e-06, beta = 0.97)
{
        if(int == T)
                x <- cbind(1, x)
        yw <- big
        k <- 1
        while(k <= kmax & yw - crossprod(y, w) > eps) {
                d <- pmin(theta  - w, 1 - theta + w)
                z <- lsfit(x, y, d^2, int = F)
                yw <- sum(rho.rq(z$resid,theta))
                k <- k + 1
                s <- z$resid * d^2
```

```
                alpha <- max(eps, pmax(s/(theta  - w), -s/(1 - theta + w)))
                w <- w + (beta/alpha) * s
        }
        coef <- z$coef
        return(coef, w)
}
"model.step.rq"<-
function(lambda, t0, step, model, theta)
{
        sum(rho.rq(model(t0 + lambda * step)$f, theta))
}
```

Appendix A

Test problem 1 (Wormersley, 1986)

The results of temperature accelerated life tests on electrical insulation in 40 motorettes are recorded in Table A.1. This data is originally from Schmee and Hahn (1979). Ten motorettes were tested at each of four temperatures. Testing was terminated at different times at each temperature. The model used to fit the data is

$$log_{10} H = x_1 + \frac{1000 x_2}{(T + 273.2)} + \epsilon,$$

where H is the failure time and T is the temperature.

Table A.1 : Data for motorettes example

	Test temperature $T^o C$			
	150	170	190	200
Failure times H in hours		1764	408	408
		2772*	408	408
		3444	1344	504
		3542	1344	504
		3780	1440	504
		4860		
		5196		
Termination time $\bar{H}$	8064	5448	1680	528
	10 units	3 units	5 units	5 units

∗ Wormersley gives the second failure time at 170° as 2722, but his results are consistent with the value recorded here from Schmee and Hahn.

At each temperature there is an upper bound $\bar{H}$ (the time at which testing was stopped) on the observed failure times, so the logarithms of the observed failure times are given by

$$min\left(log_{10}\bar{H}, \ x_1 + \frac{1000x_2}{(T+273.2)} + \epsilon \right).$$

Test problem 2 (Bard, 1970)

$$f_i(x) = y_i - \left(x_1 + \frac{u_i}{v_i x_2 + w_i x_3} \right)$$

where $i = 1, 2, \cdots, 15$, $u_i = i$, $v_i = 16 - i$, $w_i = min(u_i, v_i)$, and

i	y_i	i	y_i	i	y_i
1	0.14	6	0.32	11	0.73
2	0.18	7	0.35	12	0.96
3	0.22	8	0.39	13	1.34
4	0.25	9	0.37	14	2.10
5	0.29	10	0.58	15	4.39

Test problem 3 (Beale, 1958)

$$f_i(x) = y_i - x_1(1 - x_2^i),$$

where $i = 1, 2, 3$, $y_1 = 1.5$, $y_2 = 2.25$ and $y_3 = 2.625$.

Test problem 4 (Biggs, 1971)

$$f_i(x) = x_3\, exp(-t_i x_1) - x_4 exp(-t_i x_2) + x_6 exp(-t_i x_5) - y_i,$$

where $i = 1, \cdots,\ 13$, $t_i = i/10$ and

$$y_i = exp(-t_i) - 5 exp(-10 t_i) + 3 exp(-4 t_i).$$

Test problem 5 (Brown and Dennis, 1971)

$$f_i(x) = (x_1 + t_i x_2 - exp(t_i))^2 + (x_3 + x_4 sin(t_i) - cos(t_i))^2,$$

where $i = 1, \cdots,\ 20$, and $t_i = i/5$.

Test problem 6 (El-Attar 5.1, 1979)

$$f_1(x) = x_1^2 + x_2 - 10$$
$$f_2(x) = x_1 + x_2^2 - 7$$
$$f_3(x) = x_1^2 - x_2^3 - 1$$

Test problem 7 (El-Attar 5.2)

$$f_1(x) = x_1^2 + x_2^2 + x_3^2 - 1$$
$$f_2(x) = x_1^2 + x_2^2 + (x_3 - 2)^2$$
$$f_3(x) = x_1 + x_2 + x_3 - 1$$
$$f_4(x) = x_1 + x_2 - x_3 + 1$$
$$f_5(x) = 2x_1^3 + 6x_2^2 + 2(5x_3 - x_1 + 1)^2$$
$$f_6(x) = x_1^2 - 9x_3$$

Test problem 8 (Madsen, see Overton and Murray, 1981)

$$f_1(x) = x_1^2 + x_2^2 + x_1 x_2$$

$$f_2(x) = sin(x_1)$$

$$f_3(x) = cos(x_2)$$

Test problem 9 (Osborne 1, 1972)

$$f_i(x) = y_i - (x_1 + x_2 \, exp(-t_i x_4) + x_3 \, exp(-t_i x_5))$$

where $i = 1, 2, \cdots, 33$, $t_i = 10(i - 1)$, and

i	y_i	i	y_i	i	y_i
1	0.844	12	0.718	23	0.478
2	0.908	13	0.685	24	0.467
3	0.932	14	0.658	25	0.457
4	0.936	15	0.628	26	0.448
5	0.925	16	0.603	27	0.438
6	0.908	17	0.580	28	0.431
7	0.881	18	0.558	29	0.424
8	0.850	19	0.538	30	0.420
9	0.818	20	0.522	31	0.414
10	0.784	21	0.506	32	0.411
11	0.751	22	0.490	33	0.406

Test problem 10 (Osborne 2)

$$f_i(x) = y_i - \big(x_1 \, exp(-t_i x_5) + x_2 \, exp(-(t_i - x_9)^2 x_6)$$
$$+ x_3 \, exp(-(t_i - x_{10})^2 x_7) + x_4 \, exp(-(t_i - x_{11})^2 x_8)\big)$$

where $i = 1, 2, \cdots, 65$, $t_i = (i-1)/10$, and

i	y_i	i	y_i	i	y_i
1	1.366	23	0.694	45	0.672
2	1.191	24	0.644	46	0.708
3	1.112	25	0.624	47	0.633
4	1.013	26	0.661	48	0.668
5	0.991	27	0.612	49	0.645
6	0.885	28	0.558	50	0.632
7	0.831	29	0.533	51	0.591
8	0.847	30	0.495	52	0.559
9	0.786	31	0.500	53	0.597
10	0.725	32	0.423	54	0.625
11	0.746	33	0.395	55	0.739
12	0.679	34	0.375	56	0.710
13	0.608	35	0.372	57	0.729
14	0.655	36	0.391	58	0.720
15	0.616	37	0.396	59	0.636
16	0.606	38	0.405	60	0.581
17	0.602	39	0.428	61	0.428
18	0.626	40	0.429	62	0.292
19	0.651	41	0.523	63	0.162
20	0.724	42	0.562	64	0.098
21	0.649	43	0.607	65	0.054
22	0.649	44	0.653		

Test problem 11 (Powell, 1962)

$$f_1(x) = x_1 + 10x_2$$

$$f_2(x) = 5^{1/2}(x_3 - x_4)$$

$$f_3(x) = (x_2 - 2x_3)^2$$

$$f_4(x) = 10^{1/2}(x_1 - x_4)^2$$

Test problem 12 (Rosenbrock, 1960)

$$f_1(x) = 10(x_2 - x_1^2)$$

$$f_2(x) = 1 - x_1$$

Test problem 13 (Watson, see Kowalik and Osborne, 1968)

$$f_i(x) = \sum_{j=2}^{n}(j-1)x_j t_i^{j-2} - \left(\sum_{j=1}^{n} x_j t_i^{j-1}\right)^2 - 1,$$

where $i = 1, \cdots, 29$, $t_i = i/29$, $f_{30}(x) = x_1$ and $f_{31}(x) = x_2 - x_1^2 - 1$.

Test problem 14 (Wood, see Colville, 1968)

$$f_1(x) = 10(x_2 - x_1^2)$$

$$f_2(x) = 1 - x_1$$

$$f_3(x) = 90^{1/2}(x_4 - x_3^2)$$

$$f_4(x) = 1 - x_3$$

$$f_5(x) = 10^{1/2}(x_2 + x_4 - 2)$$

$$f_6(x) = 10^{-1/2}(x_2 - x_4)$$