Bayesian Quantile Regression

Tony Lancaster[1] and Sung Jae Jun

Brown University

December 2005

1. **Introduction**:

Recent work by Schennach(2005) has opened the way to a Bayesian treatment of quantile regression. Her method, called Bayesian exponentially tilted empirical likelihood (BETEL), provides a likelihood for data y subject only to a set of m moment conditions of the form $Eg(y, \theta) = 0$ where θ is a k dimensional parameter of interest and k may be smaller, equal to or larger than m. The method may be thought of as construction of a likelihood supported on the n data points that is minimally informative, in the sense of maximum entropy, subject to the moment conditions. Specifically the probabilities $\{p_i\}$ attached to the n data points are chosen to solve

$$\max_p \Sigma_{i=1}^n - p_i \log p_i \text{ subject to } \Sigma_{i=1}^n p_i = 1 \text{ and } \Sigma_{i=1}^n p_i g(y_i, \theta) = 0 \tag{1}$$

The solutions of this problem, well known in the maximum entropy literature, e.g. Jaynes(2003, p.357), take the form

$$p_i(\theta) = \frac{\exp\{\lambda(\theta)' g(y_i, \theta)\}}{\Sigma_{j=1}^n \exp\{\lambda(\theta)' g(y_i, \theta)\}} \tag{2}$$

where m vector λ, dependent on θ, satisfies

$$\lambda(\theta) = \arg\min_\eta n^{-1} \Sigma_{i=1}^n \exp\{\eta' g(y_i, \theta)\}. \tag{3}$$

The $\{\lambda_i\}$ are the Lagrange multipliers corresponding to the m constraints in the problem (1). For every θ (3) is a convex minimization problem and computationally straightforward.

The resulting likelihood for i.i.d data is $\Pi_{i=1}^n p_i(\theta)$ and this may be combined with a prior density on θ to yield the posterior density

$$p(\theta|Y) = p(\theta) \Pi_{i=1}^n p_i(\theta) \tag{4}$$

on a support such that 0 is in the convex hull of the $g(y_i, \theta)$.

In this paper we explore the application of this method to the case where the moments $g(y_i, \theta)$ correspond to those of quantiles or quantile regression functions.

We first consider the quantiles of a random variable and give an explicit form for the posterior distributions and a comparison with the Bayesian bootstrap posterior. We then consider posterior distributions for quantile regression functions. Finally we consider the model studied by Chernozhukov and

[1]Correspondence to first author, Department of Economics, Brown University, Providence RI 02912; Tony_Lancaster@brown.edu

Hansen(2004) of quantile regression with endogenous regressors. We show that the Schennach approach leads to an alternative way to do inference about structural quantile functions than that proposed by these authors. For each class of models we consider how to construct marginal posterior densities.

2. **Quantiles**:

Consider first Bayesian inference about quantiles. The τ'th quantile θ_τ satisfies the moment restriction $E(g) = 0$ where

$$g = 1(y \leq \theta_\tau) - \tau$$

where $1(.)$ is the indicator function. Given a random sample of size n and for any θ in $[\min y, \max y)$ the Lagrange multiplier λ solving the problem (3) satisfies the equation

$$\Sigma_{i=1}^n g_i e^{\lambda g_i} = 0$$

with solution

$$\begin{aligned} e^{\lambda(\theta)} &= \frac{\tau n_0}{(1-\tau)n_1} \\ \text{where } n_1 &= \sharp(y_i \leq \theta_\tau) \text{ and } n_0 = n - n_1. \end{aligned}$$

Substituting this solution into the expression for the posterior density (4) and assuming a uniform prior gives

$$p(\theta_\tau|y) \propto \frac{\phi^{n_1}}{n_1^{n_1} n_0^{n_0}}, \quad \phi = \frac{\tau}{1-\tau}. \tag{5}$$

This is a piecewise constant density supported on $[\min(y_i), \max(y_i))$.

The density (5) may be sampled as follows. There are $n-1$ pieces, if the observations are all distinct, forming a partition of the interval from $\min(y_i)$ to $\max(y_i)$. Sample each piece according to its probabilitity. If the sampled piece is bounded by $y_{(j)}$ and $y_{(j+1)}$ sample a random variable uniformly distributed on this interval.

Figure 1 shows the posterior density of the median from a random sample of size $n = 40$ using a uniform prior. The step function (red) is (5) and the continuous curve (blue) is the posterior density corresponding to a double exponential density of the form

$$f(y|\theta) \propto \exp\{-|y-\theta|\}.$$

The vertical line shows the sample median. Both curves have been normalized to integrate to one. Posterior distributions are always piecewise constant, as in the figure. Highest posterior density intervals with (possibly approximate) 95% etc. probability content are straightforward to construct.

3. **Comparison with the Bayesian Bootstrap:**

The Bayesian bootstrap of Rubin(1981) takes the data to be iid multinomial with probabilities $\{p_i\}$. An improper Dirichlet prior on these probabilities leads

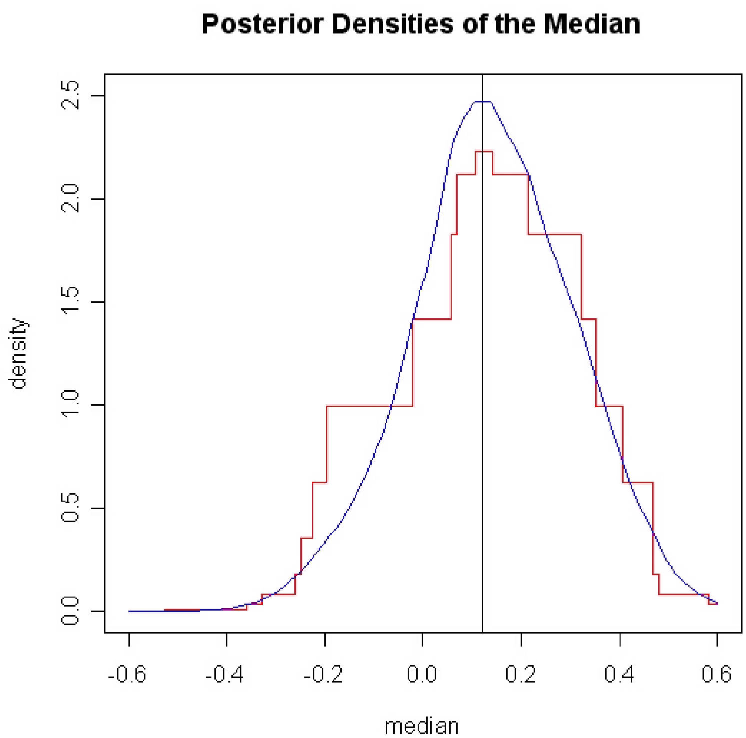
Posterior Densities of the Median
density
0.0
0.5
1.0
1.5
2.0
2.5
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
median

Figure 1:

to a Dirichlet posterior that assigns positive probability only to the distinct sample observations and in this respect is similar to BETEL. Paremeters such as quantiles that can be expressed as functionals of the data distribution have posterior distributions that can be calculated by repeatedly simulating from the posterior distribution of the $\{p_i\}$ and calculating the parameter of interest. As Chamberlain and Imbens(2003) point out, the posterior distribution of quantile regression parameters can be simulated by repeatedly solving the problem

$$\theta = \arg\min_t \Sigma_{i=1}^n c_\tau(V_i y_i - t' V_i x_i), \quad \{V_i\} \sim \text{iid} \ \ E(1). \tag{6}$$

Here c_τ is the check function defined by $c_\tau(u) = 2u(\tau - I(u < 0))$. This problem may be solved using Koenker's quantile regression R function rq(y˜x, tau, weights=..) by letting the weights be n unit exponentials.

To compare the Bayesian bootstrap and BETEL (uniform prior) posteriors we consider inference about the median – $\tau = 0.5$ – using a sample $n = 500$ standard normal variates. To get the BB posterior we solved the problem (6) 10,000 times and drew the histogram of the realizations. This is shown in figure (2). Note the sparsity of the Bayeian bootstrap distribution which reflects the fact that there were only 162 distinct realizations among the 10,000 draws even though the sample size was 500. This arises because the criterion function in (6) is a piecewise linear function with knots at the data points so solutions of the problem will always lie at one of the data points[2]. Hence there can be at most n points of support for the Bayesian bootstrap distribution and with $n = 500$ most of these will have probability so low that they will not occur in a sample of 10,000 realizations. By contrast, the BETEL distibution, shown in red, provides positive probability density over the relevant interval.[3]

It seems that the Bayesian bootstrap provides a less appealing posterior than Schennach's method in this application. This is in addition to to its other well known drawbacks.

4. **Regression Quantiles:**

The τ'th quantile regression is such that

$$\Pr(Y \leq \alpha(\tau) + \beta(\tau)X|X) = \tau$$

and so satisfies the moment conditions

$$\begin{aligned} E(1(Y &\leq \alpha(\tau) + \beta(\tau)X) - \tau|X) = 0 \\ E(X1(Y &\leq \alpha(\tau) + \beta(\tau)X) - \tau|X) = 0 \end{aligned}$$

If we now define

$$\begin{aligned} g_{1i} &= 1(y_i \leq \alpha + \beta x_i) - \tau \\ \text{and} \quad g_{2i} &= x_i(1(y_i \leq \alpha + \beta x_i) - \tau) \end{aligned}$$

[2] This assumes that if there is a flat section at the minimum the solution is chosen as one of the two data points between which the function is flat.

[3] These BB realizations were computed in R according to the program g <- rexp(n); g <- g/sum(g); coef(rq(y ˜1, weights = g)). The sample median was 0.078.

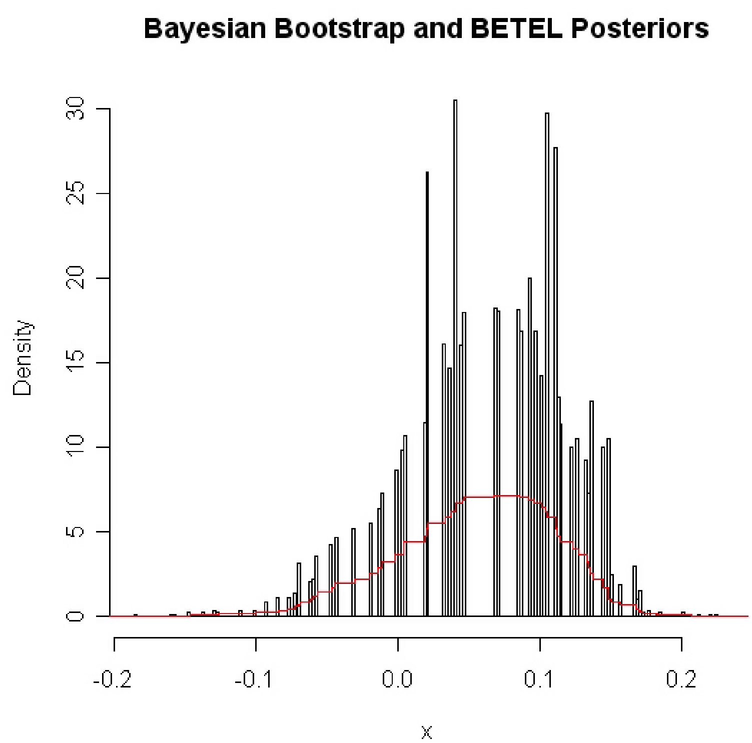
Bayesian Bootstrap and BETEL Posteriors
Density
0
5
10
15
20
25
30
-0.2
-0.1
0.0
0.1
0.2
x

Figure 2:

we may compute the Lagrange multipliers λ_1 and λ_2 by

$$\lambda = \arg\min_{\eta} \Sigma_{i=1}^{n} \exp\{\eta_1 g_{1i} + \eta_2 g_{2i}\}$$

and then calculate the posterior density according to (2).

Figure 3 plots the joint posterior density of $\alpha(0.5)$ and $\beta(0.5)$ from a sample of $n = 111$ observations and under a uniform prior. The data are Graddy's(1995) fish market observations with Y as log quantity traded and X as log price. To construct figure 3 the density was evaluated on a 100×100 grid of α, β values. As the figure shows the density consists of adjoining flat surfaces. The marginal densities of α and β may be calculated by summing this grid across rows or across columns. Figure (4) shows the marginal density of the price elasticity of demand at the 0.5 quantile found in this way. The quantile regression estimate of $\beta(0.5)$ using the R program previously mentioned was -0.41[4] which can be seen to be close to the marginal posterior mode.

4. **Quantiles With Endogenous Covariates.**

Quantile regression applied to the observations on price and quantity neglects the simultaneity of these variables when the market is in equilibrium. This can be surmounted by use of instrumental variables.

Recall that if the τ'th quantile is denoted $\alpha(\tau)$ then we can represent our random variable as

$$Y = \alpha(U), \quad U \sim \text{Uniform}(0,1), \quad \alpha(.) \text{ strictly increasing,}$$

since

$$\Pr(Y \leq \alpha(\tau)) = \Pr(\alpha(U) \leq \alpha(\tau)) = \Pr(U \leq \tau) = \tau.$$

and so $\alpha(\tau)$ is the quantile function.

Similarly, a regression quantile can be represented by

$$Y = \alpha(U) + \beta(U)X, \quad U|X \sim \text{Uniform}(0,1), \quad \alpha(\tau) + \beta(\tau)X \text{ strictly increasing in } \tau.$$

with τ'th conditional quantile equal to $\alpha(\tau) + \beta(\tau)X$.

Following Chernozhukov and Hansen(2004) consider the model

$$Y = D'\alpha(U) + X'\beta(U), \quad U|X, Z \sim \text{ Uniform}(0,1)$$

in which D is statistically dependent on U, $D'\alpha(\tau) + X'\beta(\tau)$ is strictly increasing in τ, and Z is a set of instrumental variables that are independent of U but statistically dependent on D. Then $D'\alpha(\tau) + X'\beta(\tau)$ is the τ'th quantile of Y conditional on X, Z. That is,

$$\Pr(Y \leq D'\alpha(\tau) + X'\beta(\tau)|X, Z) = \tau \quad (7)$$

The expression

$$D'\alpha(\tau) + X'\beta(\tau)$$

[4] This was computed using `rq(Q~P)` in R, where Q and P are logarithms of quantity and price.

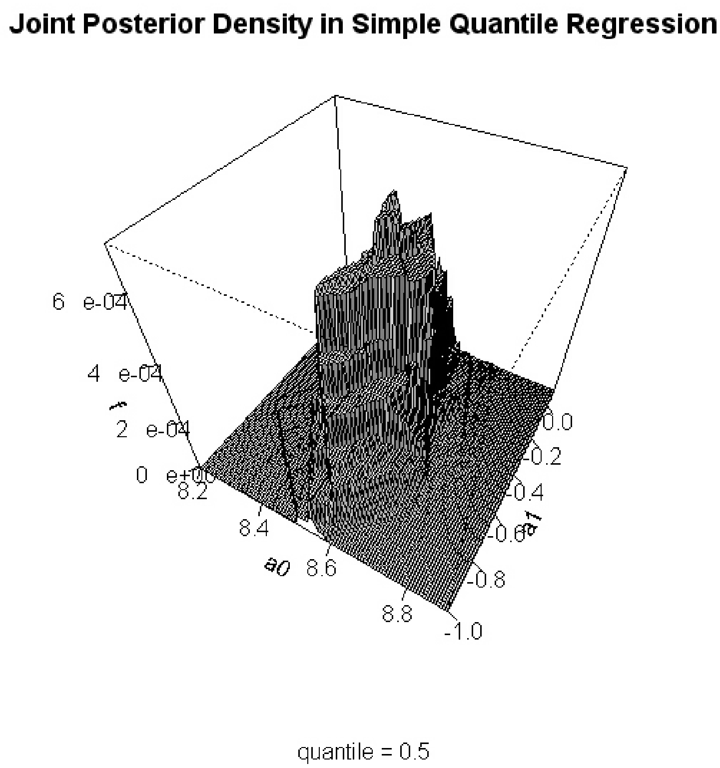
Joint Posterior Density in Simple Quantile Regression
6 e-04
4 e-04
2 e-04
0 e+00
f
8.2
8.4
8.6
8.8
a0
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
a1
quantile = 0.5

Figure 3:

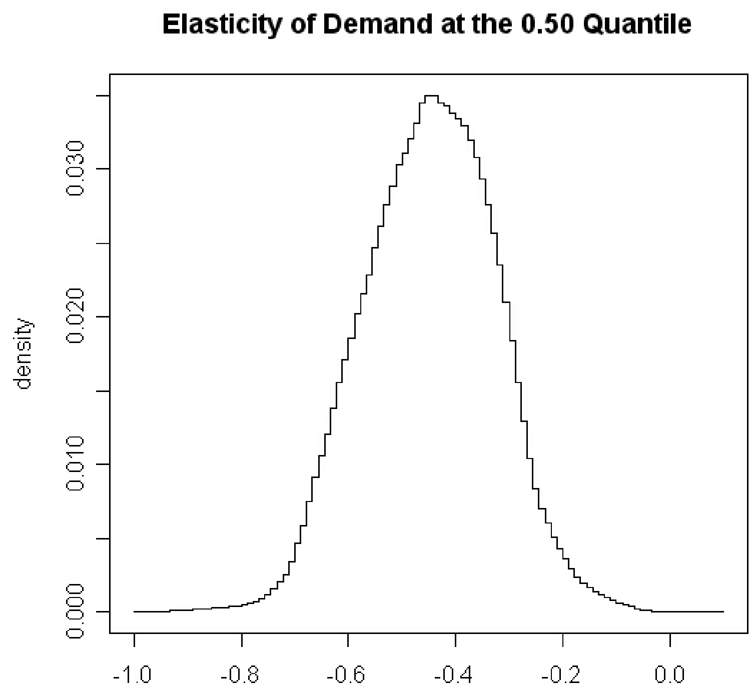
Elasticity of Demand at the 0.50 Quantile
density
0.000
0.010
0.020
0.030
-1.0
-0.8
-0.6
-0.4
-0.2
0.0

Figure 4:

is what Chernozhukov and Hansen refer to as the structural quantile function. The fact (6) then leads to conditional moments of which the simplest are of the form

$$\begin{aligned} X_i'(1(y_i &\leq D_i'\alpha(\tau) + X_i'\beta(\tau)) - \tau) \\ Z_i'(1(y_i &\leq D_i'\alpha(\tau) + X_i'\beta(\tau)) - \tau) \end{aligned}$$

Example 1: Simulated Data: In the following example the data were generated with X null and eight instruments which are the columns of Z. The Y data was generated by

$$\begin{aligned} y_i &= \alpha_0(U) + D_i\alpha_1(U), \\ D_i &= \gamma_0 + Z_i\gamma_1 + V_i, \\ \alpha_0(U), V &\sim N([0,0],[1,0.8,0.8,1] \end{aligned}$$

τ was set equal $0.25; n = 500$; and for the first experiment the eight elements of γ_1 were set equal to 1 and in the second they were set equal to 0.1 The latter choice was intended to represent weak instruments. Figures (5) and (6) show the joint posterior densities of $\alpha_0(0.25)$ and $\alpha_1(0.25)$, the slope and intercept of the strucrtural quantile function at the 0.25 quantile. Figure (4), with $\gamma_1 = 1$ shows a well behaved joint posterior density centered round the truth. Figure (5), with weaker instruments shows a thicker tailed distribution with apparent multiple modes. Further experiments not shown here show that as the coefficients on the instruments approach zero the joint density shows many modes.

Example 2: Demand for Fish: In this example we again use Graddy's data, also used by Chernozhukov and Hansen. Specifically, we use 111 observations on quantities of fish traded and their price. We also use observations on two weather variables which might be supposed to affect the supply of fish but not the demand. These are called "stormy" and "mixed".

The tau'th structural quantile function is

$$Q = \alpha_0(\tau) + \alpha_1(\tau)P$$

and this is estimated using three moment equations corresponding to stormy, mixed and the unit variable. Figure (6) shows the joint posterior density of $\alpha_0(0.5)$ and $\alpha_1(0.5)$. The density was evaluated on a 40×40 grid. It can be seen that there is only limited evidence of weak instruments in the suggestion of a secondary mode. The marginal densities can be found by summing over rows or columns and renormalising. The marginal density of the slope – elasticity of demand – is shown in figure (7). An approximate 95% highest posterior density interval runs from 0.1 to -2.5 and is marked in red. Chernozhukov and Hansen report a point estimate of -0.9 (marked in blue) using these same instruments with a 95% confidence interval running from 0 to -1.8. Note the minor mode in the marginal posterior density.

5. **Conclusions**

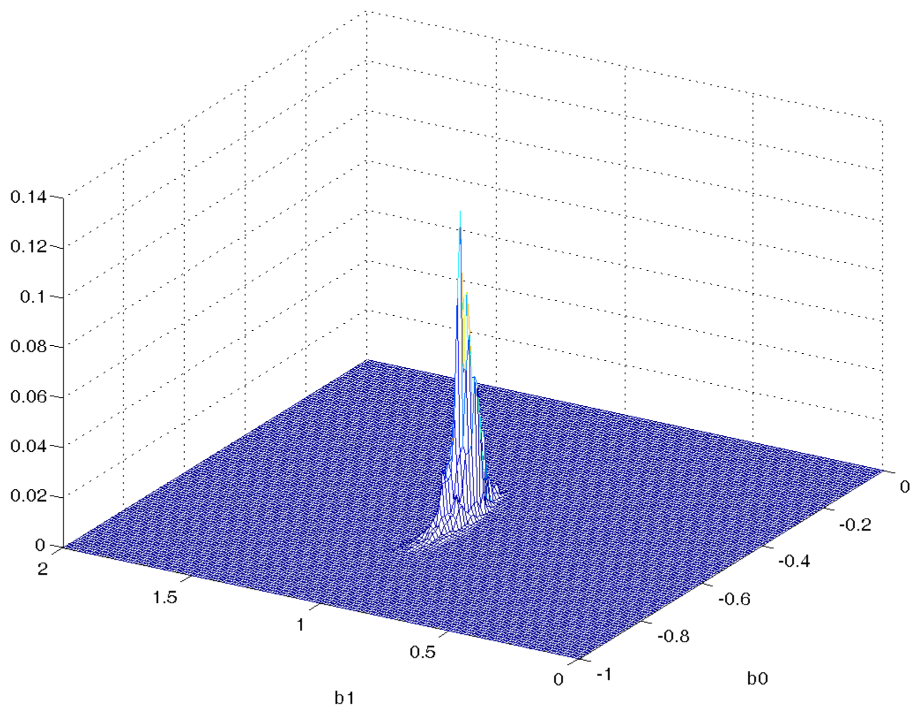

Figure 5:

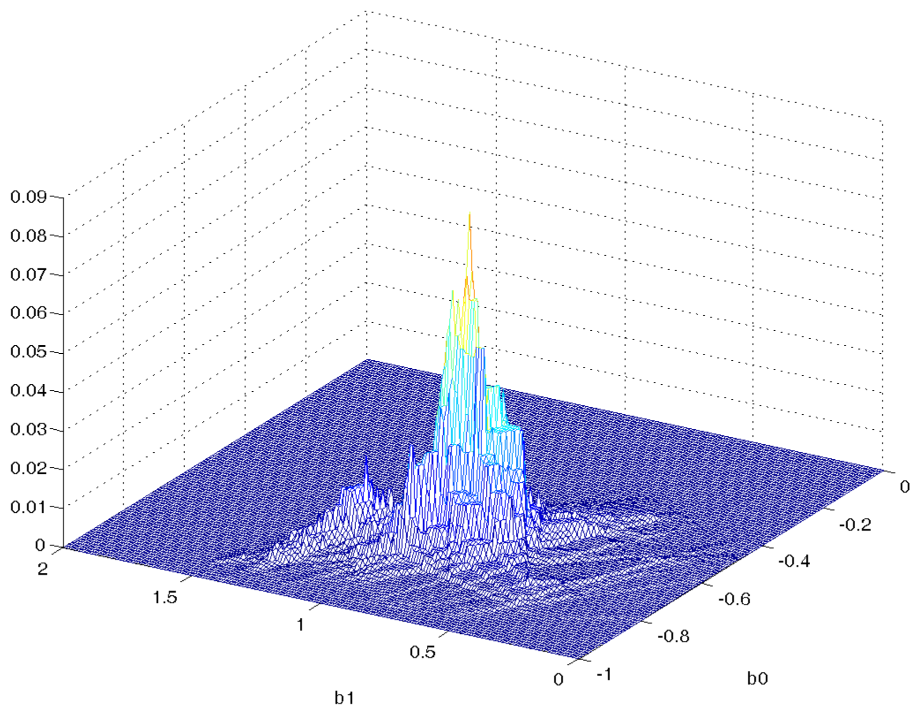

Figure 6:

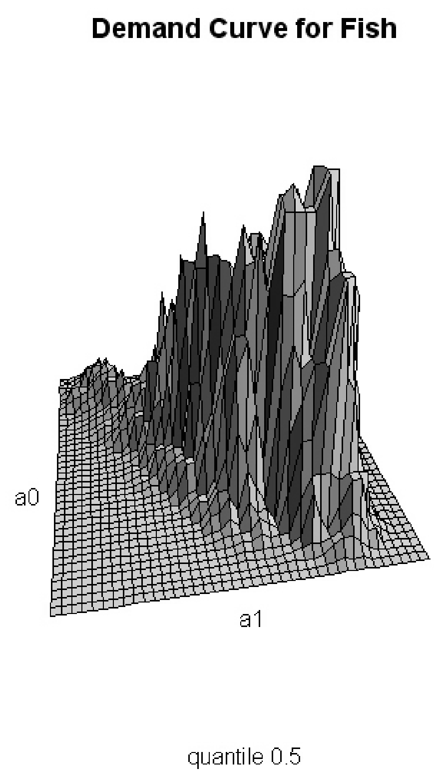

Figure 7:

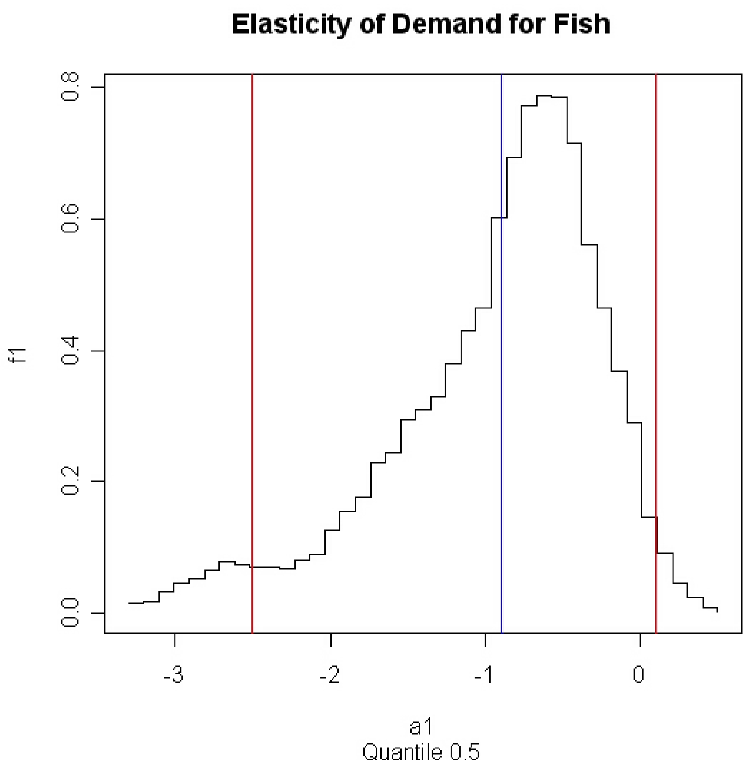
Elasticity of Demand for Fish
0.8
0.6
0.4
0.2
0.0
f1
-3
-2
-1
0
a1
Quantile 0.5

Figure 8:

Schennach's method represents a way of doing Bayesian inference about quantile regression models without using a potentially restrictive parametric likelihood. When the number of parameters of interest is small, as in the illustrations used in this paper, marginal posterior distributions can be easily evaluated by evaluating the joint posterior density on a grid, summing over the unwanted dimensions and renormalizing if necessary. When the dimensionality of the parameter of interest is larger it is likely to be preferable to use MCMC methods. In initial experiments not reported here we have used a Metropolis/Hasting sampler with a proposal distribution formed from an overdispersed version of the asymptotic multivariate normal posterior density. This appears to be satisfactory except when identification is very weak and the posterior density has multiple modes. Results on this point will be given in future versions of this paper.

References

Chamberlain, G. and G. W. Imbens *Nonparametric applications of Bayesian inference*, Journal of Business and Economic Statistics, 2003, 21, 1, 12-18.

Chernozhukov, V. and C. Hansen *Instrumental variable quantile regression*, unpublished ms, version of December 2004.

Graddy, K. *Testing for imperfect competition in the Fulton fish market*, Rand Journal of Economics, 1995 26(1),75-92.

Jaynes, E. T. *Probability theory:the logic of science* Cambridge University Press 2003.

Rubin, D. *The Bayesian bootstrap*, The Annals of Statistics, 1981, 9, 130-134.

Schennach, S. M. *Bayesian exponentially tilted empirical likelihood*, Biometrika 2005 92 (1), 31-46.