An approach to isotone regression using quantiles

Markus Vasquez

July 29, 2017

1 Introduction

We are interested in a model of the form

$$X_i = Q_i(U_i)$$

for $i = 1, \ldots, p$ with $U = (U_1, \ldots, U_p) \sim \mathcal{N}(0, \Sigma)$ for some Σ with ones along the diagonal. Also,

$$Y = R(\theta^T U + \delta) + \epsilon,$$

where we will assume that δ and ϵ are both mean-zero, normal and independent of everything else. We assume that the functions $Q_1, \ldots, Q_p, R$ are all unknown, strictly increasing continuous functions. The parameter θ is also unknown. We assume that θ and δ are normalized so that $\theta^T U_i + \delta$ is standard normal (this is without loss of generality since we only require that R be monotone.) Our goal is to estimate the regression function $S(X) = E[Y|X] = E[R(\theta^T U + \delta)|X]$ after making N observations $(X^{(n)}, Y^{(n)})$ for $n = 1, \ldots, N$.

Let Φ be the cumulative distribution function for the standard normal distribution. For any random variable A let F_A be the cdf of A.

First, since each Q_i is increasing, we know that

$$F_{X_i}(X_i) = F_{U_i}(U_i) = \Phi(U_i).$$

Solving for U_i we have

$$U_i = \Phi^{-1} F_{X_i}(X_i).$$

Let $\tilde{Y} = Y - \epsilon$. Then, using similar reasoning we find that

$$\theta^T U_i + \delta = \Phi^{-1} F_{\tilde{Y}}(\tilde{Y}).$$

If we knew F_{X_i} and $F_{\tilde{Y}}(\tilde{Y}^{(n)})$ for each n, we would then be left with a classic linear regression problem to find θ and $\mathrm{var}(\delta)$.

How can we estimate these quantities?

We can estimate F_{X_i} using either the empirical cdf or by using a kernel estimate. We can do whatever we think is best (probably use kernel estimators) to find an estimate $\hat{F}_{X_i}$.

For the second, we will use the approximation $F_{\tilde{Y}}(\tilde{Y}) \approx F_Y(Y)$. We can then use either the empirical cdf or a kernel estimate to find $\hat{F}_Y$.

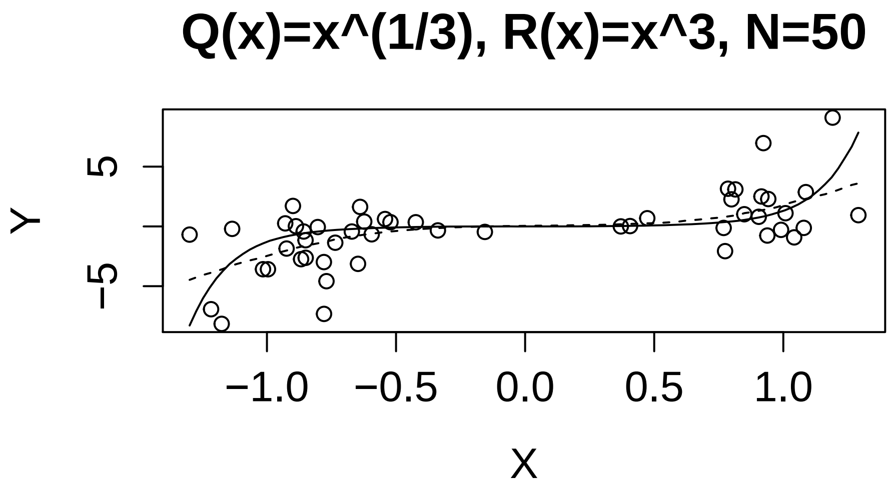
Then, letting $\hat{U}_i = \Phi^{-1}\hat{F}_{X_i}(X_i)$ and $\hat{W} = \Phi^{-1}\hat{F}_Y(Y)$, we can regress $\hat{W}$ on $\hat{U}$ to find estimates $\hat{\theta}$ for θ and $\hat{\sigma}_\delta$ for $\mathrm{sd}(\delta)$ (or even better, we can store the residuals for δ).

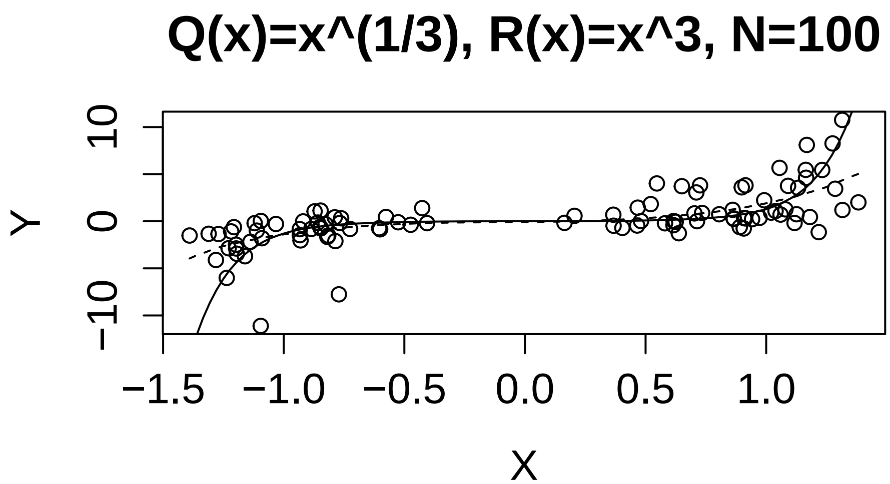
Putting this all together into our estimate for S we have

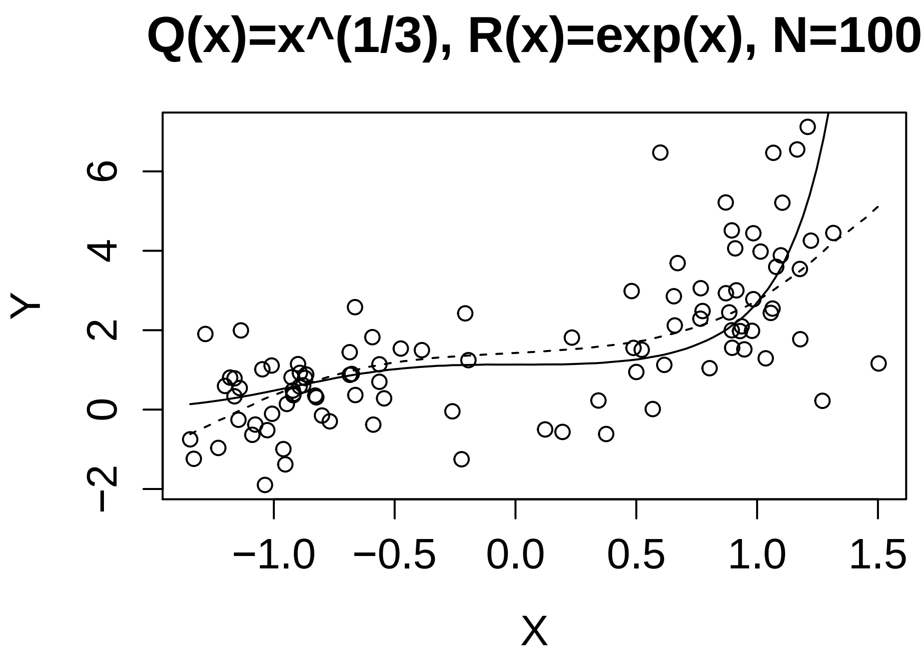
$$\hat{S}(X) = \frac{1}{N}\sum_{i=1}^{N} \hat{F}_Y^{-1}\Phi\left(\sum_{j=1}^{p}(\hat{\theta}_j \Phi^{-1}\hat{F}_{X_j}(X_j)) + \delta^{(i)}\right).$$

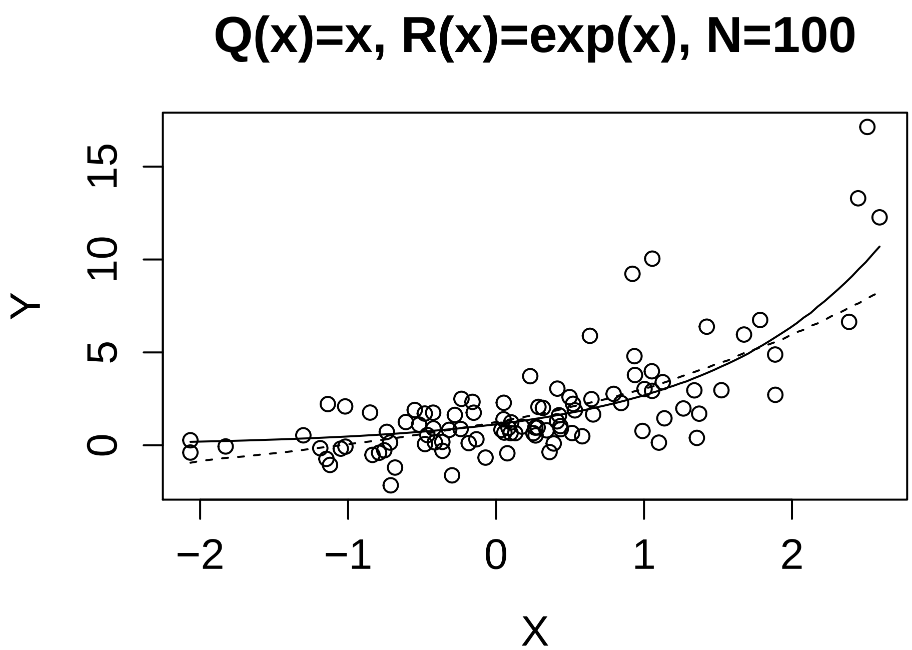
2 How does it do?

In each of the following, we are choosing the standard deviation of δ to be $1/2$ and $\theta = \sqrt{3}/2$. The circles are the observations. The solid line is the true regression function. The dashed line is our estimate. As we see, things tend to go better when Q is a contraction because the error introduced by δ tends not to be magnified.

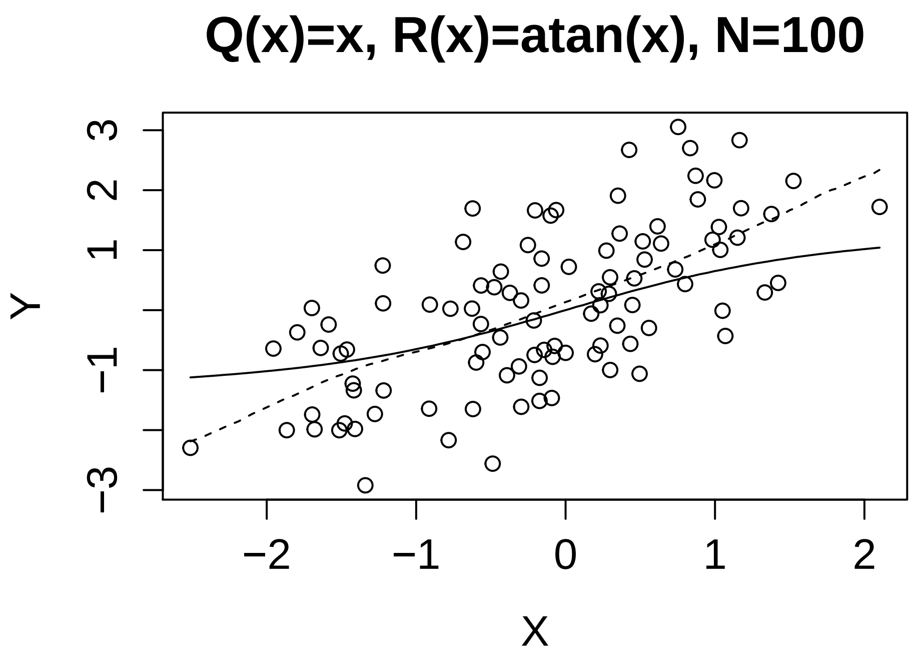

Q(x)=x^(1/3), R(x)=x^3, N=50
Y
−5
5
−1.0
−0.5
0.0
0.5
1.0
X

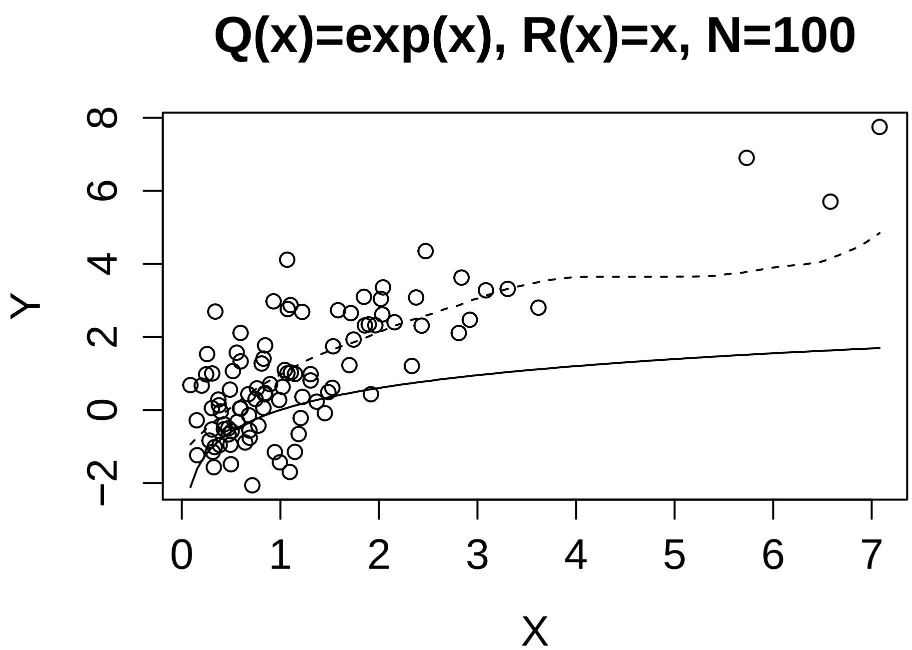

Q(x)=x^(1/3), R(x)=x^3, N=100
Y
−10
0
10
−1.5
−1.0
−0.5
0.0
0.5
1.0
X


Q(x)=x^(1/3), R(x)=exp(x), N=100
Y
−2
0
2
4
6
−1.0
−0.5
0.0
0.5
1.0
1.5
X

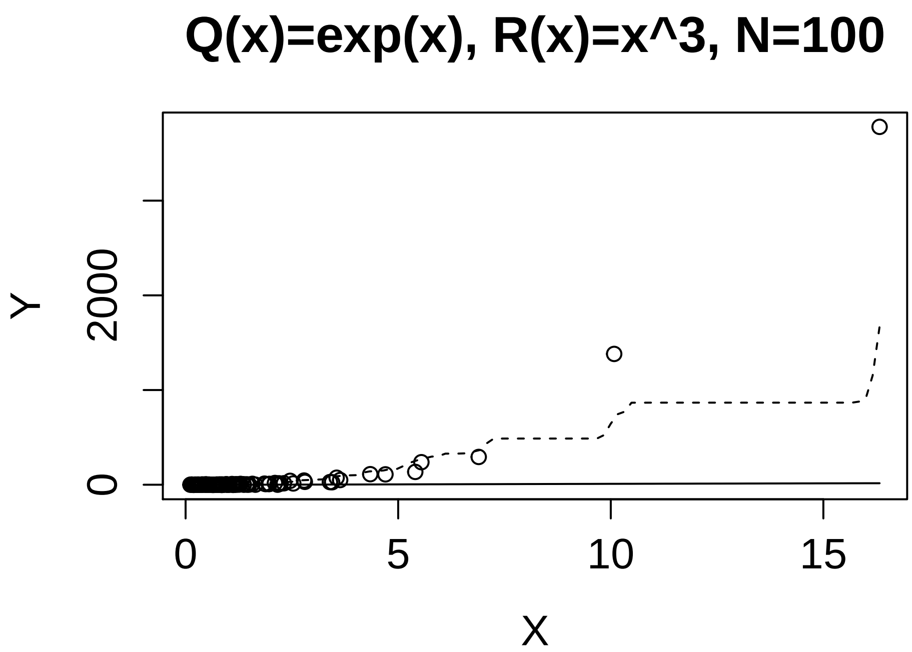


Q(x)=x, R(x)=exp(x), N=100
Y
15
10
5
0
−2
−1
0
1
2
X


Q(x)=x, R(x)=atan(x), N=100
Y
3
2
1
−1
−3
−2
−1
0
1
2
X



Q(x)=exp(x), R(x)=x, N=100
Y
−2
0
2
4
6
8
0
1
2
3
4
5
6
7
X



Q(x)=exp(x), R(x)=x^3, N=100
Y
0
2000
0
5
10
15
X