

# ON DISTRIBUTIONAL VS. QUANTILE REGRESSIONS

ROGER KOENKER

## 1. INTRODUCTION

Recent work by Firpo, Fortin, and Lemieux (2010) and Chernozhukov, Fernández-Val, and Melly (2010) has suggested that estimating families of binary response models using varying “cutoffs” to construct the binary response may provide an attractive alternative to estimating conditional quantile functions. In some very simple iid error settings the two approaches can be directly compared; we undertake such a comparison for two leading forms of the binary response model.

## 2. MODELS AND ESTIMATORS

Consider the linear model

$$(2.1) \quad Y_i = \alpha + x_i^\top \beta + u_i \quad i = 1, \dots, n,$$

with iid errors  $\{u_i\}$  with distribution function,  $F$ , survival function,  $S = 1 - F$ , and density,  $f$ , we can estimate models for the family of binary response models

$$\begin{aligned} S^{-1}(P(Y_i > y|x_i)) &= S^{-1}(P(u_i > y - \alpha - x_i^\top \beta_0)) \\ &= S^{-1}(1 - F(y - \alpha - x_i^\top \beta_0)) \\ &= y - \alpha - x_i^\top \beta_0. \end{aligned}$$

That is, we can define the indicator functions  $I(Y_i > y)$  for a particular choice of the cutoff,  $y$  and estimate the binary response model with link function,  $S$ . The effect of changing the cutoff,  $y$ , is simply to shift the intercept of the model. Estimating a family of such models for a variety of cutoffs is termed “distributional regression” in Chernozhukov, Fernández-Val, and Melly (2010).

It is well known, e.g. McCullagh and Nelder (1989), that the limiting distribution of this binary response estimator,  $\hat{\beta}_n(y)$ , is Gaussian with covariance matrix  $\hat{V}_n = (X^\top W X)^{-1}$  where  $W = \text{diag}(w_i)$ , and  $w_i = f^2(\nu(y))/(F(\nu(y))(1-F(\nu(y))))$  where  $\nu(y) = y - \alpha - x_i^\top \beta_0$ . For the corresponding quantile regression estimator,

$$\check{\beta}_n(\tau) = \text{argmax}_b \sum_{i=1}^n \rho_\tau(Y_i - x_i^\top b)$$

we have the limiting covariance matrix,  $\check{V}_n = (\tau(1-\tau)/f^2(F^{-1}(\tau)))(X^\top X)^{-1}$ . If we take  $y = F^{-1}(\tau)$  in the respective formulas we see that the two covariance matrices look quite similar. Indeed, if  $\beta_0 = 0$  then they actually coincide. However, when  $\beta_0 \neq 0$  the situation is a bit more complicated and we provide some numerical comparisons along with some simulation evidence on comparative performance in the next section.

## 3. A LOGISTIC SPECIFICATION

Consider the model (2.1) with  $\{u_i\}$  iid from the logistic  $df$ ,  $F(u) = (1 + e^{-u})^{-1}$ . This model gives rise to logistic binary response model

$$\begin{aligned} \text{logit}(P(Y_i > y|x_i)) &= \text{logit}(1 - F(y - \alpha - x_i^\top \beta)) \\ &= \text{logit}(F(\alpha - y + x_i^\top \beta)) \\ &= \alpha - y + x_i^\top \beta \end{aligned}$$

At each value of the cutoff,  $y$ , we have a logistic regression model with intercept  $\alpha - y$ . In survival analysis this model is often called the Bennett model. A valuable general discussion of the correspondence between survival models and families of binary response models is given in Doksum and Gasko (1990). In contrast we have the quantile regression model

$$Q_{y_i|x_i}(\tau|x_i) = \alpha + F_u^{-1}(\tau) + x_i^\top \beta$$

So both models have the same linear slope parameters, but the intercepts are different. However, for any choice of  $\tau$ 's we can choose corresponding  $y$ 's to be  $y_k = -F^{-1}(\tau_k)$ .

A simulation of this logistic model with  $n \in \{100, 500, 1000, 5000\}$ ,  $\{x_i\}$  iid  $\mathcal{N}(0, 1)$ ,  $\alpha = 0$ , and  $\beta = 1$  yields the results reported in Table 1. Reported root mean squared errors are scaled by  $\sqrt{n}$  so that they are comparable to the limiting standard errors given by the asymptotic theory described above. The limiting root mean squared errors were computed by numerically evaluating the limiting covariance matrices at the simulation settings with  $n = 10,000$ . The results provide some evidence for the superior efficiency of the quantile regression approach.

	Quantile Regression				Distributional Regression			
	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
$\alpha$								
$n = 100$	2.553	2.043	2.135	2.528	2.963	2.343	2.241	2.895
$n = 500$	2.507	1.981	2.010	2.556	2.894	2.140	2.209	2.832
$n = 1000$	2.427	2.008	2.024	2.417	2.719	2.225	2.212	2.724
$n = 5000$	2.501	1.940	1.965	2.565	2.928	2.245	2.210	2.791
$n = \infty$	2.500	2.042	2.042	2.500	2.796	2.249	2.253	2.808
$\beta$								
$n = 100$	2.699	2.209	2.194	2.708	3.438	3.037	2.949	3.268
$n = 500$	2.450	2.033	1.942	2.395	3.096	2.651	2.629	2.972
$n = 1000$	2.571	2.121	2.155	2.599	3.104	2.732	2.661	3.003
$n = 5000$	2.377	2.093	2.111	2.594	3.024	2.743	2.746	2.996
$n = \infty$	2.493	2.035	2.035	2.493	2.975	2.666	2.658	2.946

TABLE 1. Root Mean Squared Error in 1000 Replications: Bennett Model

## 4. A GUMBEL SPECIFICATION

A variety of other models can be accommodated in a similar fashion. As another illustration, consider the linear model (2.1) with  $\{u_i\}$  iid from the Gumbel, or Type 1 extreme value distribution,  $F(u) = 1 - e^{-e^u}$ . Then

$$\begin{aligned} P(Y_i > y|x_i) &= P(u_i > y - \alpha - x_i^\top \beta) \\ &= 1 - F(y - \alpha - x_i^\top \beta), \end{aligned}$$

so

$$\log(-\log(P(Y_i > y|x_i))) = y - \alpha - x_i^\top \beta.$$

This is the complementary log-log binary response model and can be easily estimated using this link function, McCullagh and Nelder (1989). In survival analysis this model corresponds to the Cox proportional hazard model.

Again, we can compare performance of direct quantile regression estimation of the model with estimation of the model via the corresponding binary response estimator. In Table 2 we report results from another simulation exercise structured exactly as in the logistic case except that  $u_i$ 's are Gumbel and the binary response uses the complementary log-log link function.

	Quantile Regression				Distributional Regression			
	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
$\alpha$								
$n = 100$	1.248	1.315	1.587	2.188	1.662	1.550	1.866	2.796
$n = 500$	1.242	1.371	1.652	2.275	1.674	1.573	1.832	2.565
$n = 1000$	1.237	1.368	1.630	2.264	1.542	1.571	1.855	2.675
$n = 5000$	1.292	1.384	1.603	2.253	1.562	1.589	1.818	2.514
$n = \infty$	1.243	1.337	1.599	2.241	1.552	1.541	1.821	2.544
$\beta$								
$n = 100$	1.316	1.419	1.695	2.427	2.645	2.387	2.530	2.951
$n = 500$	1.245	1.309	1.548	2.089	2.264	2.131	2.143	2.452
$n = 1000$	1.233	1.275	1.586	2.190	2.302	2.082	2.078	2.343
$n = 5000$	1.232	1.339	1.659	2.266	2.205	2.046	2.037	2.351
$n = \infty$	1.239	1.333	1.594	2.234	2.250	2.082	2.078	2.339

TABLE 2. Root Mean Squared Error in 1000 Replications: Cox Model

## 5. DICTA AND CONTRADICTA

Given the parameterization we have considered we have seen that the quantile regression estimator is considerably more accurate than the distributional regression estimator. Since the latter estimator also requires us to make a potentially controversial choice of an appropriate link function, this would seem to be a compelling argument for the quantile regression approach.

On the contrary, it should be remembered that two swallows do not make a summer, and likewise other choices of the iid error distribution and its corresponding link function may well yield different conclusions. Perhaps more crucially, we may prefer to evaluate performance of the two methods differently: according to how well they do in predicting conditional probabilities, rather than – as implicitly done here – conditional quantiles, and this too may produce contradictory evidence.

## REFERENCES

- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2010): “Inference on counterfactual distributions,” Available from <http://www.cemmap.ac.uk/wps/cwp0909.pdf>.
- DOKSUM, K., AND M. GASKO (1990): “On a correspondence between models in binary regression and survival analysis,” *Intl. Stat. Rev.*, 58, 243–252.
- FIRPO, S., N. FORTIN, AND T. LEMIEUX (2010): “Unconditional Quantile Regressions,” *Econometrica*, forthcoming.
- MCCULLAGH, P., AND J. NELDER (1989): *Generalized Linear Models*. Chapman and Hall, London, second edn.