ON DISTRIBUTIONAL VS. QUANTILE REGRESSIONS

ROGER KOENKER

1. Introduction

Recent work by Peracchi (2002), Firpo, Fortin, and Lemieux (2009), Chernozhukov, Fernández-Val, and Melly (2013) and Rothe and Wied (2013) 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

$$Y_i = \alpha_0 + x_i^\top \beta_0 + u_i \quad i = 1, \cdots, n, \tag{1}$$

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

$$\begin{aligned} S^{-1}(P(Y_i > y|x_i)) &= S^{-1}(P(u_i > y - \alpha_0 - x_i^\top \beta_0)) \\ &= S^{-1}(1 - F(y - \alpha_0 - x_i^\top \beta_0)) \\ &= y - \alpha_0 - x_i^\top \beta_0. \end{aligned}$$

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 (2013).

Under mild conditions, e.g. McCullagh and Nelder (1989), the binary response estimator, $\hat{\theta}_n(y) = (\hat{\alpha}_n(y), \hat{\beta}_n(y))$, is asymptotically 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 $\nu(y) = y - \alpha_0 - x_i^\top \beta_0$. The matrix X has typical row $(1, x_i^\top)$.

For the corresponding quantile regression estimator,

$$\check{\theta}_n(\tau) = (\check{\alpha}_n(\tau), \check{\beta}_n(\tau)) = \text{argmin}_{a,b} \sum_{i=1}^n \rho_\tau(Y_i - a - x_i^\top b)$$

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

Version: March 31, 2011.

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

	Quantile Regression				Distributional Regression			
	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
α								
$n = 100$	3.633	2.866	2.961	3.497	4.186	4.248	4.743	6.750
$n = 500$	3.219	2.619	2.631	3.245	3.916	3.865	4.414	5.973
$n = 1000$	3.253	2.654	2.661	3.243	3.771	3.790	4.316	5.974
$n = 5000$	3.141	2.491	2.516	3.255	3.929	3.880	4.361	6.070
$n = \infty$	2.500	2.042	2.042	2.500	2.796	2.249	2.253	2.808
β								
$n = 100$	1.583	1.291	1.331	1.601	2.925	3.736	5.009	9.048
$n = 500$	1.206	0.997	0.988	1.230	2.753	3.354	4.410	6.948
$n = 1000$	1.389	1.108	1.128	1.337	2.713	3.301	4.172	6.568
$n = 5000$	1.141	0.954	0.967	1.223	2.815	3.410	4.383	6.744
$n = \infty$	2.493	2.035	2.035	2.493	2.975	2.666	2.658	2.946

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 (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 τ'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 $n = 10,000$. Table 2 reports results for an identical experiment, except that the $\{x_i\}$ are iid log-normal. The results provide some evidence for the superior efficiency of the quantile regression approach.

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

	Quantile Regression				**Distributional Regression**			
	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
α								
$n = 100$	3.633	2.866	2.961	3.497	4.186	4.248	4.743	6.750
$n = 500$	3.219	2.619	2.631	3.245	3.916	3.865	4.414	5.973
$n = 1000$	3.253	2.654	2.661	3.243	3.771	3.790	4.316	5.974
$n = 5000$	3.141	2.491	2.516	3.255	3.929	3.880	4.361	6.070
$n = \infty$	2.500	2.042	2.042	2.500	2.796	2.249	2.253	2.808
β								
$n = 100$	1.583	1.291	1.331	1.601	2.925	3.736	5.009	9.048
$n = 500$	1.206	0.997	0.988	1.230	2.753	3.354	4.410	6.948
$n = 1000$	1.389	1.108	1.128	1.337	2.713	3.301	4.172	6.568
$n = 5000$	1.141	0.954	0.967	1.223	2.815	3.410	4.383	6.744
$n = \infty$	2.493	2.035	2.035	2.493	2.975	2.666	2.658	2.946

4. A Gumbel Specification

A variety of other models can be accommodated in a similar fashion. As another illustration, consider the linear model (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 3 we report results from another simulation exercise structured exactly as in the first logistic case except that u_i's are Gumbel and the binary response uses the complementary log-log link function.

5. Discrete Response

Another rationale that has been advanced for preferring the $\hat{F}$ approach over the $\hat{Q}$ approach involves the existence of mass points for the response variable, for example points corresponding to the minimum wage in labor economics wage regressions. The details of this argument are a bit unclear, but it may be worth noting that in the simplest setting of iid observations from a discrete random variable, the sample quantiles converge faster than $\sqrt{n}$ except on a subset of measure zero of $(0, 1)$, while $\hat{F}$ converges at its usual $\sqrt{n}$ rate. This observation extends naturally to p-sample data, and to regression settings as well unless there is some form of misspecification. There must be more to this than meets the eye.

Table 3. Root Mean Squared Error in 1000 Replications: Cox Model

	Quantile Regression				**Distributional Regression**			
	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
α								
$n = 100$	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
β								
$n = 100$	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

6. 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 (2013): "Inference on counterfactual distributions," *Econometrica*, forthcoming.

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 (2009): "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973.

McCullagh, P., and J. Nelder (1989): *Generalized Linear Models*. Chapman and Hall, London, second edn.

Peracchi, F. (2002): "On estimating conditional quantiles and distribution functions," *Computational Statistics & Data Analysis*, 38, 433–447.

Rothe, C., and D. Wied (2013): "Misspecification testing in a class of conditional distribution models," *J. Am. Stat. Assoc*, 108, 314–324.