Bayesian Analysis (2012) **7**, Number 2, pp. 273–276

Rejoinder

Kristian Lum* and Alan E. Gelfand†

We very much appreciate these three diverse discussions with virtually no overlap across them. We first take up the comments of Guhaniyogi and Banerjee (henceforth GB). With regard to the association structure under the asymmetric Laplace process (ALP), perhaps our presentation was not as clear as it should have been. Working with say isotropic covariance functions, we find that, regardless of the specification for the $\xi(s)$, the resulting correlation depends only on the distance between locations and is symmetric in p away from .5. Explicitly,

$$\text{corr}(\epsilon(s), \epsilon(s')) = \frac{\rho(||s-s'||)E(\sqrt{U(s)U(s')} + b_p \text{corr}(U(s), U(s')))}{1+b_p}$$

where marginally, the $U(s) \sim \text{Exp}(1)$ and $b_p = \frac{(1-2p)^2}{2p(1-p)}$. Note that b_p is minimized at 0 when $p = .5$ and tends to ∞ as $p \to 0, 1$. With a common $U(s) = U$, we see that regardless of s and s', the correlation can not go to 0, taking its minimum at $p = .5$, tending to 1 as $p \to 0, 1$. We don't employ this case. With a copula spatial process model for $\xi(s)$, equivalently, $U(s)$, we find that, for any p, correlation will go to 0 as $||s - s'|| \to \infty$. Again, it will take its minimum at $p = .5$, tending to $\text{corr}(U(s), U(s'))$ as $p \to 0, 1$, given s and s'. We don't employ this case either. For the case of i.i.d. $\xi(s)$, the second term in the numerator disappears and the expectation in the first term is constant ($\pi/4$). So now, for any p, correlation will go to 0, as determined by ρ and is strongest at $p = .5$, tending to 0 regardless of s and s' as $p \to 0, 1$. This behavior seems to be what would be desired for the $\epsilon(s)$ process.

With regard to the asymmetric Laplace predictive process (ALPP), we liked the novel form of the "bias" adjustment that arises due to the constraint that var$\ddot{Z}(s)$ must be 1. The tapered adjustment form in Sang and Huang (2012) is attractive but, we agree that its use is not likely to affect the inference in the present context. We concede that employing the double Gaussian process, drawing from Kottas and Krnjajić (2009) would be more flexible than the ALP and is investigated in the thesis of Lum (2010). Its properties and implementation issues are discussed there but presentation was beyond the scope of this paper. Finally, we do like the GB idea of joint modeling of spatial quantiles, imagining an application for modeling quantiles of ozone and $PM_{2.5}$ exposure.

We must take issue with the discussion of Lin and Chang (henceforth LC). They present a simulation example which seemingly reveals some shortcomings of the ALP. They claim that because our method does not perform well under the loss function they suggest, $SSE(p) = (q_p - \hat{q}_p)^2$, it does not provide the same flexibility as its semi-parametric frequentist cousin.

First, we disagree that we are using a mean regression model. We are certainly not

*Unaffiliated, kristianlum@gmail.com
†Department of Statistical Science, Duke University, Durham NC, alan@stat.duke.edu

 DOI:10.1214/12-BA708REJ

and further note that this points to a limitation of their example. More precisely, a mean regression model provides quantiles $q_p(X)$ in the form of $q_p(X) = \mu(X) + F^{-1}(p)$ where F is the cdf of the mean 0 error distribution. Such models imply that $q_p(X) - q_{p'}(X)$ is free of X regardless of the choice of error distribution. Furthermore, in their simulation, LC conveniently choose $\mu(X(s))$ to be strictly monotone as s moves away from $\mathbf{0}$. In practice, covariate surfaces are not monotone. The above assumptions are favorable to the squared error loss comparison they offer. Our conditional quantile specification is built from $P(Y_X < \mu_p(X)) = p$, i.e., 0 is the pth quantile of our error distribution. For us, $\mu_p(X) - \mu_{p'}(X)$ is not free of X. This is apart from our further objective of allowing our regression parameters to vary with p, in order to see how the nature of the quantile relationship changes across p. The approach of LC is not capable of making such assessment, so it is not a viable alternative for our purposes. Lastly, we will certainly concede that the flexibility of a spatial mean regression model can be sufficient to accommodate many generative models– including some with long tails.

A second point of contention is that the choice of loss function for LC, $SSE(p)$, is not an ideal loss function for evaluating quantiles. To really assess whether the pth quantile model, q_p, is fitting well, one must ask what the probability is that an observation drawn from the distribution will be less than the fitted quantile, $\hat{q}_p$. If this quantile is fitting well, that probability ought to be roughly p. Check loss has also been used as a convenient metric for assessing the fit of quantiles because it has the property that it is minimized when $100p\%$ of the data is below the fitted value. It has the added advantage over the sample proportion in that it has some sensitivity to how close the prediction is to the actual value– it penalizes predictions very far away from the data value more than those that are close but retains the property that it is small when the sample proportion is near p.

We also provide a simulation example which shows that maximum likelihood quantile regression estimates also perform poorly under $SSE(p)$. In fact, we replicate datasets, avoiding concerns regarding average performance associated with a single dataset as in LC. For simplicity, we remove the spatial component so that the classic frequentist version may be fit. We simulate $y_i \sim N(0, 1)$ for $i = 1 : 100$. We fit an intercept-only regression and compare the fit of the quantile estimated from frequentist quantile regression (as implemented in the quantreg package for R) to those implied by a least squares regression (lm in R). For the least squares regression, we estimate the pth quantile as $\hat{q}_p = \hat{\mu} + \hat{\sigma}\Phi^{-1}(p)$. This is to be compared to the maximum likelihood based quantile estimates. Here, the true $q_p = \Phi^{-1}(p)$ is known. We assess the fit to the true quantile using LC's $SSE(p)$, and summarize the results in Table 1, which shows the sum of $SSE(p)$ across simulations. We find that frequentist quantile regression also performs poorly compared to a standard linear regression under this model and loss function.

We re-evaluate the above simulation example using the sample proportion less than the fitted quantile and check loss. The results are shown in Table 2. We find that, although the normal regression still performs well, under the check loss function, it generally does not do as well as quantreg.

Table 1: SSE(p) for lm and quantreg

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
lm	18.79	13.87	11.56	10.39	9.98	10.23	11.22	13.34	17.98
quantreg	28.70	21.11	18.13	15.89	13.09	15.89	17.84	20.13	28.60

Table 2: Proportion less than fitted quantile and check loss for lm and quantreg

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
lm-prop	0.09	0.23	0.29	0.39	0.46	0.60	0.72	0.85	0.92
quantreg-prop	0.10	0.20	0.30	0.40	0.49	0.59	0.69	0.79	0.89
lm-check	124	80	57	45	40	44	57	82	127
quantreg-check	110	83	57	45	40	44	56	74	105

Finally, we also simulate from the Gaussian-Log-Gaussian (GLG) model using the same parameters as LC. We fit the spatial mean regression using the default settings in the spBayes package in R. We run each for 10,000 iterations. Table 3 shows the results for these models. We find that the ALP performs better than Bayesian spatial mean regression (BSMR) from the perspective of the check loss function. Quantreg, again, does quite well considering that it has many fewer parameters than the other two models– it has no spatial component. In this case, it is competitive with the spatial models primarily because LC chose a parameter for the spatial covariance that results in weak spatial structure (LC use a range of .25 relative to a maximum distance of $\sqrt{2}$ over the unit square).

Table 3: Proportion less than fitted quantile and check loss for BSMR, ALP, and quantreg

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
BSMR-prop	0.07	0.12	0.23	0.39	0.56	0.71	0.81	0.87	0.91
ALP-prop	0.12	0.22	0.28	0.38	0.47	0.54	0.63	0.73	0.90
quantreg-prop	0.09	0.20	0.31	0.39	0.49	0.59	0.69	0.79	0.89
BSMR-check	56	80	93	100	106	110	107	96	69
ALP-check	45	68	83	95	98	101	99	87	62
quantreg-check	56	78	94	105	110	112	112	103	81

Lastly, Ferreira makes it easy for us. In light of the foregoing discussion regarding induced dependence structure, we think the i.i.d. case is preferred. Moreover, since we have spatial dependence in the $\epsilon(s)$ process arising from that in the $Z(s)$ process, it is unclear whether the data would enable us to distinguish the case of i.i.d. $\xi(s)$ from the spatially structured case. Should we seek to implement the copula-based model for $\xi(s)$, the challenge is primarily computational, as updating each $\xi(s)$ would require recomputing the likelihood of all $\xi(s)$'s at every iteration, i.e. n expensive computations per iteration.

Again, we thank the reviewers for their time in preparing discussions for us. As LC note, quantile regression is an area with a rapidly increasing literature. There is room for contribution in both the conditional and unconditional cases, in both parametric and nonparametric specifications, and in revealing the inferential benefits of working within the Bayesian framework.

References

Kottas, A. and Krnjajić, M. (2009). "Bayesian Semiparametric Modelling in Quantile Regression." *Scandinavian Journal of Statistics*, 36: 297–319. 273

Lum, K. (2010). "Bayesian Spatial Quantile Regression." Ph.D. thesis, Duke University. 273

Sang, H. and Huang, J. Z. (2012). "A full scale approximation of covariance functions for large spatial data sets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1): 111–132. 273