Rejoinder

Yunwen Yang [1], Huixia Judy Wang [2], and Xuming He [3]

[1] *Google Inc., Seattle, WA, USA*

E-mail: yunweny@google.com

[2] *George Washington University, Washington, D.C., USA*

E-mail: judywang@gwu.edu

[3] *University of Michigan, Ann Arbor, MI, USA*

E-mail: xmhe@umich.edu

We are fortunate to have comments from a group of experienced and thoughtful colleagues on what we have done for the posterior inference in Bayesian quantile regression. We are humbled by the encouraging comments they made, and we are certainly glad to hear that our work has merit. We would like to take this opportunity to add some of our thoughts in response to those comments.

1 Can we trust the working model?

Meng is very direct in pointing out that the asymmetric Laplace working likelihood is simply too artificial; in general it does not provide a decent approximation to the underlying likelihood. This sentiment is shared by Smith. In fact, two such working likelihoods at two values of τ would conflict with each other. The fact that we are not using a likelihood that approximates the data generating mechanism explains the lack of validity of the posterior inference from such working models. Even though we have shown that an adjustment in the posterior variance leads to asymptotically valid posterior inference, by no means can we defend the choice of the likelihood from the modeling point of view. The choice of the asymmetric Laplace likelihood is mainly motivated by computational convenience. The connection to the "Bartlization" of Meng (2009) and to the quasi-likelihood approach of Kim (2014) can certainly help understand our proposed posterior adjustment. Along this line, Meng asked whether a better and more flexible parametric working model can be identified in the quantile regression setting. The answer to the question is likely to depend on how much we are willing to assume about the conditional distributions of the response variable Y given the covariate X. If we assume only a linear quantile function at a given level τ, the semiparametric efficiency result of Newey and Powell (1990) suggests that properly weighted asymmetric Laplace working likelihood is "optimal". On the other hand, if $\epsilon_i = Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_\tau$ are i.i.d., then more choices of the working likelihood become available. We refer to Gilchrist (2000) for choices of parametric models with quantile functions. Those models would approximate the likelihood through a parametric family. To take a semiparametric approach to quantile regression, any reasonable parametric likelihood would need to incorporate parameters that depend on $\mathbf{X}$, which would lead to a less desirable problem to solve when $\mathbf{X}$ is multivariate.

Is the asymmetric Laplace working likelihood harmful? We hope we have made it clear: yes it is harmful if one performs posterior inference blindly, but there is a simple remedy. Wang and Sherwood extended this approach to a more challenging problem: quantile regression with missing covariates. Once again, the adjusted posterior inference proves to be useful (and asymptotically justified) in that problem.

2 More about Wald-type confidence intervals

Koenker found it disturbing that the Wald-based "nid" confidence intervals performed poorly in one of our examples. Furthermore, Koenker demonstrated that if quadratic quantile functions are used to fit the model, the performance of the Wald intervals in that case would become quite acceptable. As we know, the Wald-type of intervals are asymptotically valid. The main difficulty with the Wald-type intervals is the lack of robustness; the results tend to depend quite critically on how the asymptotic variance-covariance is estimated. We take Koenker's experiment as another piece of evidence for the lack of reliability from the "nid" method. To take it further, we use the same model and examine the performance of the method with the quadratic quantile models at $\tau = 0.75$ instead of $\tau = 0.5$; see Table 1 for the results based on 1000 repetitions where the format of the Table follows that of Koenker's Table 2. We note that the "nid" method no longer holds up, resulting in a low coverage for b_4.

The proposed method BAL_{adj} remains competitive across multiple studies, but we do notice that the resulting intervals are sometimes conservative, and we cannot claim it is always the best performer. For linear quantile models with complete data, several competitive methods exist. When the response is censored or when a covariate has missing values, the usual asymptotic variance-covariance becomes even harder to estimate reliably. In those cases, fewer satisfactory solutions exist, making us believe that the Bayesian methods become more valuable relatively to the others in the problems with incomplete data.

3 Does the method work for more general models?

Both Smith and Wang and Sherwood raised a good question whether the posterior adjustment would work well in the presence of a high dimensional covariate or under a broader class of models. Obviously, we have no experience yet when the number of covariates p is as large as the sample size n. The asymptotic theory needs to be done properly to address such questions. Here we would report one experiment with a partially linear model where the nonparametric function is approximated by a B-spline function. In this case, we went ahead to treat the problem as an approximately linear quantile problem, allowing us to examine the performance of the posterior interval estimates for the linear regression coefficients.

We generate data from the following model:

$$y_i = x_{i1} + x_{i2} + 2\sin(2x_{i3} + 2) + (1 + 0.5x_{i1})e_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $x_{ik}, k = 1, 2, 3$, are independent Uniform variables on (-1,1), and $e_i \sim N(0, 1)$ are white noise. The τth conditional quantile of Y is

$$Q_\tau(Y|x_1, x_2, x_3) = b_0(\tau) + b_1(\tau)x_1 + b_2(\tau)x_2 + m(x_3),$$

where $b_0(\tau) = \Phi^{-1}(\tau)$, $b_1(\tau) = 1 + 0.5\Phi^{-1}(\tau)$, $b_2(\tau) = 1$ and $m(x_3) = 2\sin(2x_3 + 2)$. The nonparametric function $m(x_3)$ is approximated by a quadratic B-spline function with three internal knots. Table 2 summarizes the coverage probabilities and average lengths of 90% confidence intervals for the parametric coefficients $b_k(0.5)$, $k = 0, 1, 2$. The results show that the proposed Bayesian method remains useful for correcting the problem in the unadjusted BAL intervals. Interval estimates on the nonparametric component of the model would require a more careful study of the bias-variance trade-off, and we leave the investigation to future research.

4 Corrections

We would like to correct two errors that have appeared in our paper. First, the definition of the quantile loss function $\rho_\tau(\mu)$ just after the display (2.2) in the paper should read $\rho_\tau(\mu) = \mu\{\tau - I(\mu < 0)\}$. Second, as pointed out by Hobert and Khare in their comments, we should have said in Section 2.2 that "Khare and Hobert (2012) showed that the Markov chain underlying this three-variable Gibbs sampling algorithm converges at a geometric rate at the median regression." We could extrapolate that this property holds for all $\tau \in (0, 1)$, but there is no proof yet. We are sorry for any possible confusion that our original statements may have caused.

5 Do we have better methods?

We are grateful to Hobert and Khare for discussing better Bayesian computational algorithms in their comments. They showed that a sandwich algorithm is superior to the data augmentation algorithm at any quantiel level. Although we do not have empirical experience with the sandwich algorithm yet, their work points to a better future for Bayesian inference methods for quantile regression.

Smith discussed limitations of the asymmetric Laplace working likelihood, and mentioned a number of other approaches to quantile modeling, especially for more complex problems.

If we perform analysis at multiple quantile levels, we have found in Yang and He (2012) that the empirical likelihood is theoretically attractive as a working likelihood in the Bayesian framework. The Bayesian empirical likelihood provides valid posterior inference and can improve efficiency by borrowing strength across quantiles, but it is computationally more difficult.

It is clear from the literature that more and more researchers have turned to Bayesian quantile regression. We hope that our article can help make the yellow light and the red light more visible to all the drivers, just to follow Meng's yellow/red light comment. More importantly, we believe that the green light will be flashing with possibly better working models and much more efficient algorithms that work for a wider range of problems than we have discussed so far.

References

Gilchrist, W. (2000), *Statistical Modelling with Quantile Functions*, Chapman & Hall/CRC.

Kim, J. (2014), "An alternative quasi likelihood approach, Bayesian analysis and data-based inference for model specification," *Journal of Econometrics*, 178, 132–145.

Meng, X. (2009), "Decoding the H-likelihood," *Statistical Science*, 24, 280–293.

Newey, W. and Powell, J. (1990), "Efficient Estimation of Linear and Type I Censored Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

Yang, Y. and He, X. (2012), "Bayesian empirical likelihood for quantile regression," *The Annals of Statistics*, 40, 1102–1131.

Table 1: Empirical coverage probabilities in 1000 trials for Model 2 in Section 4.1 of the main paper: Quadratic Model at $\tau = 0.75$.

	$n = 200$				$n = 500$			
	BAL_{adj}	BAL	RQ_{rank}	RQ_{nid}	BAL_{adj}	BAL	RQ_{rank}	RQ_{nid}
b_0	0.918	0.865	0.865	0.876	0.903	0.852	0.859	0.886
b_1	0.930	0.795	0.911	0.908	0.915	0.764	0.896	0.886
b_2	0.930	0.885	0.878	0.888	0.918	0.867	0.903	0.879
b_3	0.878	0.669	0.891	0.853	0.891	0.668	0.895	0.872
b_4	0.933	0.870	0.879	0.841	0.914	0.875	0.882	0.829

Table 2: Empirical coverage probabilities and mean lengths of confidence intervals with nominal level 90% for the partially linear model (3.1) at $\tau = 0.5$ in 1000 trials. The standard errors for EML are in the range of 0.001 to 0.017 in this table.

		$100 \times$ ECP			EML		
n	Method	$b_0(\tau)$	$b_1(\tau)$	$b_2(\tau)$	$b_0(\tau)$	$b_1(\tau)$	$b_2(\tau)$
200	BAL_{adj}	91	92	92	1.40	0.52	0.50
	BAL	84	82	86	1.10	0.39	0.39
	RQ_{rank}	89	87	90	1.36	0.47	0.46
	RQ_{nid}	84	86	88	1.29	0.47	0.44
500	BAL_{adj}	92	90	92	0.88	0.32	0.30
	BAL	87	80	84	0.71	0.25	0.24
	RQ_{rank}	90	87	90	0.85	0.29	0.29
	RQ_{nid}	87	86	88	0.83	0.30	0.28