**Issue with Linear Regression**

Prediction intervals based on standard linear model assumptions may not be appropriate when the data displays some of the following characteristics: 1) heteroscedasticity 2) non-normality 3) has outliers. In the picture below, we see that: 1) the mean is highly influenced by the outliers and 2) the 90% prediction interval is clearly incorrect.

**Quantile Regression**

For data as described previously, it may be more appropriate to use Quantile Regression developed by Koenker and Bassett [1] (the chart below was created using package ‘quantreg’ [2]).

Quantile regression is based on loss function:

\[ L_{\tau}(u) = \begin{cases} (\tau - 1)u & (u < 0) \\ \tau u & (u \geq 0) \end{cases} = u(\tau - I(u < 0)) \]

and the vector of parameters can be estimated using:

\[ \beta^{QR} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} L_{\tau}(y_{i} - x_{i}'\beta) \]

**Quantile Additive Models (QAMs)**

Quantile Additive Models (QAMs) [3] chap. 5, [4]) are non-linear multivariate extensions of Quantile Regression.

The QAM is defined as:

\[ g_{QAM} = \arg\min_{g} \frac{1}{n} \sum_{i=1}^{n} L_{\tau}(y_{i} - g(x_{i})) \]

where

\[ g(x_{i}) = g_{1}(x_{11}) + g_{2}(x_{21}) + \ldots + g_{p}(x_{p1}) + g_{1}(x_{12}x_{22}) + g_{13}(x_{11}x_{31}) + \ldots + g_{123}(x_{11}x_{22}x_{33}) + \ldots \]

First, function \( g \) can be represented as a weighted sum of basis functions leading to a representation as: \( g(x_{i}) = X_{i}'\beta \). Second, as the model is overparameterized, we introduce smoothing penalties on the \( g_{j} \) to control the bias-variance tradeoff. For a smooth curve this might be: \( \int (g''(x))^{2}dx \) which, given a basis can be re-written as \( \beta'^{T}S_{\beta}\beta \)

\[ \beta^{QAM-Penalty} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} L_{\tau}(y_{i} - X_{i}'\beta) + \frac{1}{2} \sum_{j=1}^{p} \lambda_{j}\beta_{j}^{2} \]

**GACV for Hyperparameters**

Yuan [5] based on the work of Craven and Wahba [6] and Nychka et al. [7] propose an approximation of Leave One Out Cross Validation (LOOCV) where only the knowledge of \( \beta_{h} \) is required to determine the vector of hyperparameters / smoothing parameters \( \Lambda = (\lambda_{1}, \ldots, \lambda_{p}) \)

\[ GACV_{n,\tau}(\Lambda) = \sum_{i=1}^{n} L_{\tau}(y_{i} - X_{i}'\hat{\beta}(\Lambda)) \]

\( L_{\tau,\nu}(\hat{\beta}(\Lambda)) \) is a smooth approximation of \( L_{\tau}(u) \)

\( edf_{\alpha,\tau}(\Lambda) = \text{effective degrees of freedom} \)

As \( \alpha \to 0 \)

\[ L_{\tau,\nu}(\Lambda) \to L_{\tau}(u) \]

However, there are two issues: a) \( \Lambda \) is optimized using a grid approach and b) Reiss and Huang [8] note that the GACV tends to undersmooth for non-central quantiles. We also note that although the central quantile is often fitted correctly, it is sometimes undersmoothed. Below is an example where extreme quantiles and the median are undersmoothed.

**Prop. 1: QGACV**

The derivation of the GACV is based on averaging weights of an ACV criterion ([7], [5]) \((h_{\alpha,\tau},\nu) \text{ are diag. elements of the ‘hat’ matrix).} \]

\[ ACV_{\alpha,\tau}(\Lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1-h_{\alpha,\tau}(\Lambda)} \left( y_{i} - X_{i}'\hat{\beta}(\Lambda) \right)^{2} \]

As pointed out by Reiss and Huang [8], this averaging does not take into account that the weights on each side of the regression curve are vastly different. However, we note that the same quantile regression curve could be obtained if we resampled from the side with the highest number of points and we applied a symmetric loss function resulting in more even weights. We then derive:

\[ QGACV_{\alpha,\tau}(\Lambda) = \frac{1}{n} \sum_{i=1}^{n} \left( y_{i} - X_{i}'\hat{\beta}(\Lambda) \right)^{2} \phi = \min_{\tau, \phi} \min_{\Lambda} \]

where \( \phi \) is a quasi-convex function. We then propose a quasi-convex function result.

**Prop. 2: Graduated Optim.**

Parameter \( \alpha \) determines the degree of approximation of the loss function.

Although the QGACV criterion is a non-convex function of \( \Lambda \), we note that as \( \alpha \) increases, the QGACV function becomes closer to a quasi-convex function. We then propose to use Graduated Optimization/Non-Convexity (Blaze and Zisserman [9], chap. 7) to determine the vector of smoothing parameters \( \Lambda \). It consists in solving the optimization problem at decreasing values of \( \alpha \) (i.e. we start at a large value of \( \alpha = \alpha_{start} \) optimize, then reduce \( \alpha \) and optimize, starting from the previous minimum and so on until we reach a value \( \alpha_{opt} \) that is ‘optimal’). Once the vector of smoothing parameters \( \Lambda \) is known, we continue to decrease \( \alpha \) to determine the vector of parameters \( \beta \) until we are close to the exact loss function.

**Results/Conclusion**

Using QGACV/Graduated Optimization both the central and extreme quantile fits are closer to the true quantiles in terms of MSE.

**References/Acknowledgements**


This poster uses class ‘ biposter ’ and is based on poster examples created by Brian Amburg. http://www.brian-amburg.de/unix/poster/