Smooth Bootstrap Inference for Parametric Quantile Regression

Tatjana Kecojević and Peter Foster

Lancashire Business School, University of Central Lancashire
Preston, UK

School of Mathematics, University of Manchester
Manchester, UK

The 8th International R Users Meeting

June 15, 2012

Regression and Quantiles

- Regression modelling enables the study of the relationship between a response variable Y and a set of covariates x.
- Normal linear regression models how the mean of Y changes with x; ie. $E(Y|\boldsymbol{X} = x)$.
- However, a single mean curve may not be informative enough in certain contexts.
- Quantile Regression enables us to explore more fully the conditional distribution of the response on the covariates.

Linear Regression Quantile (Koenker and Bassett, 1978)

A random sample $\{y_1, y_2, ..., y_n\}$,

- Median = $\text{argmin}_\xi \, \Sigma \, |y_i - \xi|$;
- The τ^{th} sample quantile of y_i is a solution to:

$$R(\xi) = \underset{\xi}{\text{argmin}} \, \Sigma \rho_\tau(y_i - \xi), \tag{1}$$

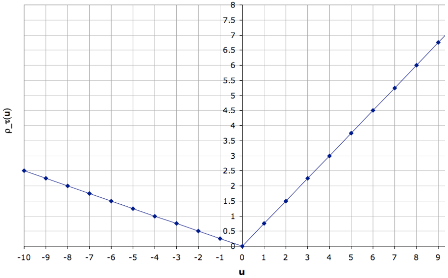
where $\rho_\tau(u)$ is a peace-wise function:

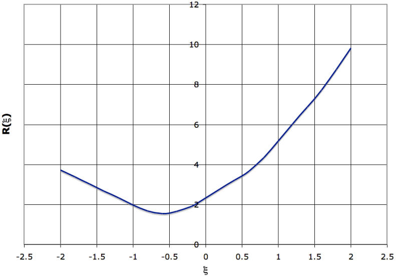
$$\rho_\tau(u) = u\big(\tau - I(u < 0)\big), \qquad 0 < \tau < 1, \tag{2}$$

where

$$I(u < 0) = 1 \quad if \quad u < 0, \quad \text{and} \;\; 0 \;\; \text{otherwise.} \tag{3}$$

Linear Regression Quantile (Koenker and Bassett, 1978)





Following the analogy of defining the sample quantiles, we can define conditional quantile functions $Q_y(\tau|x) = \boldsymbol{x}^\top \boldsymbol{\beta}(\tau)$, where

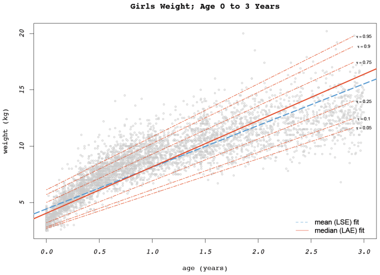
$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}^\top \boldsymbol{\beta})$$

Quantiles

- The estimated conditional quantile is $\hat{Q}_\tau(Y|\boldsymbol{X}) = \boldsymbol{X}^\top \hat{\beta}(\tau)$,
- $\hat{\beta}_\tau$ can be calculated efficiently by means of linear programming. eg. the simplex method (Koenker and D'Orey, 1987 and 1993) for moderate sample sizes.
- **R** package **quantreg** (contributed by Koenker) can be used to fit quantile rgeression models to data. The function **rq** uses the simplex method as its default.

SaudI Arabian Girls' Weight, Age Birth to 3 years

Scatterplot and Quantile Regression Fit for Girls Weight, Age Birth to 36 Months: The plot shows a scatterplot of the girls weight, age birth to 3 years, for a sample of $6,123$ observations. Superimposed on the plot are the $\{0.05,\ 0.01,\ 0.25,\ 0.50,\ 0.75,\ 0.90,\ 0.95\}$ quantile regression lines in dashed red, the median fit in a solid red line, and the least square estimate of the conditional mean function as the solid blue line.



$$Q_y(\tau|x) = \beta_0 + \beta_1 + \sigma(x)F_u^{-1}(\tau)$$

Sparsity

In the non-iid case the asymptotic distribution of $\hat{\beta}(\tau)$ is:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)\right)\sim\mathcal{N}\left(0,\tau(1-\tau)\boldsymbol{H}_n^{-1}\boldsymbol{J}_n\boldsymbol{H}_n^{-1}\right),$$

where

$$\boldsymbol{J}_n(\tau)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\top}$$

and

$$\boldsymbol{H}_n(\tau)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\top}f_i\left(\xi_i(\tau)\right).$$

The conditional density of the response, y_i, evaluated at the τ^{th} conditional quantile is given by $f_i\left(\xi_i(\tau)\right)$.

The nuisance quantity:

$$s(\tau)=\left[f\left(F^{-1}(\tau)\right)\right]^{-1},$$

known as the sparsity function, which is the derivative of the quantile function:

$$\frac{d}{d\tau}F^{-1}(\tau)=s(\tau).$$

quantreg

Computable in the **quantreg** package:

- Wald Tests - require estimation of *the sparsity parameter*
- Rank-Score Process - does not requires estimation of *the sparsity parameter*, depends on the *score function*
- Resampling Methods:

i (x, y) pair method

ii Parzen, Wei and Ying (1994) approach

iii Markov Chain Marginal Bootstrap, MCMB (Kocherginsky at el., 2005)

iv Weighted (x, y) pair method (Bose and Chatterjee, 2003)

Smooth Bootstrapping Using Conditional Variance Modelling

- We would like to used a smoothed bootstrap resampling technique to estimate CIs.
- When the errors are iid we can pool them together and estimate their common density from which to sample.
- In the non-iid case we exploit conditional variance modelling.

To motivate the methodology we consider the parametric case when the errors are Normally distributed. The idea is based on the following result:

If $X \sim \mathcal{N}(0;1)$ then $X^2 \sim \chi^2(1)$.

Now let $Y = \sigma X$, so that $Y \sim N(0;\sigma^2)$, thus

$Y^2 = \sigma^2 X^2 \sim Gamma(\frac{1}{2\sigma^2}, \frac{1}{2})$ and

$E[Y^2] = \frac{1/2}{1/(2\sigma 2)} = \sigma^2$

This suggests that for Normally distributed errors we can estimate the conditional variance function using a Gamma Generalised Linear Model.

Gamma Distribution

The density of the gamma distribution is usually given by:

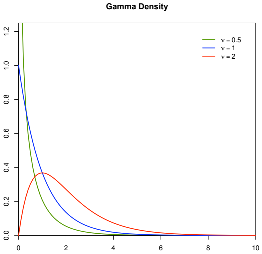
$$f(y) = \frac{1}{\Gamma(\nu)} \lambda^{\nu} y^{\nu-1} e^{-\lambda y}, \qquad y > 0,$$

where ν describes the shape and λ describes the scale of the distribution. Thus, if Y has a gamma distribution, with parameters ν and λ,

$$E[Y] = \frac{\nu}{\lambda}, \qquad Var[Y] = \frac{\nu}{\lambda^2}, \qquad and \qquad m_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\nu} \quad for \quad t < \lambda.$$

For gamma family the *canonical parameter* is inverse, $\theta = -1/\mu$, so that the *canonical link* is

$$\eta = g(\mu) = -\mu^{-1} = -\frac{\nu}{\lambda}.$$



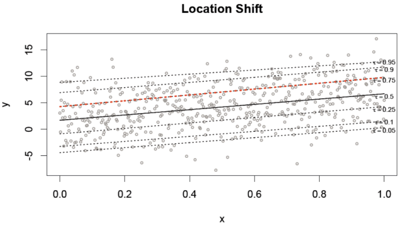
Example: Models with the error from Normal distribution

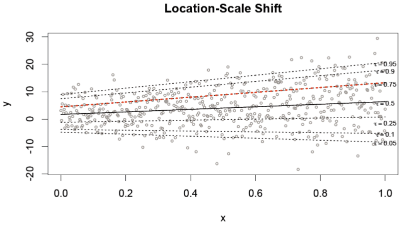
Let us consider the following two models:

$$\begin{aligned} \text{homoscedastic Model 1} &: \quad y_i = 2 + 5x_i + e_i \quad \text{and} \\ \text{heteroscedastic Model 2} &: \quad y_i = 2 + 5x_i + \sigma(x_i)e_i \end{aligned}$$

where $x \in [0,1]$ and $x_i = i/n$ for $n = 500$, and with $\{e_i\}$ iid from $\mathcal{N} \sim (0, 16)$ and $\sigma(x) = \sqrt{1+4x}$, $\tau = 0.75$.

i) Estimate the τ^{th} quantile function of interest: $\hat{Q}_y(\tau|x) = \boldsymbol{x}^\top \hat{\beta}(\tau)$;





Example: Smooth Bootstrapping Using Conditional Variance Modelling

ii) Obtain the residuals $\{u_1(\tau), \ldots, \ u_n(\tau)\}$: $u_i(\tau) = y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}(\tau)$;

iii) Using the estimate of the mean function of the residuals,

$\hat{E}\Big[u_i(\tau|x)\Big] = \tilde{u}_i(\tau) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_{\hat{u}} + \epsilon_{\hat{u}}$,

construct a set of centered and squared residuals: $su_i(\tau) = \Big(u_i(\tau)) - \tilde{u}_i(\tau)\Big)^2$;

iv) The conditional mean of these squared residuals is equal to the conditional variance

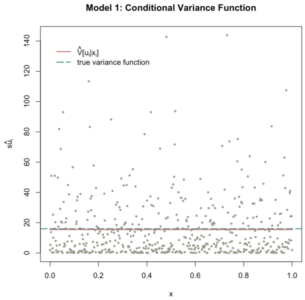
i.e. $\hat{V}\Big(u_i(\tau)\Big) = \hat{E}\Big(su_i(\tau)\Big)$

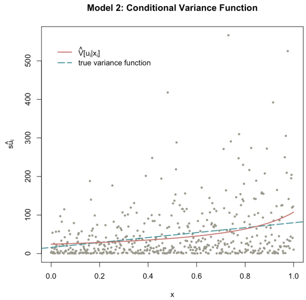
Example: Conditional Variance

We can parametrically estimate the conditional variance function using a gamma GLM function in **R**:

M1 : `V(u|x)=glm(`$s\hat{u}_i \sim$ `1, family=Gamma(link="inverse"))`

M2 : `V(u|x)=glm(`$s\hat{u}_i \sim$ `1 + x, family=Gamma(link="inverse")).`





Example: Kernel Density Estimate

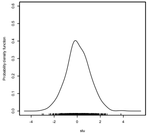
which enables the standardisation of the residual

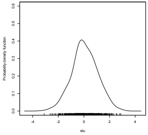
$$stu_i(\tau) = \hat{u}_i(\tau)/\hat{V}[u_i(\tau)];$$

v) Construct the kernel density estimate of the standardised residuals:

$$\hat{f}_{u(\tau)}(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t - stu_i(\tau)}{h}\right),$$

where K is the kernel function (eg. the $N(0;1)$ pdf) and h is the smoothing parameter, $h > 0$. $\hat{f}$ is consistent in MISE if $h \to 1$ (see Silverman (1986)).





The figures plot the kernel density estimate of the standardised residuals for the two models: the location-shift and the location-scale models respectively, for the 75^{th} quantile regression line.

Example: Silverman's Algorithm (1986)

vi) Draw a sample of standardised residuals from $\hat{f}(stu)$ using Silverman's (1986) algorithm:

Step 1: Chose I uniformally with replacement from $\{1, \ldots, n\}$;

Step 2: Generate ϵ to have probability density function K;

Step 3: Set $stu_i^* = stu_I(\tau) + h\epsilon$, or in the case when the realisations stu_i are transformed to reflect the first and second moment properties observed in the sample $\{stu_1, \ldots, stu_n\}$, use

$$stu_i^* = \mu_{stu_i} + (stu_i - \mu_{stu_i} + h\epsilon)/\sqrt{(1 + h^2\sigma_K^2/\sigma_{stu_i}^2)};$$

vii) Scale the standardised residuals to their original scale and construct a smooth bootstrap sample $y_i^* = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}(\tau) + stu_i^* \sqrt{\hat{V}[u_i|x_i]}$;

viii) Re-fit the quantile regression model to the bootstrap data

$$Q_{y_b^*}(\tau|x) = \boldsymbol{x}^\top \boldsymbol{\beta}_b^*(\tau)$$

to obtain a new set of parameter estimates $\hat{\boldsymbol{\beta}}_b^*(\tau)$.

ix) Repeat this resampling process to build up the distribution of the parameter estimates empirically.

x) Use this bootstrap distribution to make inferences about $\boldsymbol{\beta}(\tau)$.

Example: Simulation Study

Let us focus on the problem of the confidence intervals for the median regression ($\tau = 0.5$) parameters as it is most straight-forward when comparing the performance between the different methods and the true parameters values. We will consider the two given models: pure location-shift model (Model 1) and location-scale model (Model 2), using:

- `glm.Gamma.Inv_ksm`: redefine kernel smooth bootstrap in which the density estimate has the same mean and the variance as the empirical data used to estimate the density.

and the following **R** based functions:

- `xy`: $(x,\ y)$ pair bootstrap method.
- `pwy`: Parzen-Wai-Ying bootstrap method.
- `mcmb`: markov chain marginal bootstrap method.
- `wxy`: generalized bootstrap method of Bose and Chatterjee (2003) with unit exponential weights.

We use 500 realisations ($R = 500$) of 500 observations ($n = 500$) with $1,000$ bootstraps ($B = 1,000$).

Example: Simulation Results - Parameter estimates

homoscedastic Model 1 : $y_i = 2 + 5x_i + e_i$ and

heteroscedastic Model 2 : $y_i = 2 + 5x_i + \sigma(x_i)e_i$

Model 1 (M1)	β_0		β_1	
Method	b_0	SE	b_1	SE
glm.Gamma.Inv_ksm	2.02677	0.01875	4.96014	0.03388
xy	2.01944	0.01842	4.96510	0.03279
pwy	2.01990	0.01849	4.96422	0.03281
mcmb	2.01956	0.01848	4.96548	0.03274
wxy	2.02018	0.01847	4.96462	0.03284
Model 2 (M2)	β_0		β_1	
Method	b_0	SE	b_1	SE
glm.Gamma.Inv_ksm	2.03913	0.02562	4.93566	0.05739
xy	2.02679	0.02452	4.94486	0.05476
pwy	2.02757	0.02460	4.94344	0.05471
mcmb	2.02895	0.02476	4.94279	0.05532
wxy	2.02791	0.02458	4.94447	0.05482

Example: Simulation Results - 95% CIs

Results: $\tau = 0.5$; column C is coverage probability and column L is average length of the 95% confidence intervals for each coefficient.

Model 1 (M1)	β_0			β_1		
Method	C	L	SE	C	L	SE
glm.Gamma.Inv_ksm	96.00%	1.75092	0.00469	95.20%	3.03332	0.00818
xy	96.40%	1.79039	0.01397	95.40%	3.13494	0.02213
pwy	96.60%	1.79458	0.01414	96.00%	3.13874	0.02229
mcmb	95.60%	1.77899	0.01209	94.60%	3.10114	0.02173
wxy	96.20%	1.78469	0.01408	95.00%	3.12381	0.02221
Model 2 (M2)	β_0			β_1		
Method	C	L	SE	C	L	SE
glm.Gamma.Inv_ksm	98.40%	2.84057	0.00759	93.40%	4.92105	0.01326
xy	96.80%	2.38126	0.01853	94.40%	5.16402	0.03728
pwy	96.80%	2.38565	0.01864	94.60%	5.16527	0.03745
mcmb	98.00%	2.60154	0.01722	95.00%	5.23220	0.03808
wxy	96.00%	2.37212	0.01863	94.20%	5.14685	0.03734

Example: Comments

1. close parameters estimates for both models and good coverage probability;
2. Model 1: kernel smooth bootstrapping, *ksm*, adjusted for the first and the second moment is very competitive compared to the other bootstrapping methods;
3. Model 2: L of β_0 slightly wider than those of other bootstrapping methods, but in case of β_1 *ksm* is outperforming the others;
4. *ksm* performs well when the error in the model is normally distributed $=>$ HOW ROBUST is this method when the error in the model is non-normal, i.e. squared residuals are non-gamma?

Robustness

Consider same models, M1 & M2, but this time we consider $\{e_i\}$ which is iid, from tree different distributions:

- $\mathcal{N}(0, 16)$,
- $t(20)$ and
- $t(10)$,

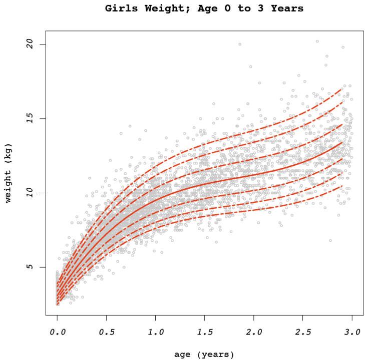
where $x \in [0, 1]$ and $x_i = i/n$ for $n = 500$, and where $\sigma(x) = \sqrt{1 + 4x}$. We again focus 50^{th} quantile, $\tau = 0.5$, thus 50^{th} quantile of the u_i variable is at 0.

- Parametric modelling of the variance function by:
  - i) Extending GLM:
    - Iteratively Re-weighted Least Squares (IRWLS) (Nelder and Wedderbur, 1972);
    - Joint Modelling of Mean and Dispersion (Nelder and Lee, 1991);
    - Tweedie GLM (Smyth and Jorgensen, 1999);
  - ii) Double GLM with Tweedie Family (Smyth, 1989),
  - iii) Quasi-Likelhood GLM (Wedderburn, 1974),
  - iv) Robust GLM (Cantoni and Ronchetti, 2001).
- Non-Parametric modelling of the variance function by:
  - i) Local Polynomial Regression (oder $p = 1$ and $p = 3$) (Hall and Carroll, 1989);
  - ii) Locally Weighted Polynomial Regression (Cleveland, 1979);
  - iii) Difference Based Variance Estimation, using first order difference, Δ_1 (Rice, 1984).

Outcomes

1. The choice of the method used for the estimation of the conditional variance function depends not only on the type of the regression model we deal with, but also on the underlying variance function itself.
2. We suggest the parametric estimation of the conditional variance function using DGLM with Tweedie family with log link to be applied to the standardisation of the residuals used in the kernel smoothing bootstrapping adjusted to have the same mean and the variance as the data from which it is constructed.
3. Despite its popularity and the development in terms of its application the issue of the bandwidth selection has not been adequately addressed for the difference-based variance estimation method making this approach complex to implement.
   - The CV approach is not just sensitive to the outliers but also to the skewness in the data.

SaudI Arabian Study: Girls' Weight, Age Birth to 3 years



$$Q_y(\tau|\boldsymbol{x}) = \beta_0(\tau) + \beta_1(\tau)x + \beta_2(\tau)x^2 + \beta_3(\tau)x^3,$$

for $\tau \in (0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95)$

Fitted models using **rq.fit** function in **R**:

$$\begin{aligned} Q_y(\tau = 0.5|\boldsymbol{x}) &= 3.100 + 10.630x - 5.166x^2 + 0.939x^3, \\ Q_y(\tau = 0.75|\boldsymbol{x}) &= 3.400 + 11.371x - 5.412x^2 + 0.973x^3. \end{aligned}$$

SaudI Arabian Study: Parameter estimates

ksm - The conditional variance function was estimated using the proposed DGLM with the Tweedie family with the log link by fitting a cubic model to the centered squared residuals

$$\text{V(u|x)} = \text{dglm}(s\hat{u}_i \sim \text{1 +x I(x^2)+I(x ^3),}$$
$$\text{dformula=} \sim \text{x, family=tweedie(var.power=2, link.power=0), method="reml")}$$

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\tau = 0.50$				
ksm	3.15664	10.62845	-5.16428	0.93900
xy	3.11342	10.56746	-5.11237	0.92739
pwy	3.11213	10.57861	-5.12668	0.93147
mcmb	3.13091	10.50225	-5.06408	0.91734
wxy	3.11211	10.57285	-5.11851	0.92895
$\tau = 0.75$				
ksm	3.47945	11.36891	-5.41055	0.97327
xy	3.41857	11.30777	-5.36420	0.96378
pwy	3.41879	11.31581	-5.37643	0.96711
mcmb	3.43122	11.26144	-5.32934	0.95660
wxy	3.41785	11.31142	-5.37044	0.96564

SaudI Arabian Study: 95% CIs

Computed 95% confidence intervals of the parameters.

Method	β_0 CI	β_1 CI	β_2 CI	β_3 CI
$\tau = 0.50$				
ksm	(3.10939, 3.20390)	(10.42505, 10.83186)	(-5.36327, -4.96530)	(0.88927, 0.98873)
xy	(3.05842, 3.16840)	(10.28178, 10.85314)	(-5.42069, -4.80405)	(0.84266, 1.01211)
pwy	(3.06065, 3.16361)	(10.29828, 10.85894)	(-5.42727, -4.82609)	(0.84959, 1.01335)
mcmb	(3.06887, 3.19295)	(10.20688, 10.79761)	(-5.37718, -4.75099)	(0.83341, 1.00127)
wxy	(3.05992, 3.16430)	(10.27896, 10.86674)	(-5.44358, -4.79344)	(0.84036, 1.01755)
$\tau = 0.75$				
ksm	(3.42662, 3.53226)	(11.14117, 11.59665)	(-5.63411, -5.18700)	(0.91735, 1.02920)
xy	(3.36006, 3.47708)	(11.00136, 11.61418)	(-5.68867, -5.03973)	(0.87810, 1.04945)
pwy	(3.35644, 3.48113)	(10.99943, 11.63219)	(-5.70829, -5.04457)	(0.88022, 1.05400)
mcmb	(3.36785, 3.49459)	(10.94250, 11.58038)	(-5.66794, -4.99073)	(0.86727, 1.04592)
wxy	(3.36006, 3.47564)	(11.01214, 11.61069)	(-5.68443, -5.05646)	(0.88346, 1.04782)

SaudI Arabian Study: Length of the 95% CIs

Lengths (L) of the computed 95% confidence intervals of the parameters.

Method	β_0 L	β_1 L	β_2 L	β_3 L
$\tau = 0.50$				
ksm	0.09451	0.40680	0.39796	0.09946
xy	0.10998	0.57136	0.61664	0.16945
pwy	0.10296	0.56066	0.60119	0.16377
mcmb	0.12408	0.59073	0.62619	0.16785
wxy	0.10438	0.58777	0.65014	0.17719
$\tau = 0.75$				
ksm	0.10563	0.45548	0.44711	0.11185
xy	0.11701	0.61282	0.64894	0.17136
pwy	0.12468	0.63276	0.66372	0.17378
mcmb	0.12674	0.63789	0.67721	0.17864
wxy	0.11557	0.59856	0.62797	0.16437

Comments

1. For iid errors the link function of the gamma glm is a constant which can be used to estimate the common error variance.
2. The Gamma glm works reasonably well for non-Normal errors.
3. When we have correlated longitudinal data it may be appropriate to use a generalized liner mixed model.
4. When a glm is not appropriate we may use a kernel regression estimator of the conditional variance function. See Hall and Carroll (1989).
5. The methodology is applicable for making inferences with other types of regression models with heteroscedastic errors.
6. The bootstrap enables the estimation of the covariance matric of the parameter estimates.

- Bose A. and Chatterjee S. *Generalized bootstrap for estimators of minimizers of convex functions.* Journal of Statistical Planning and Inference, 2003; **117**:225–239.
- Cantoni E. and Ronchetti E. *Robust Inference for Generalized Linear Models.*, Journal of the American Statistical Association, 2001; **96**:1022–1030.
- Cleveland W.S. *Robust Locally Weighted Regression and Smoothing Scatterplots.* Journal of the American Statistical Association, 1979; **74**:829–836.
- Hall P. and Carroll R.J. *Variance Function Estimation in Regression: The Effect of Estimating the Mean.* Journal of the Royal Statistical Society. Series B (Methodological), 1989; **51**:3–14.
- Kocherginsky M, He X. and Mu Y. *Practical Confidence Intervals for Regression Quantiles.* Journal of Computational and Graphical Statistics, 205; **14**:41–55.
- Koenker R. and Basset G. *Regression Quantiles.* Econometrica, 1978; **46**:33–50.
- Koenker R. and D'Orey V. *Computing Regression Quantiles.* Applied Statistics, 1987; **36**:383–393.
- Koenker R. and D'Orey V. *A Remark on Computing Regression Quantiles.* Applied Statistics, 1993; **43**:410–414.
- Parzen M.I, Wei L.J. and Ying Z. *A Resampling Method Based on Pivotal Estimating Function.* Biometrika, 1994; **81**:341–350.

- Rice J. *Bandwidth Choice for Nonparametric Regression.* The Annals of Statistics, 1984; **12**:1215–1230.
- Nelder J.A. and Lee Y. *Generalized Linear Models for the Analysis of Taguchi-Type Experiments.* Applied Stochastic Models and Data Analysis, 1991; **7**:107–120.
- Nelder J.A. and Wedderburn R.W.M. *Generalized Linear Models.* Journal of the Royal Statistical Society. Series A, 1972; **135**:370–384.
- Silverman B.W., 1986. *Density Estimation.* New York: Chapman and Hall
- Smyth G.K. *Generalized Linear Models with Varying Dispersion.* Journal of the Royal Statistical Society. Series B, 1989; **51**:47–60.
- Smyth G.K. and Jorgensen B. *Fitting Tweedie's Compound Poisson Model to Insurance Claims Data: Dispersion Modelling.*, In Proceedings of the 52nd Session of the International Statistical Institute, August 1999; 10–18.
- Wedderburn R.W.M. *Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method.* Biometrika, 1974; **61**:439–447.