AFT Models and Empirical Likelihood

Mai Zhou

Department of Statistics, University of Kentucky

Collaborators: Gang Li (UCLA); A. Bathke; M. Kim (Kentucky)

- Accelerated Failure Time (AFT) models:

$$Y = \log(T) = X\beta + \epsilon \ .$$

- The responses T, $\log(T)$ or Y are subject to right censoring.

- We study the use of empirical likelihood method (Owen 2001) for inference about β.

Owen (1991) studied the use of empirical likelihood with linear models. (no censoring) He defines the Empirical Likelihood as:

$$EL = \prod_{i=1}^{n} p_i \; ; \qquad \text{where} \qquad \sum p_i = 1; \quad p_i \geq 0 \; .$$

Is this a likelihood of the Y's or is this the likelihood for the ϵ's? Either interpretation is plausible here.

Owen (1991) identified two different linear models from the viewpoint of Empirical Likelihood (EL):

regression models and

correlation models,

which came from Freedman (1981) bootstrap study of linear models.

Correlation Model:

We observe iid random vectors $(Y_i, X_i); i = 1, \ldots, n$ where Y_i are one dimensional responses, X_i are k-dimensional covariates. Unknown parameter β is the minimizer of $g(b) = E \parallel Y - X^T b \parallel^2$.

Regression Model:

The covariates $x_i; i = 1, \ldots, n$ are fixed constants, observable k-dimensional vectors. Y_i are independent random variables with distributions having location parameters $x_i^t\beta$.

The EL for the two types of models appears identical in Owen (1991) as given above, but have different interpretations and require different proofs for the generalized Wilks theorem under different assumptions.

Owen (1991) showed the EL Wilks Theorem holds for both models:

$$-2 \log \frac{\max EL}{\prod 1/n} \longrightarrow \chi^2$$

where the max of EL is taken over p_i such that

$$0 = \sum X_i \psi(Y_i - X_i \beta_0) p_i \ .$$

Notice for a given (non-censored) data set, under either regression or correlation model the two EL ratios have the same value and the least squares estimator $\widehat{\beta}$ is also the same under either model.

So the two different models only pertain to two different sets of assumptions, under which the EL Wilks theorem hold.

With censored responses, not all Y_i are available. Estimation equations, as well as the empirical likelihood needs to be changed.

Censored data: randomly right censoring:

$$Q_i = \min(Y_i, C_i) \ ; \quad \delta_i = I[Y_i \le C_i]$$

where C_i are the censoring times.

We see two different ways of defining the empirical likelihood for censored data, corresponding to the two types of linear model:

• Define the empirical likelihood based on the censored 'residuals': $e_i(b) = Q_i - X_i b$; δ_i:

$$EL_{(res)} = \prod_{i=1}^{n} (p_i)^{\delta_i} \left(1 - \sum_{e_j \le e_i} p_j \right)^{1-\delta_i} .$$

Notice the order in the summation is based on the residuals. This empirical likelihood is more suitable for regression models.

• The second approach is to consider the empirical likelihood for (censored) vectors: (Q_i, δ_i, X_i): which we shall call it case-wise EL.

$$EL_{(case)} = \prod_{i=1}^{n} (p_i)^{\delta_i} \left(1 - \sum_{Q_j \le Q_i} p_j \right)^{1-\delta_i} \quad .$$

Notice the order in the summation of the EL is based on Q's. This empirical likelihood is more suitable for correlation models.

The two definitions of the EL is parallel to the two ways of bootstrapping in linear models:

(1) re-sample the residuals e_i, for the regression models.

(2) re-sample the vectors or cases (Y_i, X_i), for the correlation models.

See for example Freedman (1981).

We consider two types of estimates $\hat{\beta}$ suitable for the two types of EL and AFT models:

(1) Buckley-James (1979) estimate. Rely on the fact that the errors (residuals) are iid. More suitable for the regression models.

(2) Case weighted estimators (Koul, Susarla, Van Ryzin (1982) [not 1981 !], Zhou (1992), Stute (1993, 1994, 1999); Huang, Ma, Xie (2005) etc.) Idea is based on the

(iid) cases (Y_i, X_i) or (Q_i, δ_i, X_i). More suitable for the correlation models.

Let us first consider the **Buckley-James** estimator and $EL_{(res)}$.

Let $e_i(b) = Q_i - X_i b$, be the (censored) residuals.

The Buckley-James estimator of β is the solution to the estimation equation

$$0 = \sum_{i=1}^{n} X_i \left\{ \delta_i e_i(b) + (1-\delta_i) \sum_{j: e_j > e_i} \frac{e_j(b) \Delta \hat{F}(e_j)}{1 - \hat{F}(e_i)} \right\} \ ,$$

where $\widehat{F}(\cdot)$ is the Kaplan-Meier estimator computed from $(e_i(b), \delta_i)$.

The expression in blue is an estimation of the conditional expectation, $\widehat{E}(Y_i - X_i b | Q_i, \beta = b, \delta_i = 0)$.

There are two summation signs in the above equation. (one for index i one for index j). Exchange the order of the summations we have

$$0 = \sum_{j=1}^{n} \delta_j e_j(b) \left\{ X_j + \sum_{i: e_i < e_j, \delta_i = 0} \frac{X_i \Delta \widehat{F}(e_j)}{1 - \widehat{F}(e_i)} \right\} .$$

This form of Buckley-James estimation equation will lead to the constraint equations we use with the censored empirical likelihood EL_{res}.

The above defined $EL_{(res)}$ is to be maximized with and without the following added constraint equations (with $b = \beta_0$):

$$0 = \sum_{j=1}^{n} p_j \delta_j e_j(b) \left\{ X_j + \sum_{i: e_i < e_j, \delta_i = 0} \frac{X_i \Delta \hat{F}(e_j)}{1 - \hat{F}(e_i)} \right\} \frac{1}{\Delta \hat{F}(e_j)} \; .$$

Notice the same p_i appears in this constraint equations and in the definition of $EL_{(res)}$.

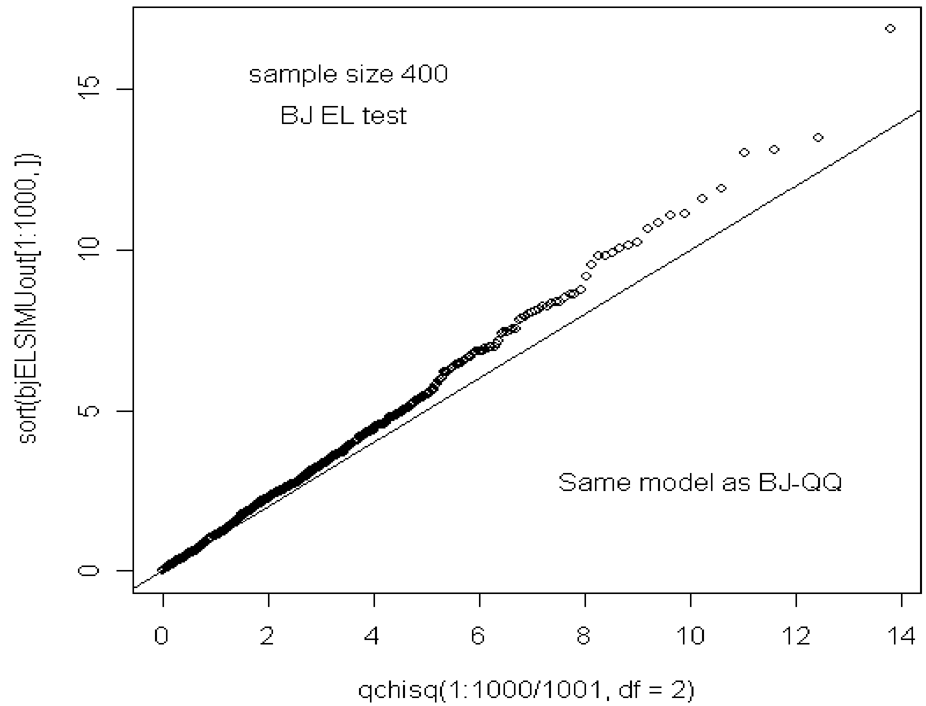
Without this constraint, the EL_{res} is maximized at the Kaplan-Meier estimator, computed from $e_i(\beta_0), \delta_i$.

Under null hypothesis, i.e. when $b = \beta_0$, the $-2logEL_{res}$ ratio has asymptotically a chi squared distribution:

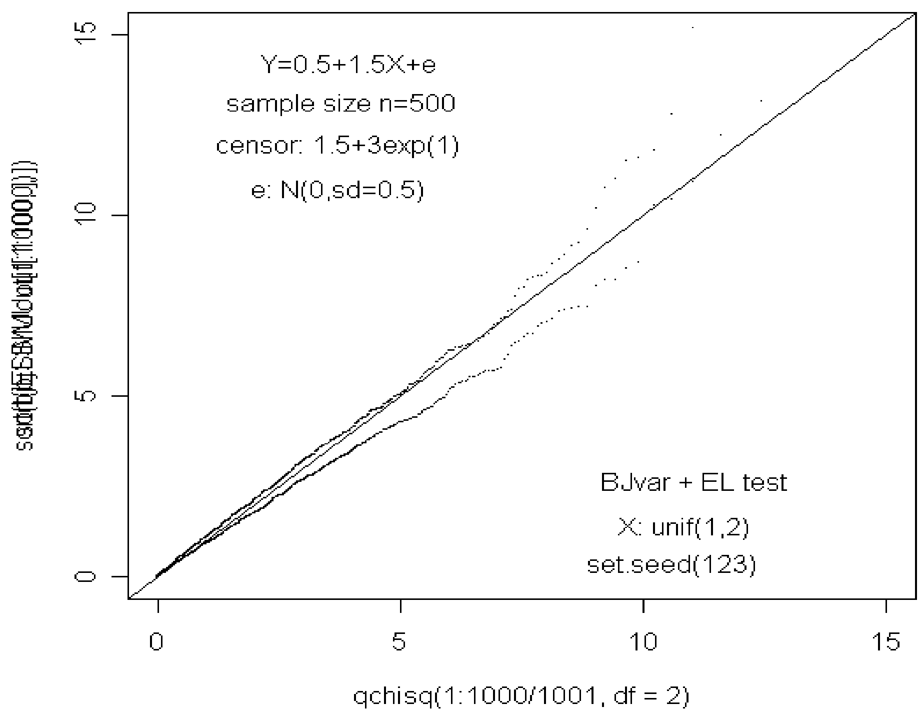
$$-2 \log \frac{\max EL_{res}}{\max EL_{res}} \longrightarrow \chi^2$$

Proof — See Zhou and Li (2004).

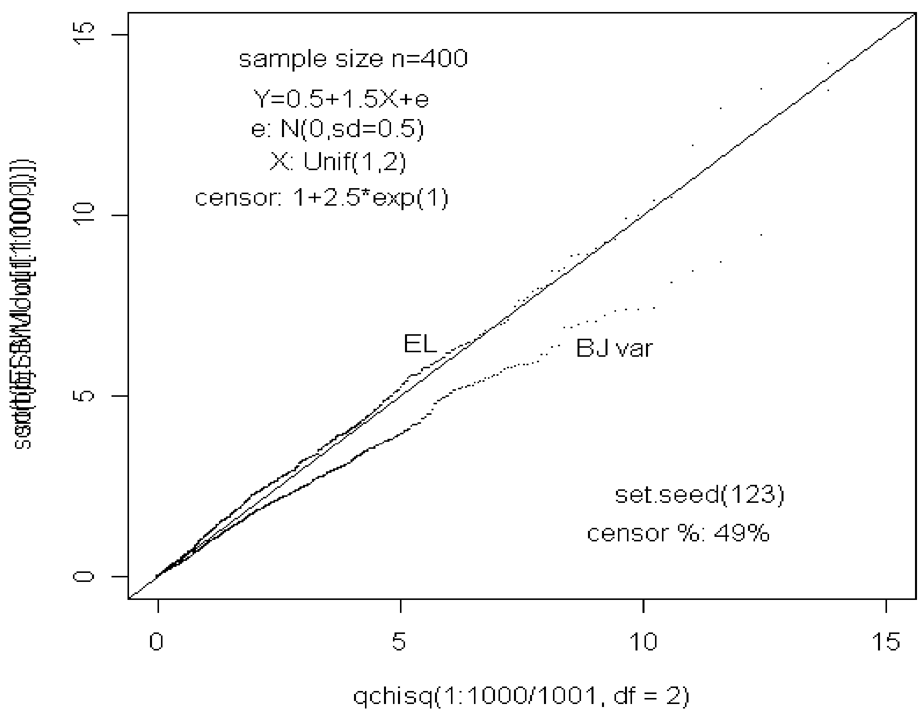
Simulation confirmation of the chi squared limit:



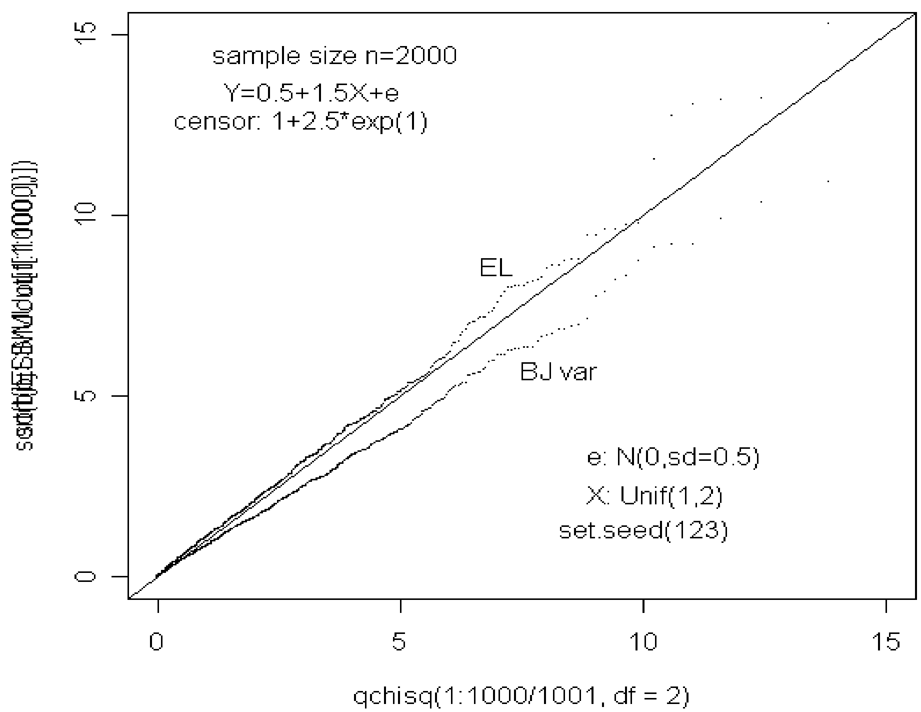
Q-Q plot of $-2 \log ELR$



Q-Q plot of $-2\log ELR$



Q-Q plot of $-2 \log ELR$



Q-Q plot of $-2\log ELR$

Now let us consider the **correlation model and the case weighted estimator**.

Many authors have proposed and studied the case-weighted estimator, including large sample property of the estimator under various assumptions. The earliest reference I can find is Koul, Susarla and Van Ryzin (1982), the latest reference is Huang, Ma and Xie (2005). In between there are: Zhou (1992), Stute (1993, 1994, 1999), Gross and Lai (1996), and van der Laan and Robins (2003): in their book they call this "inverse censoring probability weight": since

$$\Delta \hat{F}(Q_i) = \frac{\delta_i}{1 - \hat{G}(Q_i)} = w_i$$

The case weighted estimator is defined by the *estimating equations* (cf. Zhou 1992)

$$0 = \sum_{i=1}^{n} w_i \delta_i X_i \psi(Q_i - X_i b)$$

where w_i is the jump of the Kaplan-Meier estimator at Q_i computed from (Q_i, δ_i).

$\psi()$ is monotone, usually the derivative of $\rho()$.

The estimator $\widehat{\beta}$ is easy to obtain, no iteration needed.

The *constraint equations* to work with the EL_{case} are

$$0 = \sum_{i=1}^{n} p_i \delta_i X_i \psi(Q_i - X_i b) \ .$$

The $EL_{(case)}$ is to be maximized with and with out the above constraint equations.

We have a chi square limit theorem (Wilks):

Under null hypothesis, we have

$$-2 \log \frac{\max EL_{case}}{\max EL_{case}} \longrightarrow \chi^2$$

in distribution.

In the denominator, the max of EL_{case} occurs when $p_i =$ jumps of the Kaplan-Meier estimator based on Q_i, δ_i.

If we interprets the constraint equations as constraints on the marginal DF of Y, then (since the EL is also on the Y marginal), existing results apply.

The constraint equation can be thought of as a constraint on the marginal CDF of the Y

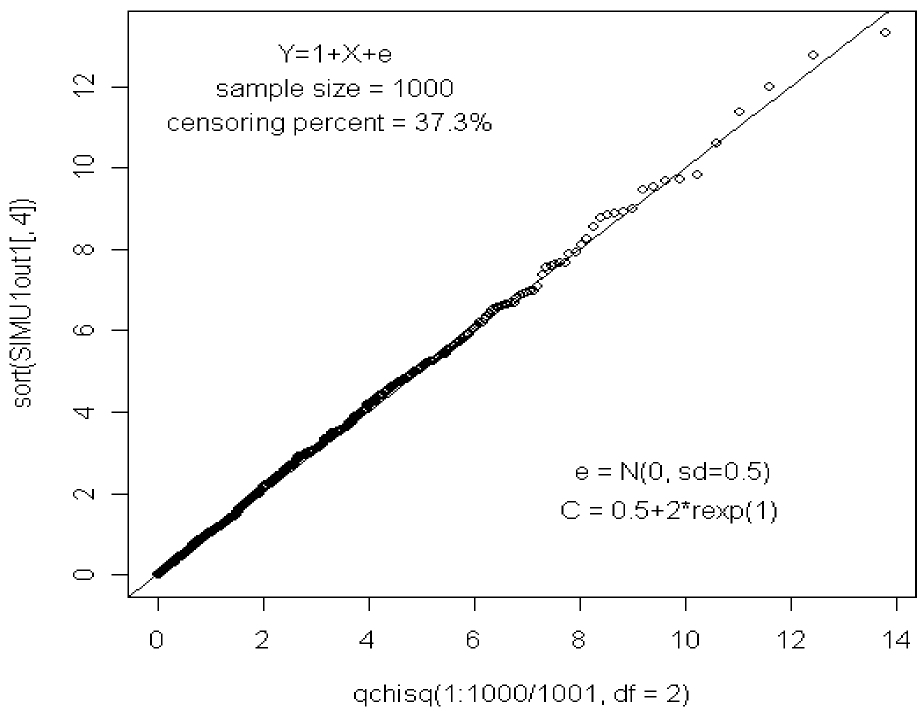
$$\int g(y, x) dF_Y(y) = 0$$

Censored Data AFT Models

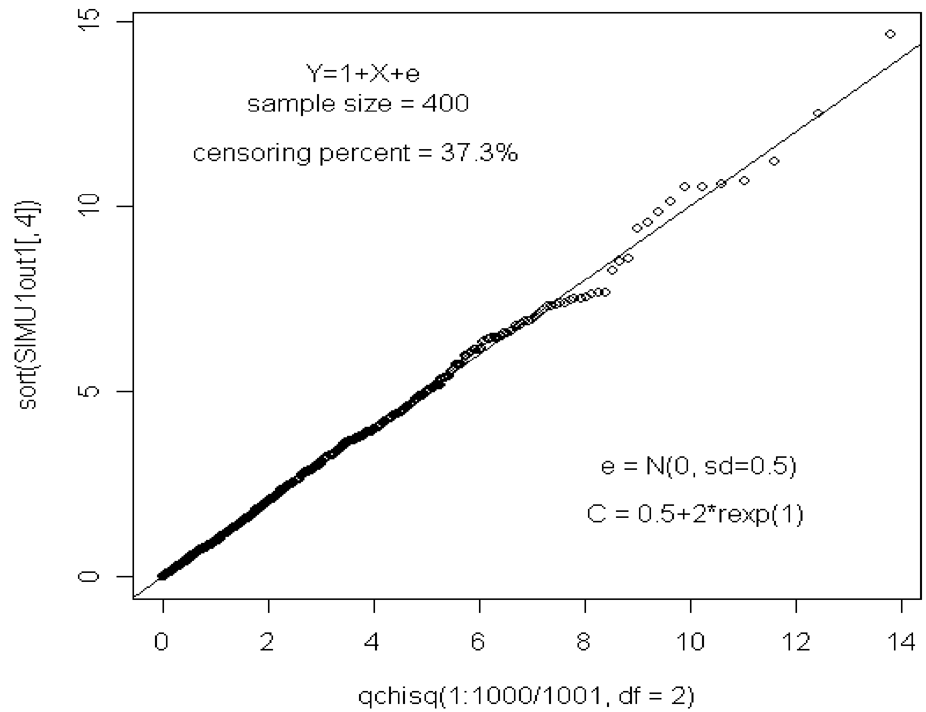
Regression Model	Correlation Model
Buckley-James Est.	Weighted Est.
EL(residual)	EL(case)
Wilks Theorem	Wilks Theorem

Notice with censored data, the two estimators are different and the two EL are also different. Yet they both have EL ratio converge to chi-square distribution under certain conditions.

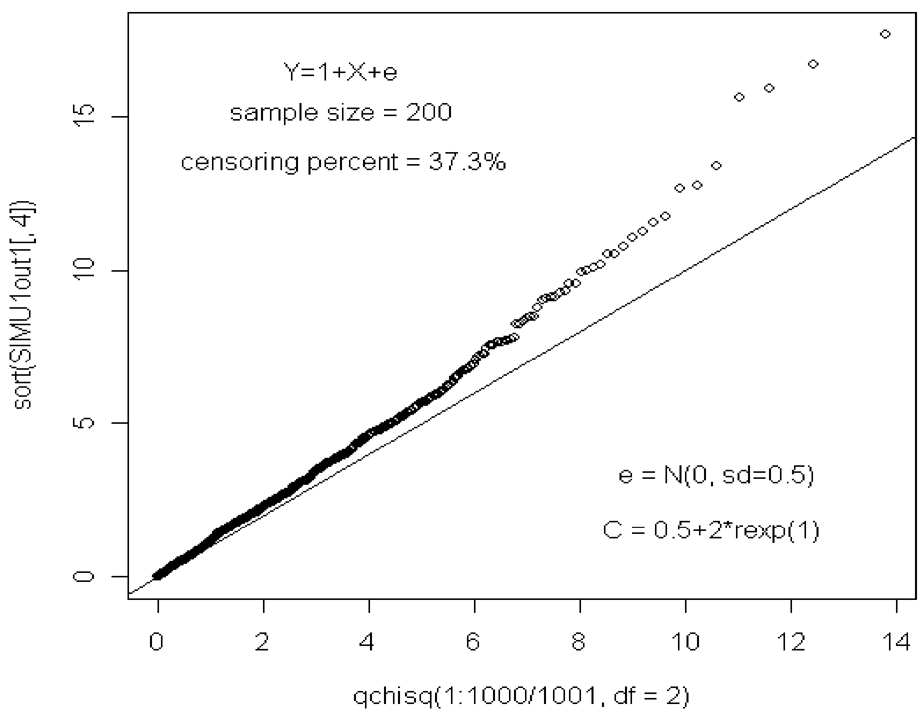
Simulation results for correlation AFT models:



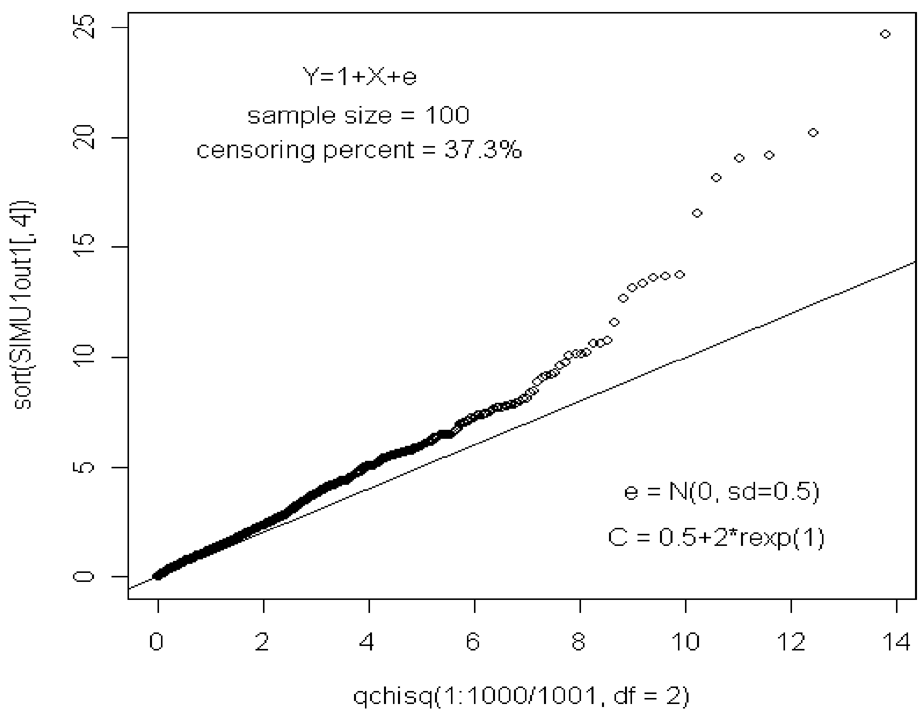
Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$

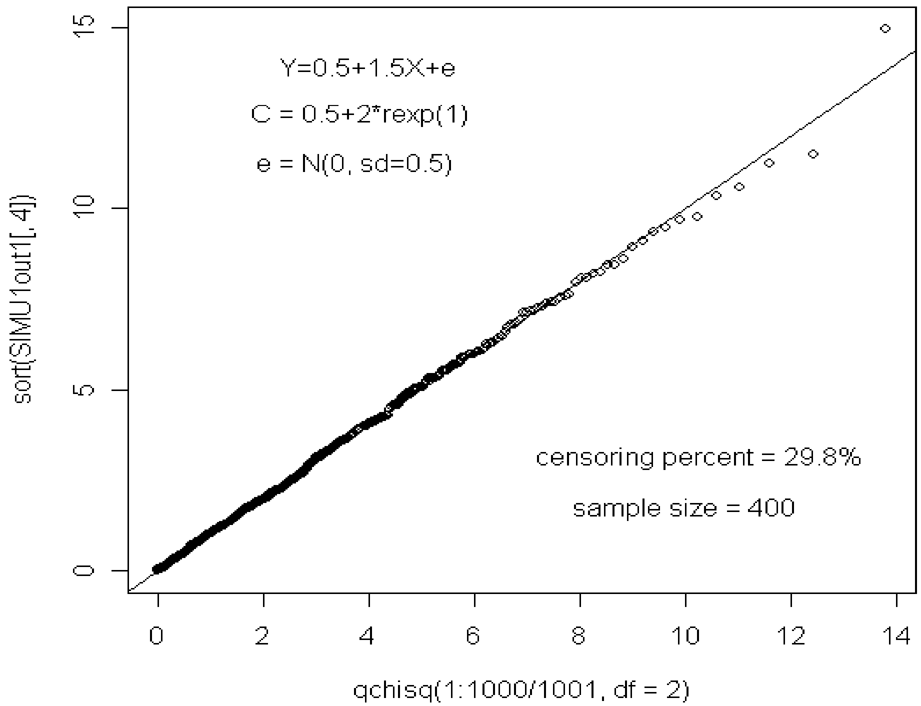


Q-Q plot of $-2\log ELR$

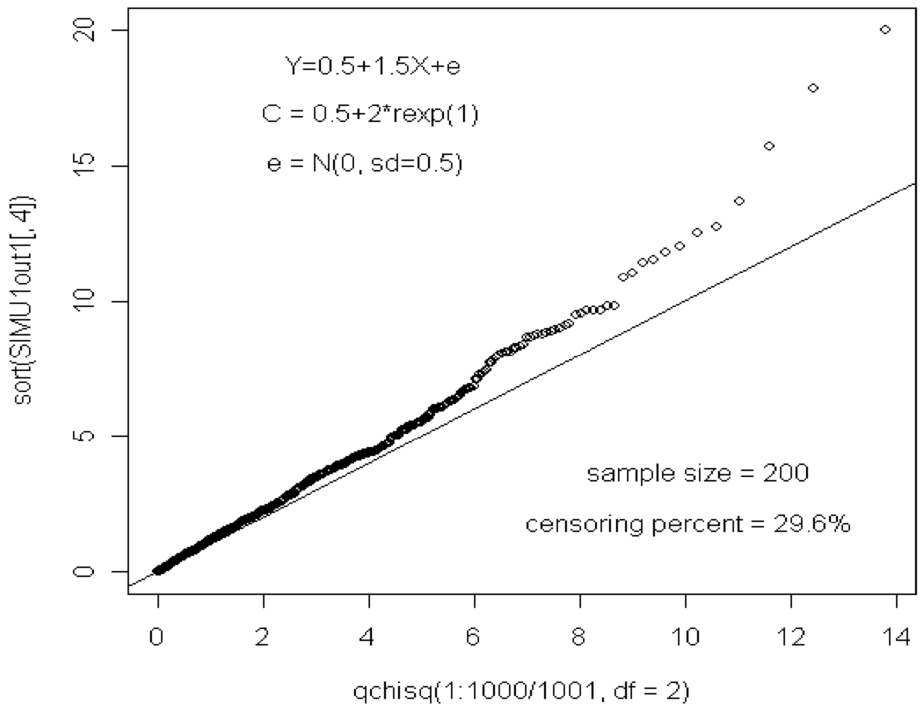


Q-Q plot of $-2\log ELR$

Less censoring:



Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$

Quantile regression is an important model in econometrics (Koenker 2005)

$$\text{Quantile of } \; Y_i = \beta X_i$$

Estimate β by

$$\arg\min \sum_{i=1}^{n} \rho(Y_i - \beta X_i)$$

where ρ is the so called ‘check function’.

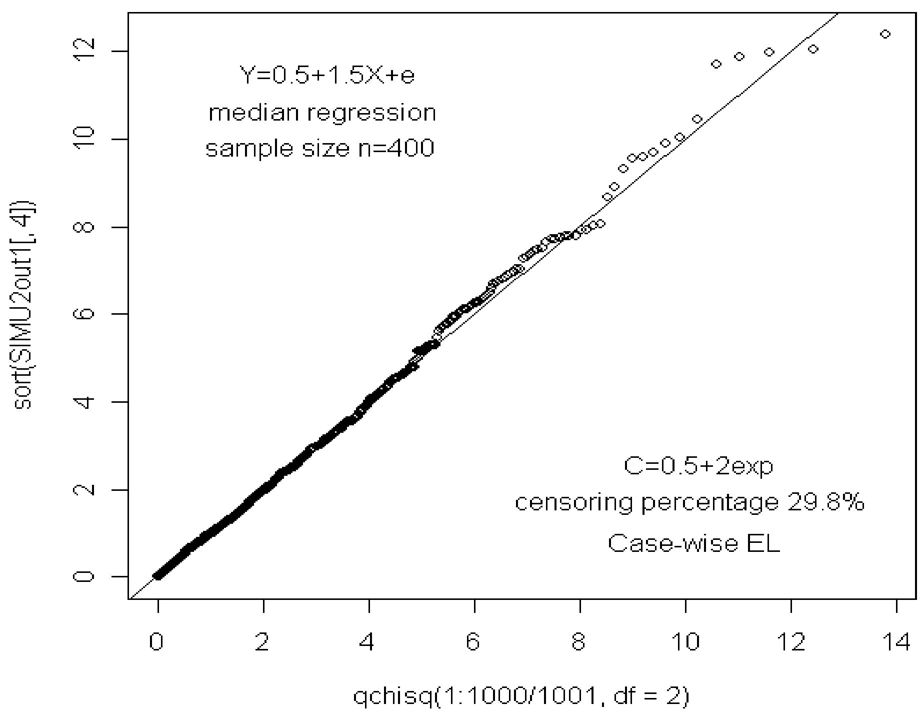
Estimating equation for β

$$0 = \sum_{i=1}^{n} X_i \psi(Y_i - \beta X_i)$$

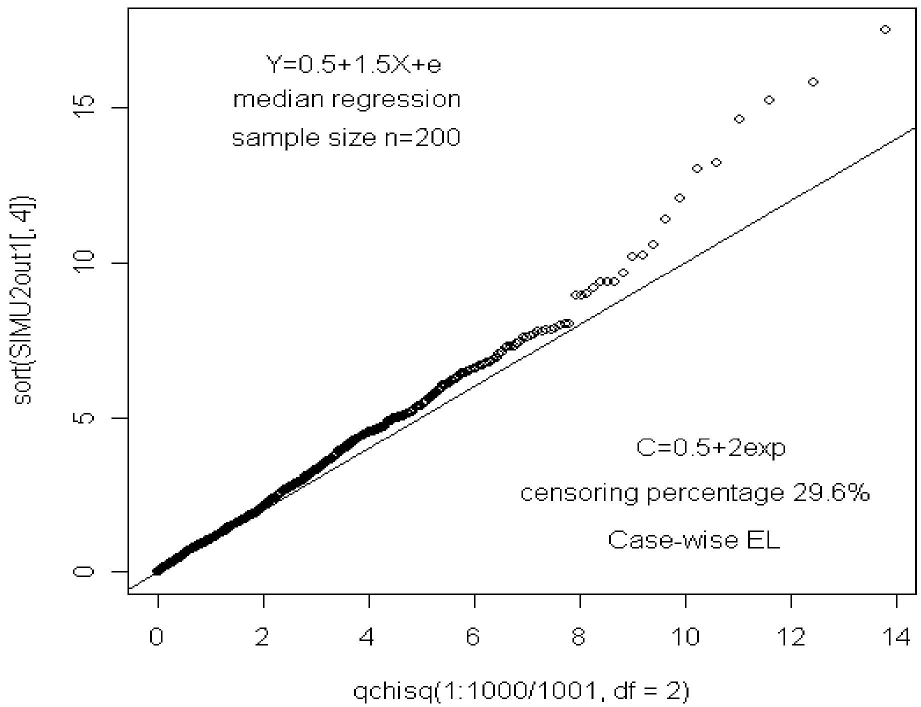
where ψ needs to be defined carefully.

Median regression: $\psi(t) = 1, 0$ or -1 depend on $t > 0, t = 0$ or $t < 0$.

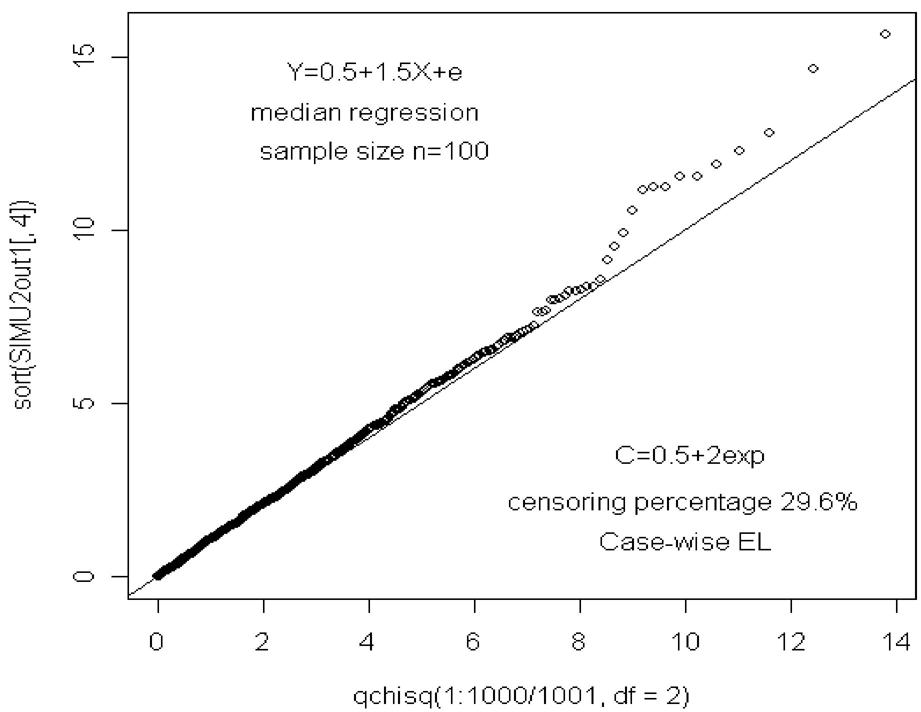
Variance of the $\widehat{\beta}$ is hard to estimate. Using EL, we do not need to estimate the variance of $\widehat{\beta}$.



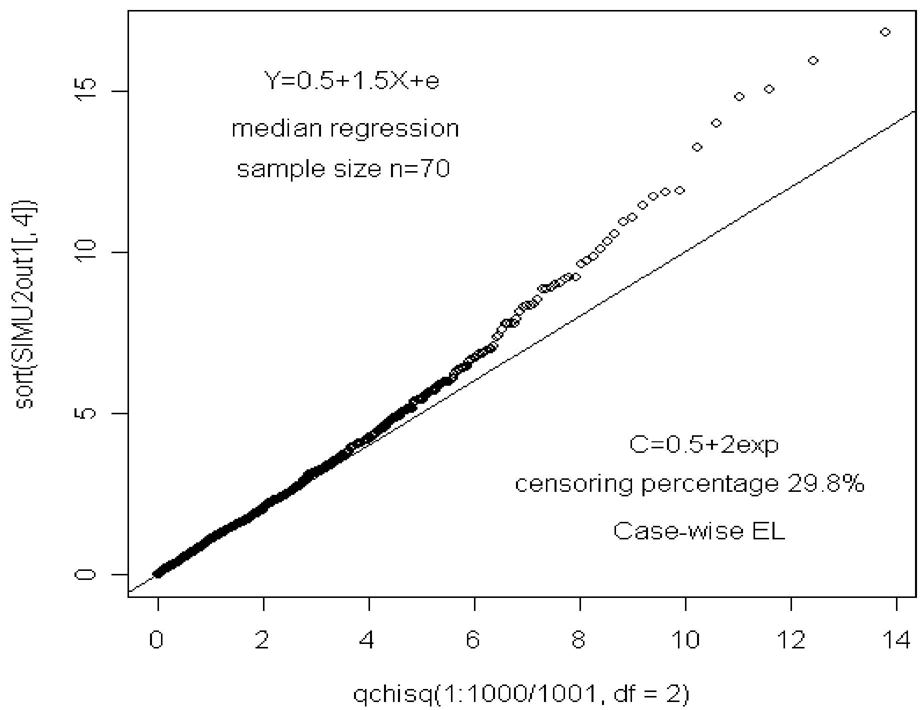
Q-Q plot of $-2\log ELR$



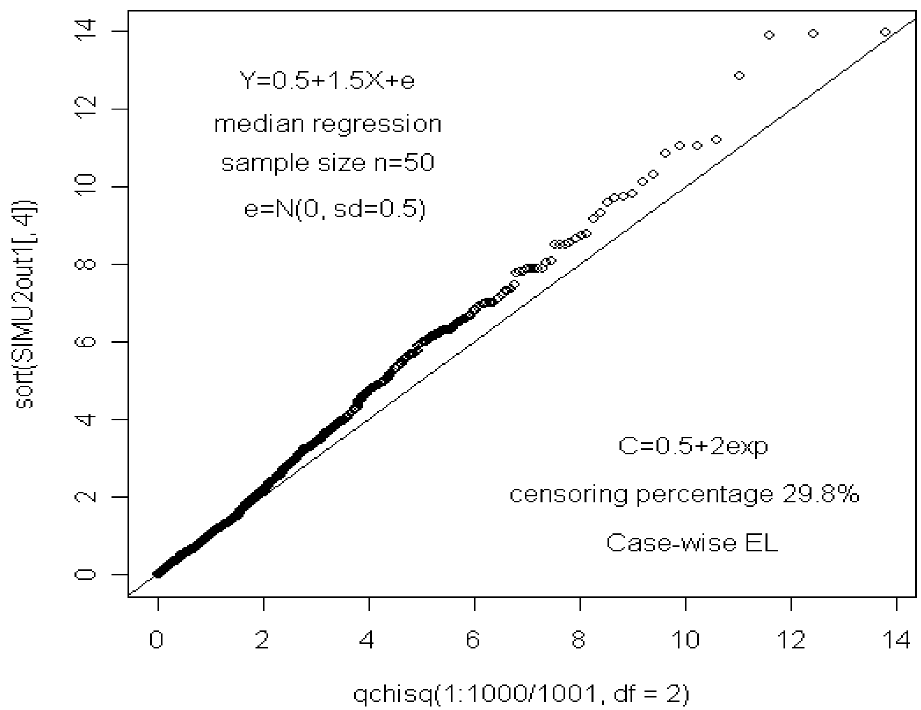
Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$

25% Quantile regression: $\psi(t) = 0.5, -0.5$ or -1.5 depend on $t > 0, t = 0$ or $t < 0$.

75% Quantile regression: $\psi(t) = 1.5, 0.5$ or -0.5 depend on $t > 0, t = 0$ or $t < 0$.

These functions equal to 2 times the derivative of the so called "check functions" defined in Koenker ≪Quantile Regression≫. We take this definition so that the median regression ψ is the more commonly used -1/+1 function.

The (heteroscedastic) model we are simulating next is generated by

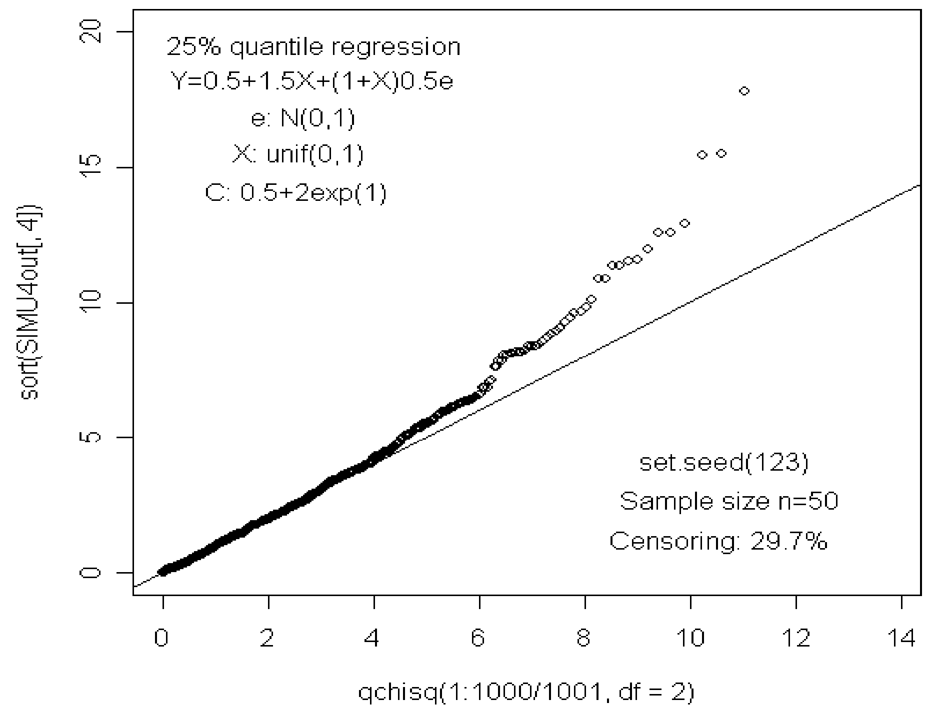
$$Y = 0.5 + 1.5X + 0.5(1 + X)\epsilon$$

where ϵ is iid Normal(0,1). X is uniform (0,1) and Y are right censored by $0.5 + 2exp(1)$.

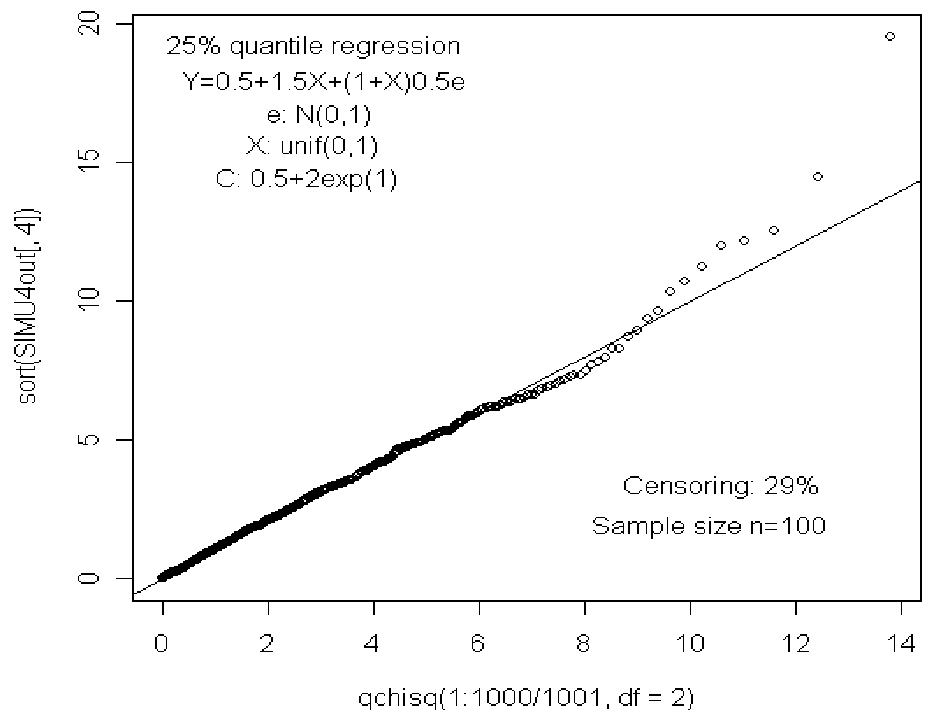
For 25% quantile regression, the true regression line is $Y = 0.162755 + 1.162755X$

For 75% quantile regression, the true regression line is $Y = 0.837245 + 1.837245X$

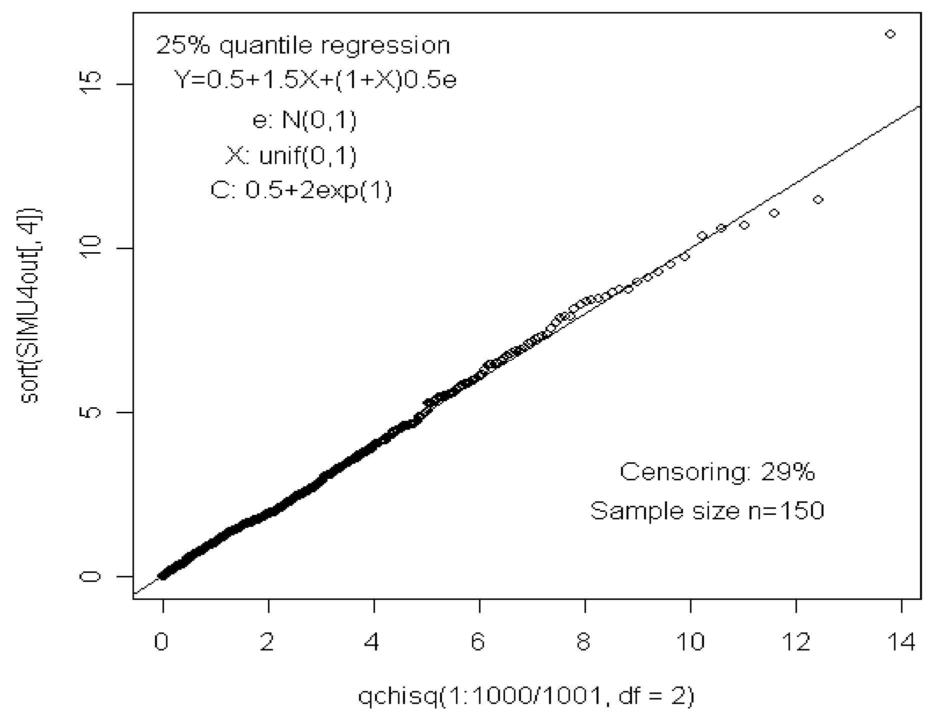
From the plot we see that the 25% quantile regression is less sensitive to the right censoring, compared to median regression and 75% quantile regression.



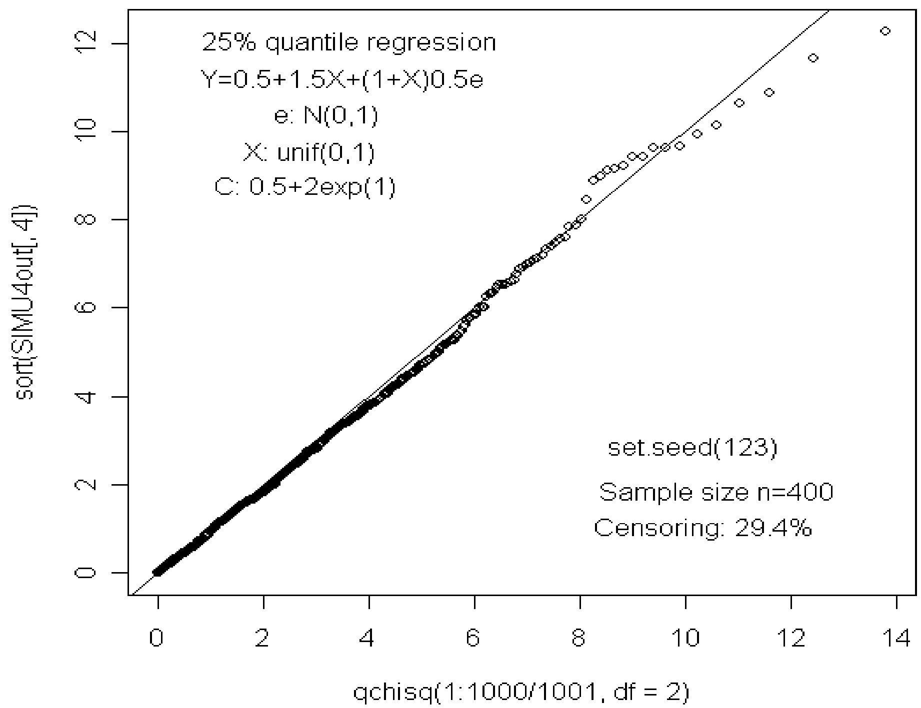
Q-Q plot of $-2\log ELR$



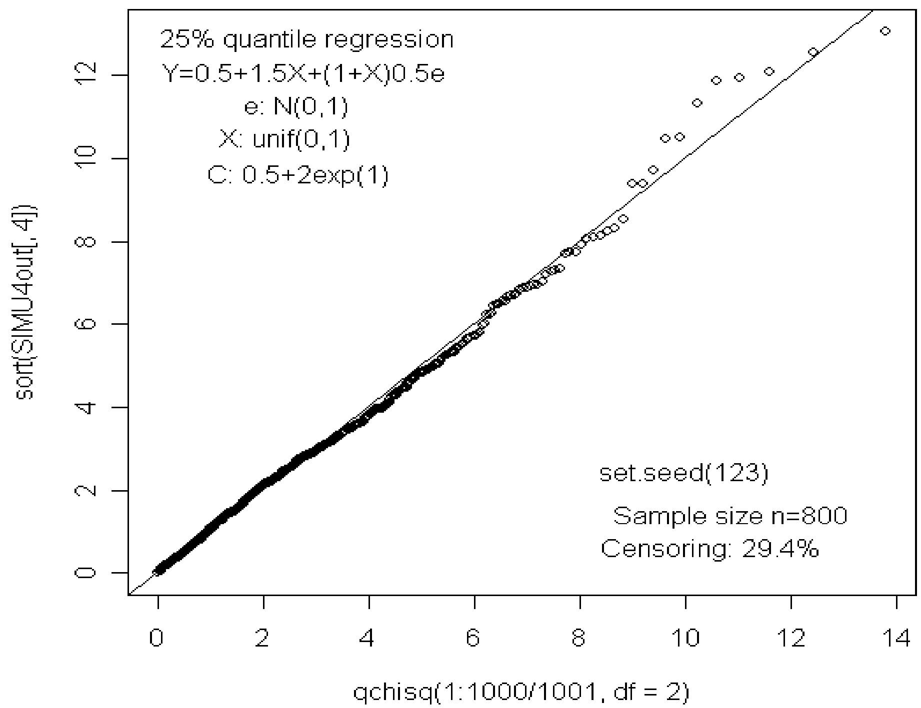
Q-Q plot of $-2\log ELR$



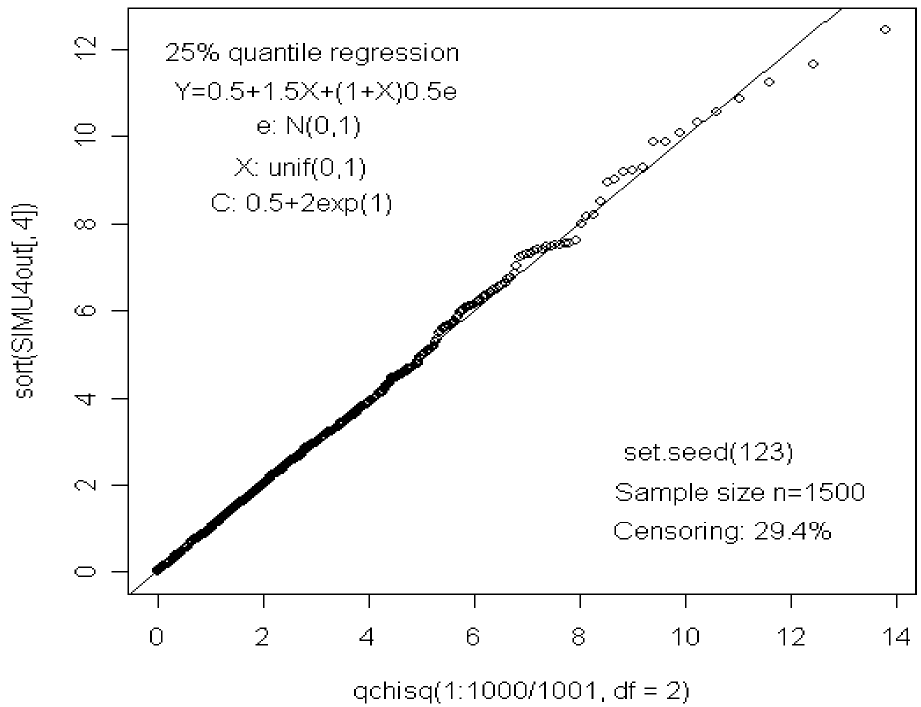
Q-Q plot of $-2\log ELR$



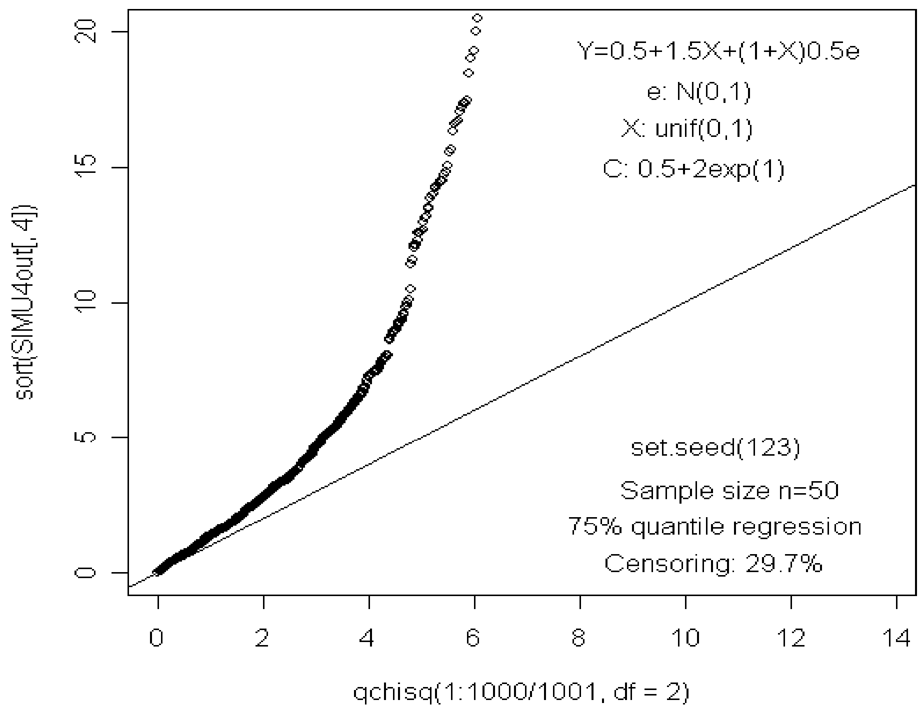
Q-Q plot of $-2 \log ELR$



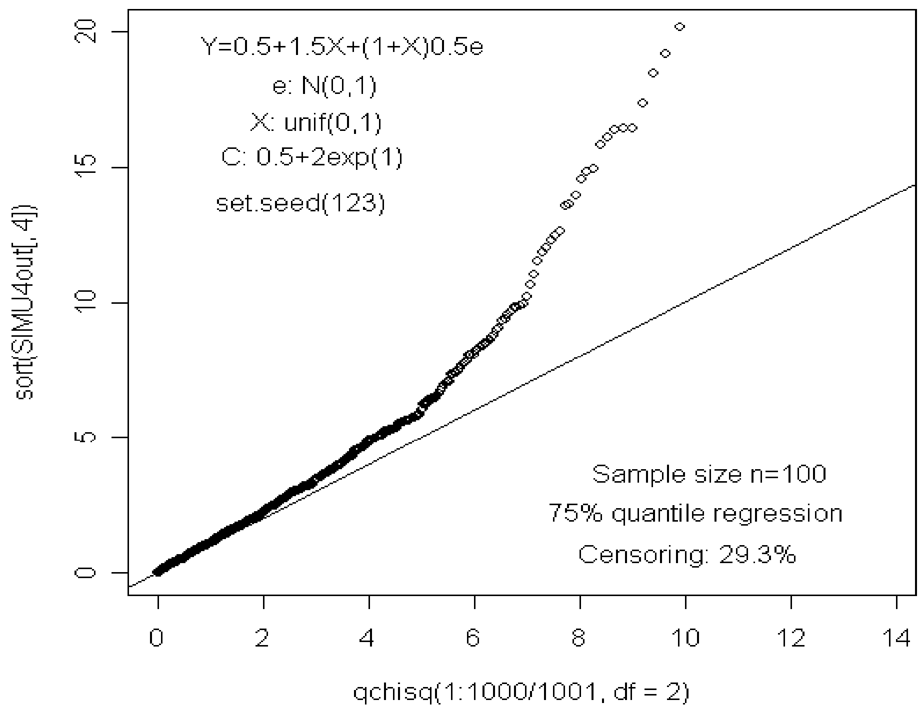
Q-Q plot of $-2\log ELR$



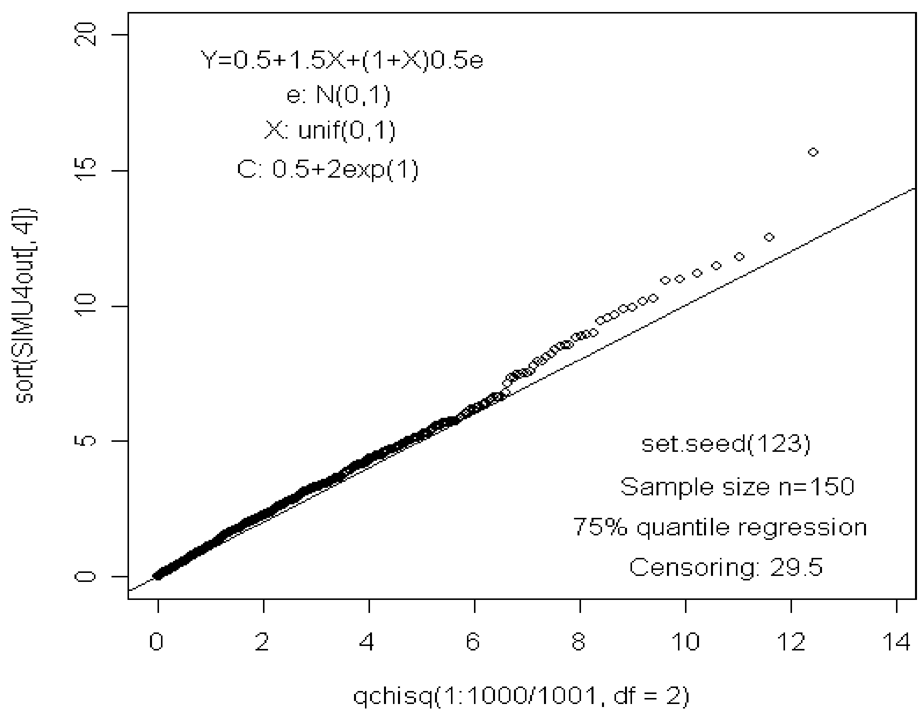
Q-Q plot of $-2\log ELR$



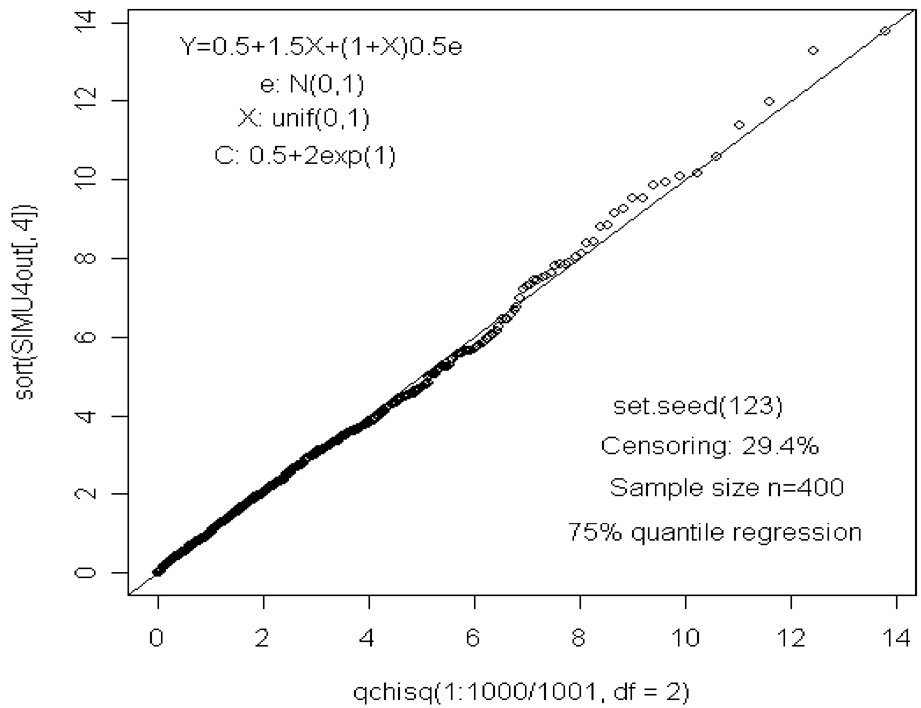
Q-Q plot of $-2\log ELR$



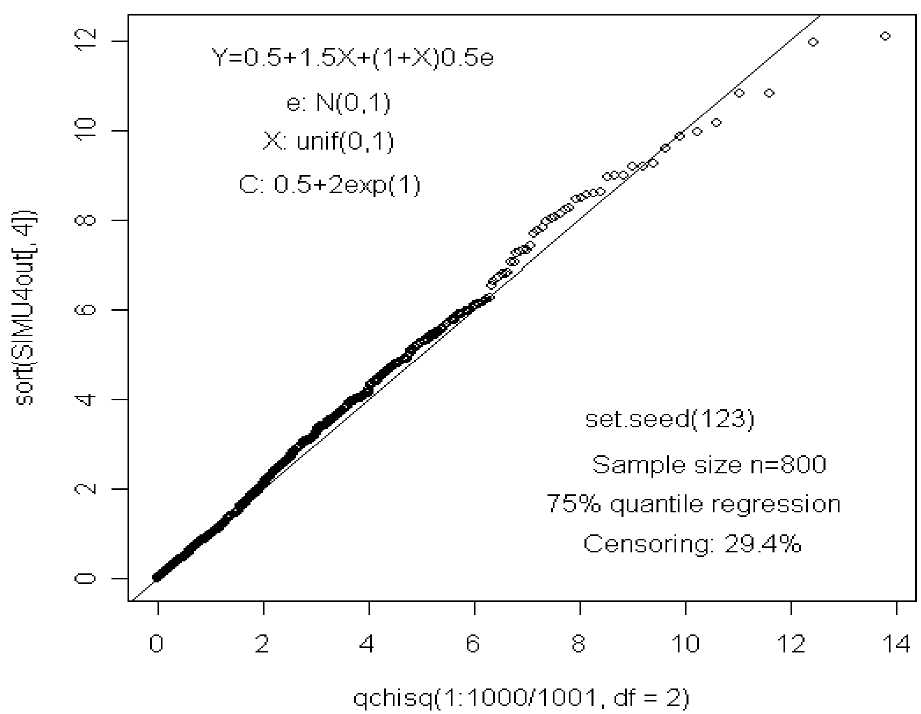
Q-Q plot of $-2\log ELR$



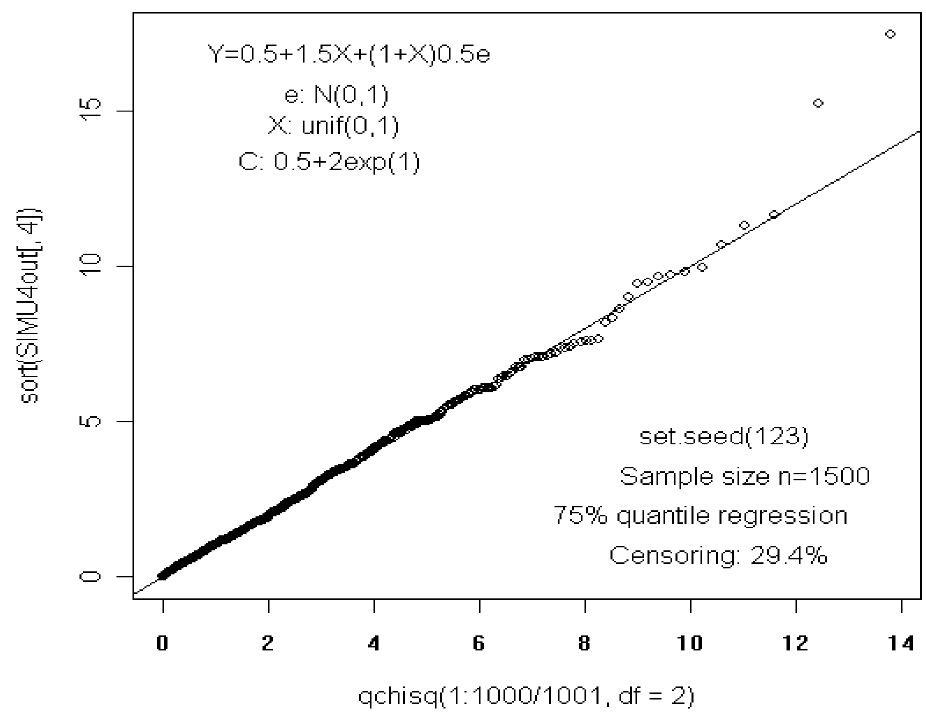
Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$



Q-Q plot of $-2\log ELR$

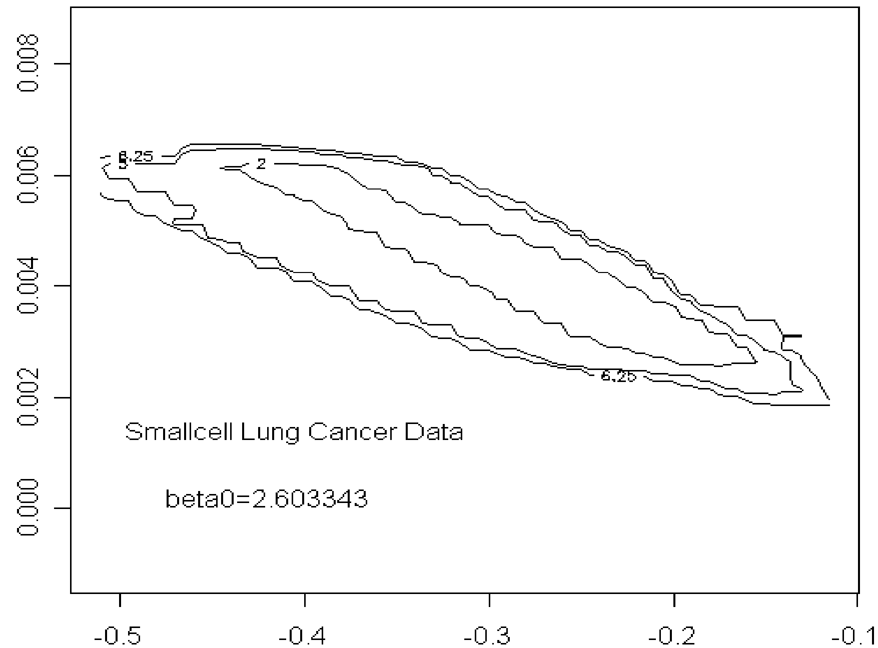


Q-Q plot of $-2\log ELR$

Example: Smallcell Lung Cancer Data (Ying, Jung & Wei 1995)

$$\log(T) = \beta_0 + X_1\beta_1 + X_2\beta_2 + \sigma(X_1, X_2)\epsilon$$

X_1 = indicator of treatment; X_2 = age at entry.



Contour plot of $-2\log ELR$ *for* $\beta_1\beta_2$; β_0 *fixed.*

Comments about heteroscedasticity. The correlation model can accommodate some heteroscedasticity.

When Y is censored, X can be missing.

Censoring: C independent of Y. (stronger than conditional independence usually assumed by regression model and Buckley-James estimator). But Buckley-James assumes iis error.

References

Buckley, J. J. and James, I. R. (1979). Linear regression with censored data. *Biometrika*, **66** 429-36.

Freedman, D.A. (1981). Bootstrapping regression models. *Annals of Statist.*, **9**, 1218-1228.

Gross, S. and Lai, T.L. (1996). Nonparametric estimation and regression analysis with left truncated and right censored data. *J. Amer. Statist. Assoc.* **91**, 1166-1180.

Huang, J., Ma, S. and Xie, H. (2005) Least Absolute Deviations Estimation for the Accelerated Failure Time Model. Tech Report #350 Univ. of Iowa, Dept. Stat. and Actuarial Science.

Koul, H., Susarla, V, and Van Ryzin, J. (1982). Least squares regression analysis with censored survival data. In *Topics in Applied Statistics*, Ed.Y.P. Chaubey and T.D. Dwivedi, pp 151-165. Marcel Dekker, New York.

Laan, M. van der, and Robins, J. M. (2003) *Unified Methods for Censored Longitudinal Data and Causality.* Springer, New York.

Owen, A (1991), Empirical likelihood for linear models. *Annals of Statist.*, **19**, 1725-1747.

Owen, A. (2001). *Empirical likelihood*. Chapman & Hall, London.

Stute, W. (1999). Nonlinear Censored Regression. *Statistica Sinica*, **9, 4**, 1089–1102.

Jin, Z, Lin, D. Y. and Ying, Z. (2006). On least-squares regression with censored data *Biometrika* **93** 147-161

Zhou, M. (1992). M-estimation in censored linear models. *Biometrika*, **79**, 837-841.

Zhou, M. and Li, G. (2004). Empirical likelihood analysis of the Buckley-James estimator.