Empirical Likelihood Analysis for the Heteroscedastic Accelerated Failure Time Model

Mai Zhou, Arne Bathke, Mi-Ok Kim *

Abstract

We study M-estimation methods and inference for the heteroscedastic AFT (Accelerated Failure Time) model in which the data is generated by a correlation model. We propose a casewise weighted scheme for the M-estimation to cope with random right censoring. Empirical likelihood is used for inference. We show that a nonparametric version of Wilks' theorem holds for the resulting empirical likelihood ratio.

The results are applicable also in censored quantile regression with heteroscedastic errors.

Simulations and an example illustrate the versatility of the method.

Key Words: Regression model; Survival analysis; Wilks' theorem; Buckley-James estimator; Censored data; Quantile regression.

*Mai Zhou is Associate Professor, and Arne Bathke and Mi-Ok Kim are Assistant Professors, Department of Statistics, University of Kentucky, Lexington, KY 40506-0027, USA (email: mai@ms.uky.edu). This work was partially supported by National Science Foundation grant DMS-0604920.

1 INTRODUCTION

1.1 Empirical Likelihood

The Empirical Likelihood (EL) method (Owen, 2001) inherits the advantages of the likelihood ratio tests but is applicable to nonparametric statistical models. EL was first proposed by Thomas and Grunkemeier (1975) to obtain better confidence intervals related to the Kaplan-Meier (1958) estimator. Owen (1988, and subsequent work) made it into a general methodology.

EL is a versatile nonparametric inference method not unlike the bootstrap. Many authors have demonstrated that the EL method and the nonparametric version of Wilks' theorem apply to estimators in various statistical models.

Empirical likelihood analysis for linear regression models was studied by Owen (1991). Owen identified two different types of linear models from the viewpoint of EL: regression models and correlation models. Following Freedman (1981) and Owen (1991), the main characteristics of these two model concepts can be summarized as follows.

Correlation Model:

We observe i.i.d. random vectors $(Y_i, X_i); i = 1, \ldots, n$, where the Y_i are one-dimensional responses and the X_i are (column vectors of) k-dimensional covariates. The unknown parameter β relating X and Y is the minimizer of $g(b) = E\rho(Y - X^t b)$, where $\rho(\cdot)$ is a convex loss function (e.g., quadratic loss or absolute loss.). It is usually assumed that $E(XX^t)$ is positive definite. If bootstrap methods are applied in a correlation model, the appropriate resampling approach consists of resampling the vectors, and not the residuals. A correlation model is used if, for example, the goal is to estimate a regression plane on the basis of a simple random sample of multivariate observations. Apart from measurement error in the response, there may be misspecification error in the covariates.

Regression Model:

The covariates $x_i; i = 1, \ldots, n$ are fixed, constant, and observable k-dimensional (column) vectors, forming a matrix of full rank. The responses Y_i are independent random variables with distributions having location parameter $x_i^t\beta$. An appropriate bootstrap for the regression model would resample the centered residuals. In general, regression models are used if measurement error of the response is the main source of uncertainty.

Typically, in regression models, the errors are assumed homoscedastic, whereas the conditional distribution of the error term, given X, in a correlation model depends on X. That is, the errors in a correlation model are typically heteroscedastic. If the errors in a correlation model are homoscedastic, then the asymptotics are the same as in a regression model (Freedman, 1981).

Notice that for a given data set, when there is no censoring, empirical likelihood ratios have identical values under either regression or correlation model (Owen, 1991). Thus, for uncensored data, the two different models mainly pertain to two different sets of assumptions under which the empirical likelihood Wilks theorem holds. They yield identical p-values or confidence intervals, though interpreted differently. However, under censoring, the two concepts lead to different estimators and empirical likelihood ratios, as we shall discuss in more detail below.

In this manuscript, we introduce a new empirical likelihood approach to linear models with censored data in a correlation model context. Thus, we assume random covariates and allow for heteroscedastic errors. The proposed method can also be applied to censored quantile regression, as demonstrated below.

1.2 Linear Model With Censored Data - AFT Model

The Accelerated Failure Time (AFT) model postulates that the location parameter of the log survival time T_i is a linear function of the covariate X_i, either assumed fixed (regression model) or random (correlation model). AFT models provide a competitive and flexible alternative to the Cox (1972) proportional hazards model. In some sense, AFT models complement the proportional hazards model, and they do have some advantages over it (see, e.g., Wei, 1992; Reid, 1994). Furthermore, an AFT type analysis can yield a more straightforward interpretation of results, such as quantification of survival times instead of the more abstract hazard rates. However, it is not as popular as it should be, mostly because inference for the AFT model has been difficult. Jin et al. (2003, p.342) write that "despite the theoretical advances, semiparametric methods for the accelerated failure time model have rarely been used in applications, mainly because of the lack of efficient and reliable computational methods". A great advantage of the empirical likelihood approach that we pursue in this manuscript is its straightforward and reliable implementation using the R package 'emplik', as demonstrated below.

We consider in this paper the correlation model in which i.i.d. vectors $(Y_i, X_i); i = 1 \ldots, n$, are observed. As a special case of the correlation model, we assume here that the following relation exists between the i.i.d. covariates X_i and the response.

$$Y_i = \log T_i = X_i \beta_0 + \sigma(X_i)\epsilon_i \tag{1}$$

where X_i are i.i.d. covariates and ϵ_i represents the i.i.d. error component. We term this model *AFT correlation model*. Note that through the term $\sigma(X_i)$, this model explicitly allows for heteroscedastic errors. The assumptions on the distributions of ϵ_i, X, and the functions $\sigma(\cdot)$ are such that $E\rho(X^t r + \sigma(X)\epsilon)$ is finite and has a unique minimum at $r = 0$. Here, $\rho(\cdot)$ is a non-negative, convex loss function (see the minimization criterion (3) below).

We further suppose that the responses Y_i are subject to censoring. Denote the censoring times by C_i. Assume that the C_i are independent of (Y_i, X_i). This independent censoring assumption is stronger than the conditional independence assumption stated usually for rank based estimators (see Kalbfleisch and Prentice, 2002, Chapter 7 and references therein, or Jin et al., 2003) or the Buckley-James estimator (Buckley and James, 1979; Lai and Ying 1991). However, the latter estimators in turn require homoscedastic errors whereas the model we propose in (1) allows for heteroscedasticity. We observe $Z_i = \min(Y_i, C_i)$ and $\delta_i = I[Y_i \leq C_i]$. X_i is assumed observed for all the uncensored observations.

Many different estimation methods have been proposed for the censored data AFT model. Two popular approaches in the literature are the just mentioned rank based estimators and the Buckley-James estimator. However, both approaches are geared towards the AFT *regression* models and assume homoscedasticity. For an empirical likelihood approach to those models, see Zhou and Li (2004) and Zhou (2005b). An alternative to the censored AFT model is censored quantile regression (Portnoy, 2003).

We shall study an estimation method here for the AFT correlation model, which we will term casewise weighted method. Such an approach was first proposed by Koul et al. (1982 – NB: This is *not* the manuscript published by the same three authors in the Annals of Statistics, 1981) for the least squares estimator. Zhou (1992) generalized it to regression M-estimators and derived the asymptotic distribution of the estimator based on the estimating equations

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

using modern counting process theory. Here, w_i is the jump of the Kaplan-Meier estimator at Z_i computed from (Z_i, δ_i), and ψ is a monotone function, usually the derivative of ρ. The estimator $\hat{\beta}$ is easy to obtain and does not require any iterative algorithms.

Gross and Lai (1996) used case weighted estimation of M-estimators for left-truncated and right-censored data. Stute (1993, 1996, 1999) studied the distributional behavior of the weights and applied the method to nonlinear regression models. In particular, Stute (1993) introduced a multivariate extension of the Kaplan-Meier estimator which allows for estimation under censoring in the presence of random, observable covariates. Stute did not assume a particular semiparametric model such as the Cox proportional hazard model or the accelerated failure time regression model, but a rather general nonlinear regression model. He postulated the following two assumptions regarding independence. (i) The censoring times C_i are independent of Y_i. (ii) Given Y_i, the covariates do not provide any further information as to whether censoring takes place or not, that is $P(Y_i \leq C_i|X_i, Y_i) = P(Y_i \leq C_i|Y_i)$ (Stute also provided the perspicuous explanation that "this is a convenient way to remind you that once Y is known, things which had been considered to be of some importance in your life, then become irrelevant," Stute, 1993, p.91). Furthermore, it was noted in Stute (1996) that in order to achieve almost sure and distributional convergence of the Kaplan-Meier integrals, it is sufficient to assume that the $(Z_i, \delta_i); i = 1, \ldots, n$, are i.i.d., and the independence of Y_i and C_i is only needed to identify the limit. In Stute (1999), the methodology was extended to estimate the joint distribution of (X_i, Y_i) in a nonlinear regression model where Y_i is subject to censoring and X_i are random, observable covariates. Other studies include, for example, Huang et al. (2005) who suggested a weighted least absolute deviations method for estimation in the AFT model with right-censored data. Like Stute (1993, 1996, 1999), Huang et al. (2005) also used the Kaplan-Meier weights to account for censoring. Finally, note that the expression "inverse censoring probability weight" used in the book by van der Laan and Robins (2003) also refers to a casewise weighting method.

Our focus here is to use the casewise weighting estimation approach and to derive the related empirical likelihood inference methods for the AFT correlation model. The results obtained in this manuscript also apply to censored quantile regression models, which can be considered an appealing alternative to conditional mean regression with the AFT model. In the censored quantile regression, the model is

$$Y_i = X_i\beta(\tau) + \epsilon_i \tag{2}$$

where the ϵ_i are independent with unspecified distribution F_{e_i} whose τth quantile is 0. With no censoring present in the data, the F_{e_i} are usually allowed to be not necessarily identical as to reflect more the essence of quantile regression. However, as presented in Portnoy (2003), the requirement of estimating the censoring probability for each observation, even for estimating a single quantile, has restricted many quantile regression approaches to be under more stringent assumptions than usually required for other approaches. Restrictions are put either on the censoring probabilities such as *constant known* censoring times for every observation (even uncensored ones), while preserving the heteroscedasticity (e.g., Powell, 1984, 1986; Buchinsky and Hahn, 1998; Chernozhukov and Hong, 2001). Or, alternatively, the heteroscedasticity is sacrificed to permit random censoring, conditional or unconditional (e.g., Ying et. al., 1995; Yang, 1999; Honoré et. al., 2002). We refer to Portnoy (2003) and reference therein for more details. An exception is Portnoy (2003) where the assumption of linearity of *all* conditional quantiles suffices to accommodate both conditional random censoring from the above (right censoring) and heteroscedasticity. The casewise weighting scheme proposed in the present manuscript, and the accordingly defined estimation and inference methods, provide a computationally simple alternative to Portnoy's (2003) estimator and the bootstrap based inference in censored quantile regression.

1.3 Empirical Likelihood for the AFT Model

Early attempts to use empirical likelihood with the AFT model include Qin and Jing (2001) and Wang and Li (2002) who studied the 'synthetic data' estimation method (Koul et al., 1981; Zheng, 1984; Leurgans, 1987). However, they used a different definition of the EL function compared to the one we use below, and the limiting distribution of the so defined EL ratio is not pivotal. It is characterized by linear combinations of chi squares, and involves unknown quantities that need to be estimated. Also, in several simulations, the resulting statistical inference did not perform so well.

The reason is that the 'synthetic data' likelihood is not the actual empirical likelihood for the censored data: the terms in the estimation equation that are given by 'synthetic data' are not independent of each other.

Zhou and Li (2004) and Zhou (2005b) have proposed and studied the use of EL with the censored AFT regression model respectively for the Buckley-James estimator and the rank based estimator. In both cases, they have proposed a censored EL function that is residual based, since the estimating equations are residual based and correctly reflect the censoring. Moreover they showed that the limit of the -2 log empirical likelihood ratio has a true chi square distribution (Wilks' theorem), assuming the regression model described on p.3 in this manuscript.

We propose and study here the EL analysis for the AFT *correlation* model using a casewise weighted estimation method. As mentioned in the introduction of this manuscript, the two different approaches of EL to the AFT models correspond exactly to the two different approaches to bootstrap for regression models: resampling the residuals for regression models and resampling the vectors (cases) (Y, X) for correlation models.

2 CASEWISE WEIGHTED ESTIMATION AND EMPIRICAL LIKELIHOOD

2.1 Censored Regression Model and M-Estimator

We shall now discuss in detail an estimation method in the censored data AFT (correlation) model that we will call casewise weighted approach.

The estimator $\hat{\beta}$ for the AFT correlation model is defined through the following minimization criterion:

$$\hat{\beta} = argmin_b \sum_i \rho(Z_i - X_i^t b) w_i \ , \tag{3}$$

where $\rho(\cdot)$ is a non-negative, convex loss function. For example, for $\rho(t) = t^2$ we obtain the least squares estimator, whereas $\rho(t) =\mid t \mid$ corresponds to median regression. The weights w_i are taken to be the jumps of the Kaplan-Meier estimator, $w_i = \Delta \hat{F}(Z_i)$, which are computed from (Z_i, δ_i).

With some careful definition of a function $\psi(\cdot)$, we may turn the minimization definition above into an estimating equation: the estimator $\hat{\beta}$ is the solution of the equation

$$0 = \sum_i X_i^t \psi(Z_i - X_i^t b) w_i \ . \tag{4}$$

In contrast to the 'synthetic data' approach of Koul et al. (1981), Leurgans (1987), and various generalizations, the 'casewise weighted' approach never creates any new response values (i.e. synthetic data). Instead, it tries to recoup the effect of censored responses by properly weighting the uncensored responses.

Another point worth noting is that there seem to exist two different weighting schemes discussed in the literature, but they are in fact equivalent: Weighting by inverse probabilities of censoring, or weighting by the jumps of the Kaplan-Meier estimator. For example, Stute (1993, 1996) uses as weights the jumps of the Kaplan-Meier estimator. He further rewrites the Kaplan-Meier estimator $\hat{F}$ as

$$\hat{F} = \sum_{i=1}^n w_i D_{Z_{(i)}} \quad \text{where}$$

$$w_1 = \frac{\delta_{(1)}}{n}, \quad \text{and} \quad w_i = \frac{\delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}}, i = 2, \ldots, n.$$

Here, $Z_{(i)}$ denotes the ith order statistic, $\delta_{(i)}$ is the censoring indicator δ corresponding to $Z_{(i)}$, and D_x denotes the Dirac measure on x.

On the other hand, inverse censoring probability weights have been used in many different places, see van der Laan and Robins (2003) or Rotnitzky and Robins (2005) for a survey. In fact, these two weighting schemes are identical. This is most easily seen since for all t

$$[1 - \hat{F}(t)][1 - \hat{G}(t)] = 1 - \hat{H}(t)$$

where $\hat{F}(t)$ is the Kaplan-Meier estimator based on (Z_i, δ_i), and $\hat{G}(t)$ is the Kaplan-Meier estimator computed by replacing δ_i by $d_i = 1 - \delta_i$. Sometimes, the latter is called the Kaplan-Meier estimator of the censoring distribution. The estimator $\hat{H}(t)$ is just the regular empirical distribution based on the Z_i's. For $t = Z_i$ with $\delta_i = 1$, both $\hat{H}(t)$ and $\hat{F}(t)$ have jumps, but $\hat{G}(t)$ does not. We therefore have, by examining the jumps of the above equation,

$$[\Delta\hat{F}(t)][1 - \hat{G}(t)] = 1/n \quad \text{for} \quad t = Z_i \;\; \text{with} \;\; \delta_i = 1 \; .$$

Finally for $t = Z_i$ with $\delta_i = 1$, we obtain

$$w_i = \delta_i \Delta\hat{F}(Z_i) = \frac{\delta_i}{n[1 - \hat{G}(Z_i)]} \; . \tag{5}$$

We see that, aside from a constant, the jumps of Kaplan-Meier weights $w_i = \Delta\hat{F}(Z_i)\delta_i$ are the same as the inverse censoring probability weights

$$w_i = \frac{\delta_i}{n[1 - \hat{G}(Z_i)]}.$$

Remark: As is conventional, we always treat the largest observation as an uncensored observation, since there are no more observations left for it to re-distribute the weights.

Notice that in the above, we do not need to assume the existence of a 'true' censoring distribution $G(t)$. $\hat{G}(t)$ is merely a notation. Since all the calculations are algebraic identities, they hold for any samples of (Z_i, δ_i).

The distinguished feature of this estimation method is that the computation of the estimator is simple, no iteration is needed to calculate the weights, as opposed to the Buckley-James estimator. Consequently, there is also no convergence problem. The (weighted) minimization equation (3) and the estimating equation (4) can be solved using available software like SAS and Splus/R. One just needs to provide the weights w_i. The weights w_i in the equations above are the jump sizes of the Kaplan-Meier estimator computed from (Z_i, δ_i). Notice that the Kaplan-Meier estimator here does not depend on the residuals, on β, or on $\hat{\beta}$ (as opposed to the Buckley-James estimator). Also, the creation of the weights does not involve the covariates which may be high dimensional, or missing.

More importantly, as we will show next, this estimation procedure permits an empirical likelihood inference. We will show that the -2 log likelihood ratio has asymptotically a regular χ^2 limit distribution under null hypothesis.

We propose to use the empirical likelihood method to obtain confidence intervals for the estimators and p-values for testing hypotheses. This has some advantages over re-sampling based methods.

Zhou (1992) proposed the above approach for a general ρ and proved that the resulting estimator has an asymptotic normal distribution for $\rho(t) = t^2$. Stute (1993) showed the consistency of this estimator, Huang et al. (2005) studied the asymptotic distribution for absolute error loss. Under fairly general conditions, the estimator has an asymptotic normal distribution. That is, the estimator $\hat{\beta}$ defined through (3) or (4) is consistent and asymptotically normally distributed.

Next, we study the empirical likelihood connection for this weighted estimation method.

2.2 Empirical Likelihood Function for the AFT Correlation Model

Under the AFT correlation model, we define the censored data EL with respect to (Z_i, δ_i):

$$EL(Z,\delta) = \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{Z_j \le Z_i} p_j) \ , \tag{6}$$

where $p_i \ge 0, \sum p_i = 1$. This is in fact the marginal EL for censored Y. We call this the *casewise EL* function.

Remark: This should be contrasted with the empirical likelihood used by Zhou and Li (2004) and Zhou (2005b). They define the censored EL as

$$EL(e,\delta) = \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{e_j \le e_i} p_j) \ , \tag{7}$$

where $e_i = Z_i - X_i b$. We will call this *residual EL*.

We want to point out that the casewise empirical likelihood $EL(Z,\delta)$ does not involve the estimator $\hat{\beta}$, whereas in the residual empirical likelihood, $EL(e(b),\delta)$ does involve $\hat{\beta}$. Also, $EL(Z,\delta)$ does not involve the covariate X_i. For very high dimensional covariates, this can be an advantage. Furthermore, in the constraint equation (7) below, the values of X_i for $\delta_i = 0$ are not needed. This makes inference possible even for samples that have missing covariates for censored responses.

Clearly, the maximum of the $EL(Z,\delta)$ is attained when p_i is the (jump of the) Kaplan-Meier estimator computed from (Z_i, δ_i). That is, $p_i = w_i$ in equation (5) above. Let us denote this maximized value of EL as EL(Kaplan-Meier).

More interestingly, we would like to maximize the $EL(Z,\delta)$ (with respect to p_i) when p_i satisfies the constraint equations

$$0 = \sum_{i=1}^{n} p_i \delta_i X_i^t \psi(Z_i - X_i^t \beta_0) \ . \tag{8}$$

A similar consideration as Owen (1988) will make us impose the following restriction on the p_i with regard to the w_i: The probability distribution given by the p_i is absolutely continuous with respect to the distribution given by the w_i. Since the w_i distribution is discrete, this means that the set of points that are assigned positive probabilities through the p_i must be a subset of the set of points with positive w_i. Let us denote this constrained maximized value of $EL(Z, \delta)$ as $EL(H_0)$.

This constrained maximization problem does not have an explicit algebraic solution, but it can be solved reliably by the modified EM algorithm of Zhou (2005a). See also the function `el.cen.EM2`, in the R package `emplik`.

Thus, the empirical likelihood ratio can be obtained readily: The numerator can be obtained using the function `el.cen.EM2`, and the denominator can be obtained with p_i = the (jump of) Kaplan-Meier estimator.

With the empirical likelihood ratio, we may test the hypothesis

$$H_0 : \beta = \beta_0 \;\; vs. \;\; H_a : \beta \neq \beta_0$$

using the following theorem.

Theorem. Under the null hypothesis above and assuming some regularity conditions, the -2 log empirical likelihood ratio has asymptotically a regular chi squared distribution (Wilks' theorem):

$$-2 \log \frac{EL(H_0)}{EL(\text{Kaplan-Meier}(Z_i, \delta_i))} \longrightarrow \chi^2_k \; ,$$

where k (degrees of freedom) equals the length of the parameter vector β. The required regularity conditions depend on the respective choice of the function ψ. For example, for median regression, see Huang et al. (2005), Murphy and van der Vaart (1997); and for least squares estimation, see Zhou (1992) for appropriate sets of conditions.

Proof: Notice here that the constraint equation involves i.i.d. observations (see the correlation model described on p.2) of which only the first dimension is censored. Known results for the multivariate Kaplan-Meier estimator (Stute, 1993) can be used. Recall that the p_i are the weights of the marginal distribution of Y (see display (6)).

Therefore, equation (8) can be interpreted as a constraint on the marginal distribution of Y, and rewritten as

$$\int g_{X_1,X_2,\ldots,X_n}(y)dF(y) = 0.$$

Thus, the result from Murphy and van der Vaart (1997) can be applied, and we obtain the asymptotic χ^2-distribution of the empirical likelihood ratio.

Remark: If instead of testing for β_0, we test for $\beta = \hat{\beta}$, where $\hat{\beta}$ is the estimator defined above, then we must have the -2 log likelihood ratio equal to zero (since the ratio is one), and thus the p-value of the test equal to 1. Therefore, the confidence interval/region is "centered" at the case weighted estimator that we defined.

$(1-\alpha)$–confidence regions can be obtained as sets of parameter values that would result in a p-value larger than α. Notice that the confidence interval based on $\hat{\beta}$ is obtained without estimating the asymptotic variance of the estimator. The estimator of the regression parameter is easy to compute (no iteration needed), but a common problem of other approaches is the estimation of the asymptotic variance.

3 SIMULATIONS AND EXAMPLE

3.1 Simulating the Empirical Likelihood Ratios

In the following, we present simulation results obtained using R (Ihaka and Gentleman, 1996; including the packages 'emplik'). We focus on three different situations that are special cases of the general approach discussed above: Mean regression and median regression with homoscedastic error, as well as quantile regression with heteroscedastic error. For each situation, quantile-quantile (Q-Q) plots are created to compare the simulated distribution of the empirical likelihood ratio with the asymptotic χ^2-distribution.

The random variables Y_i are generated according to the simple linear model $Y_i =$

$\beta_0+\beta_1 X_i+\sigma(X_i)\epsilon_i$ with different parameter values for β_0 and β_1, and different functions $\sigma(X_i)$ for simulating heteroscedastic models. The X_i are i.i.d. uniform(0,1) random variables, and the ϵ_i are i.i.d. normal. Then, the Y_i are subjected to right censoring from a random variable with shifted exponential distribution: 0.5+2exp(1). Information about sample sizes, parameter values, and censoring percentages, as well as details about error and censoring distribution are provided for each of the simulations. The number of simulation runs is always 1,000. The $-2\log$ empirical likelihood ratios are computed in each simulation run for testing the appropriate, correct null hypothesis (e.g., $H_0 : \beta_0 = 1,\ \beta_1 = 1$ for Figure 1). The resulting Q-Q plots for the 1,000 sample empirical likelihood ratios show in general a good fit to the χ^2-distribution with 2 degrees of freedom. That is, the empirical likelihood ratio is well approximated by the χ^2-distribution for moderate sample sizes. In the case of quantile regression, upper quantiles are, as expected, more affected by right-censoring than the median or lower quantiles (compare Figures 3 and 4 for 25% and 75% quantile regression). Note that the proposed method which is based on the correlation model is not affected by heteroscedasticity in the errors.

For brevity, only selected Q-Q plots are presented. More extensive simulations have been carried out, and additional results are available upon request from the authors.

Figure 1: Q-Q plot of $-2 \log ELR$ for mean regression with homoscedastic errors. Sample sizes 50, 100, 200, and 400.

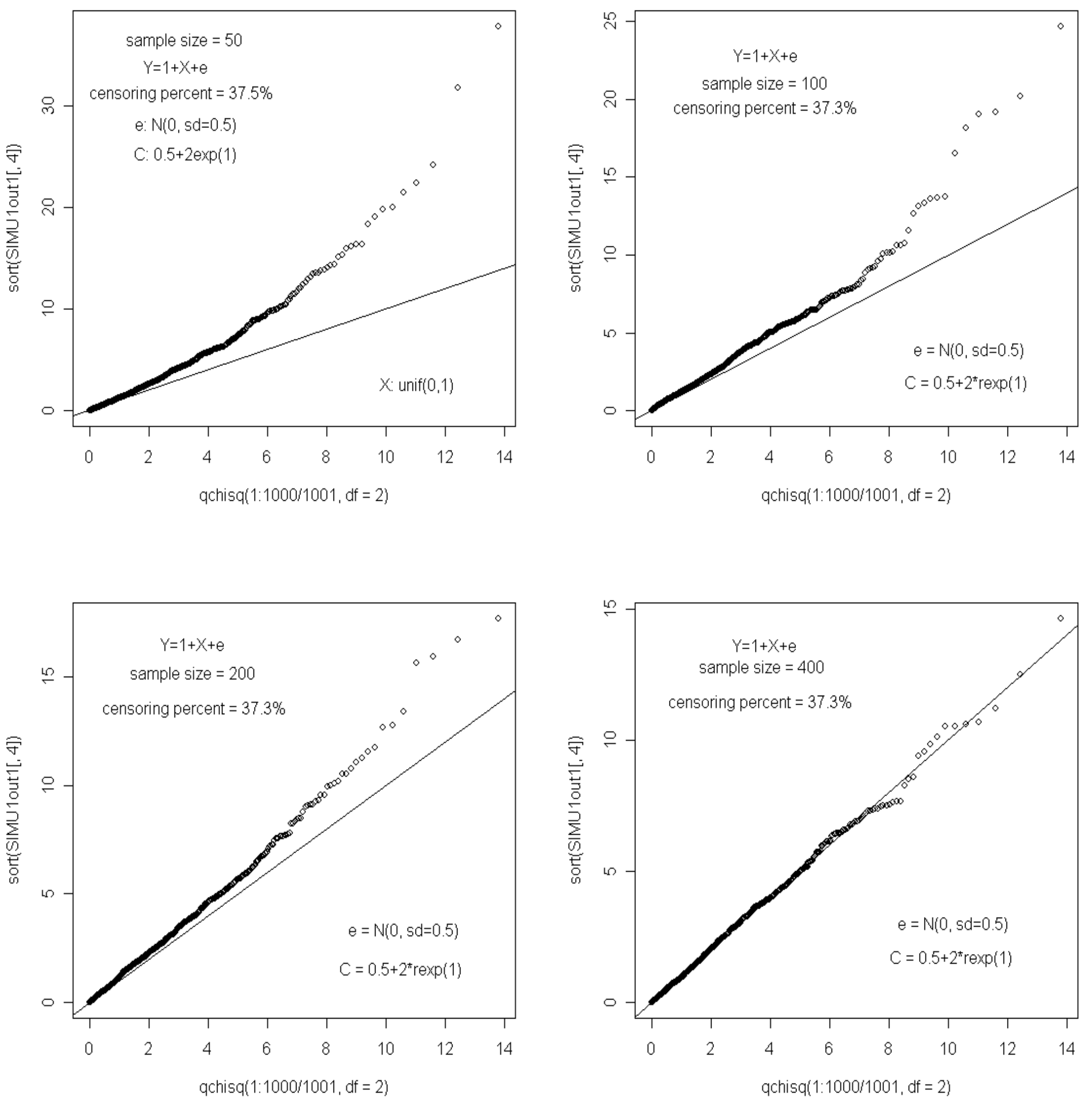


Figure 2: Q-Q plot of $-2 \log ELR$ for median regression with homoscedastic error, where $\psi(t) = 1$ or -1 depending on $t > 0$ or $t < 0$. Sample sizes 50, 100, 200, and 400.

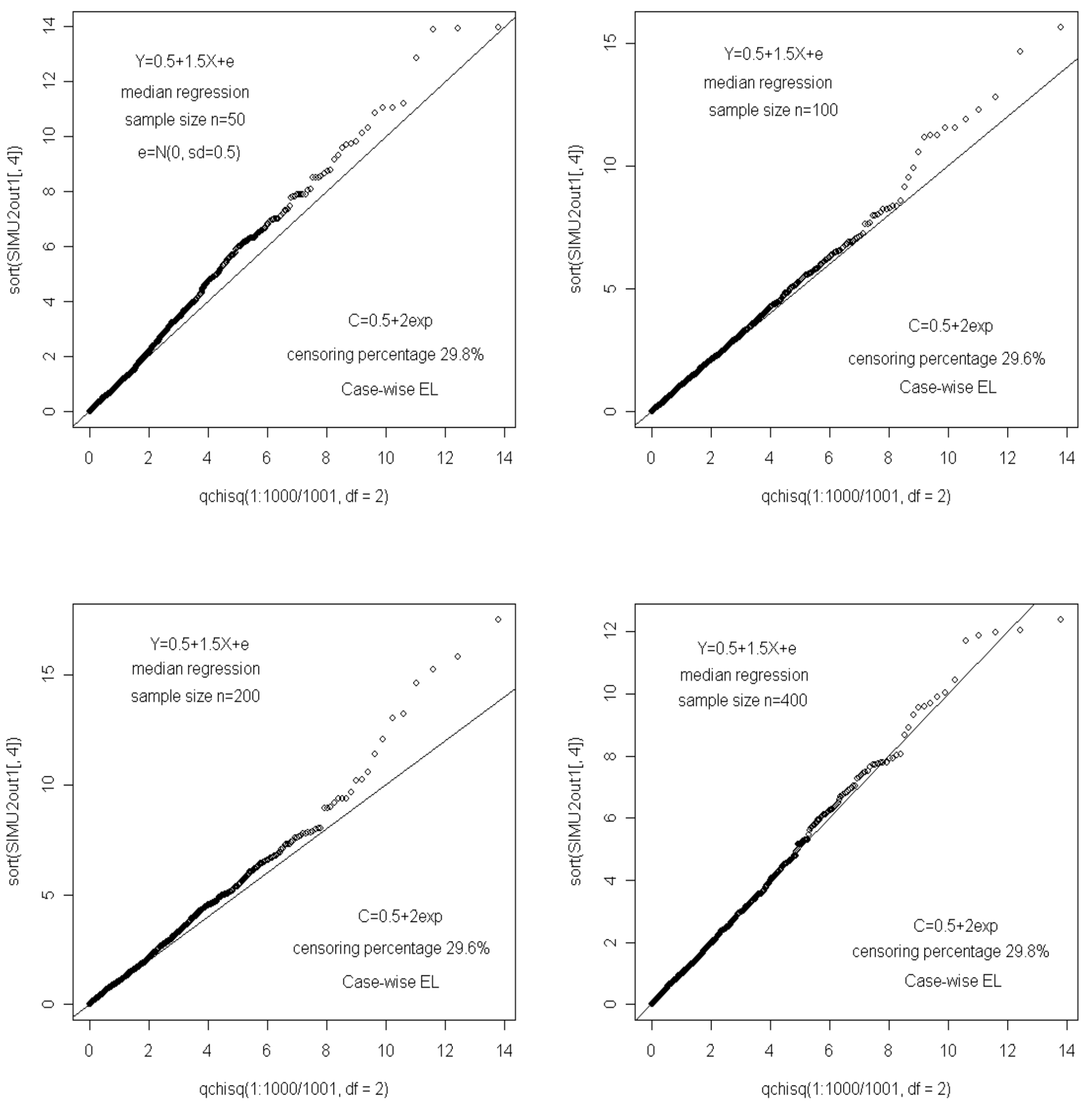


Figure 3: Q-Q plot of $-2 \log ELR$ for 25% quantile regression with heteroscedastic error. Sample sizes 50, 100, 200, and 400.

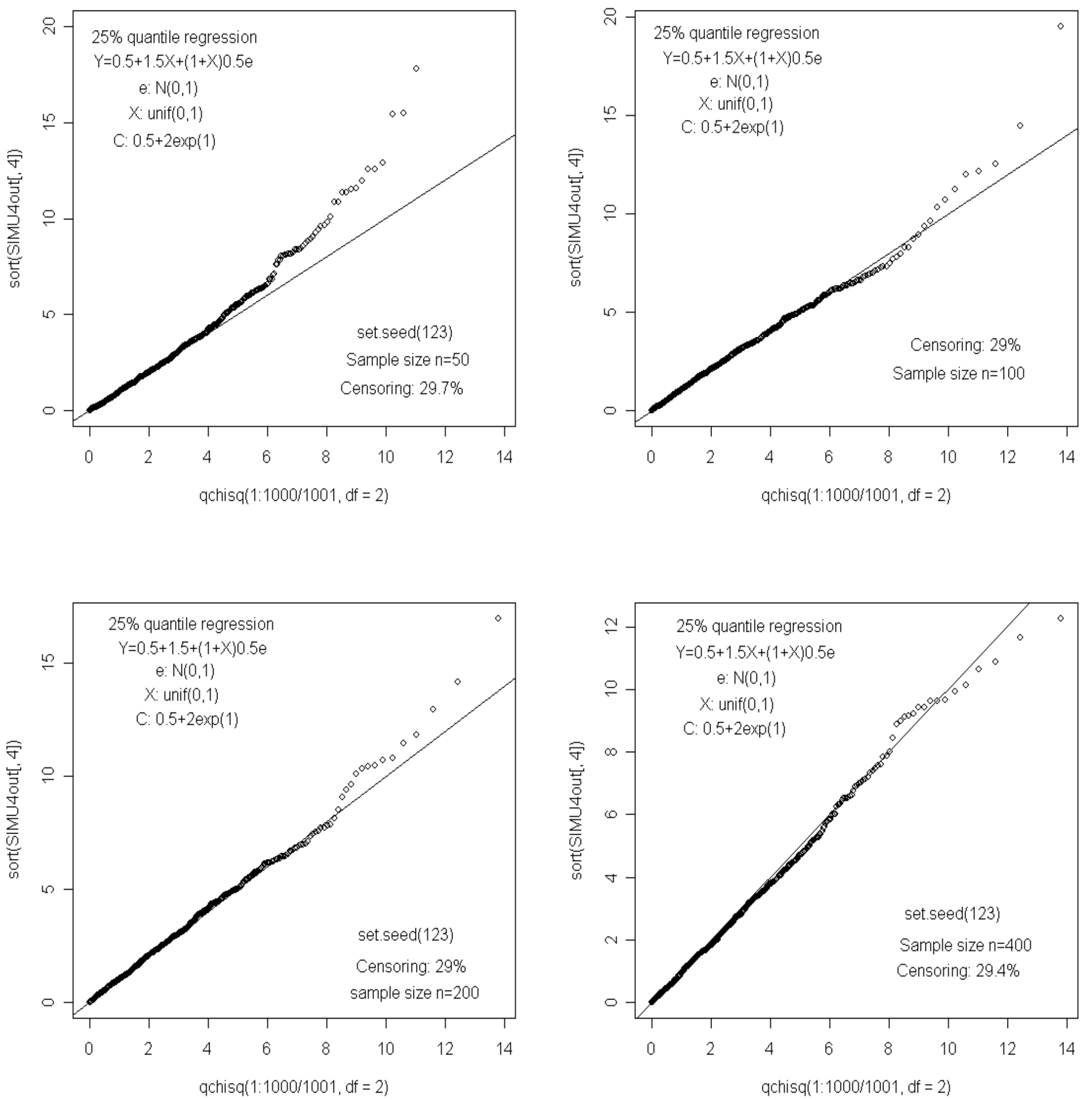
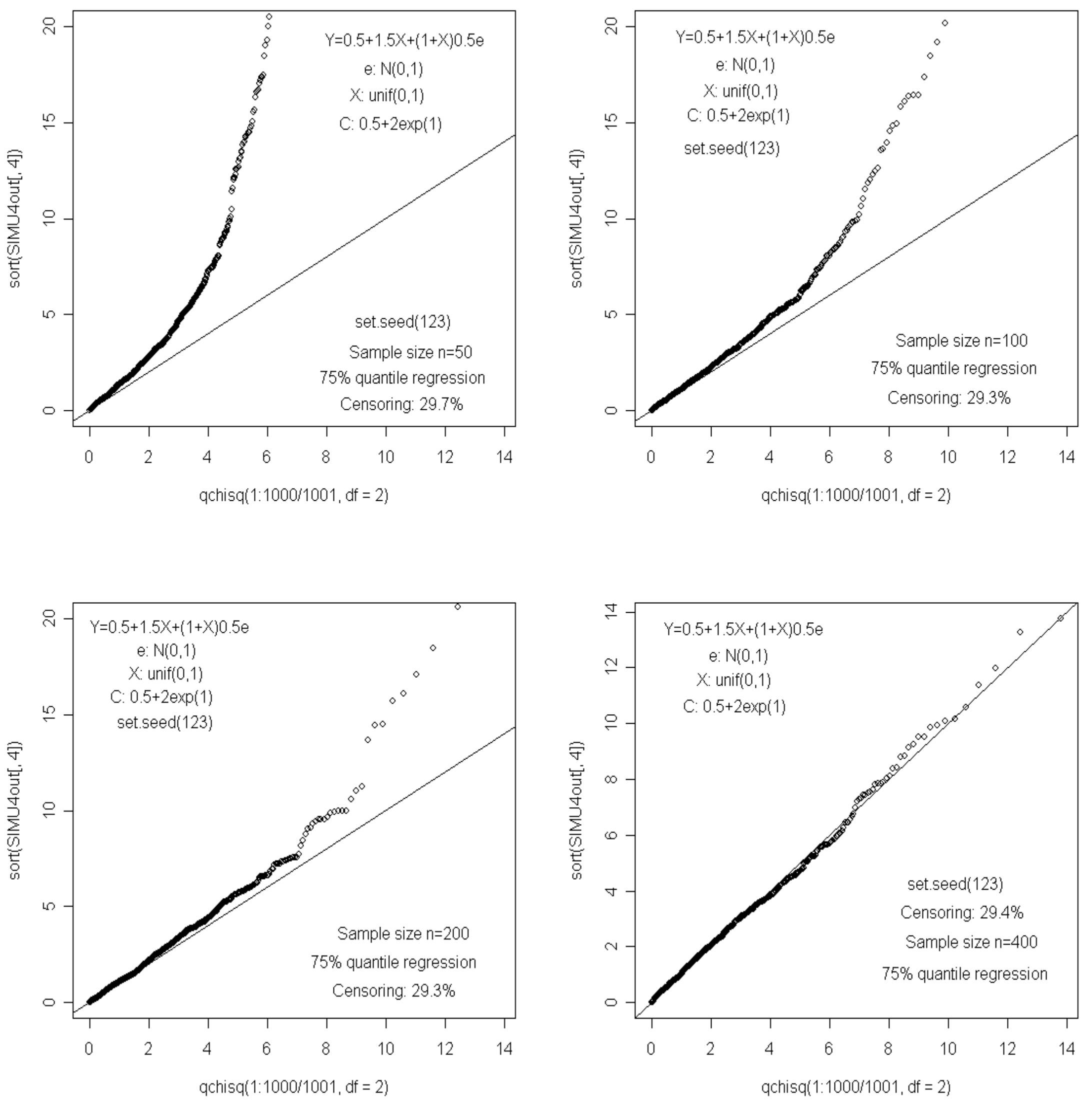


Figure 4: Q-Q plot of $-2 \log ELR$ for 75% quantile regression with heteroscedastic error. Sample sizes 50, 100, 200, and 400.



3.2 Small-Cell Lung Cancer Data

We consider a lung cancer data set (Maksymiuk et al., 1994) that has been analyzed by Ying et al. (1995) using median regression, and by Huang et al. (2005) using a least absolute deviations method in the accelerated failure time (AFT) model. In this study, 121 patients with limited-stage small-cell lung cancer were randomly assigned to one of two different treatment sequences A and B, with 62 patients assigned to A and 59 patients to B. Each death time was either observed or administratively censored, and the censoring variable did not depend on the covariates *treatment* and *age*. Denote X_{1i} the treatment indicator variable, and X_{2i} the entry age for the ith patient, where $X_{1i} = 1$ if the patient is in group B. Let Y_i be the base 10 logarithm of the ith patient's failure time. We assume the AFT model (1)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \sigma(X_{1i}, X_{2i})\epsilon_i \ .$$

Note that the estimated parameter values obtained using the approach described in this paper need to be equal to the ones from Huang et al. (2005), provided weighting has been done in the same way. The major difference is in inference about the parameters, where empirical likelihood has the advantage that it is not necessary to estimate the asymptotic variance of the estimator in order to perform hypothesis tests and to construct confidence regions. An empirical likelihood confidence region for this data set is displayed in Figure 5.

The following median regression estimates were obtained by Ying et al. (1995) and Huang et al. (2005).

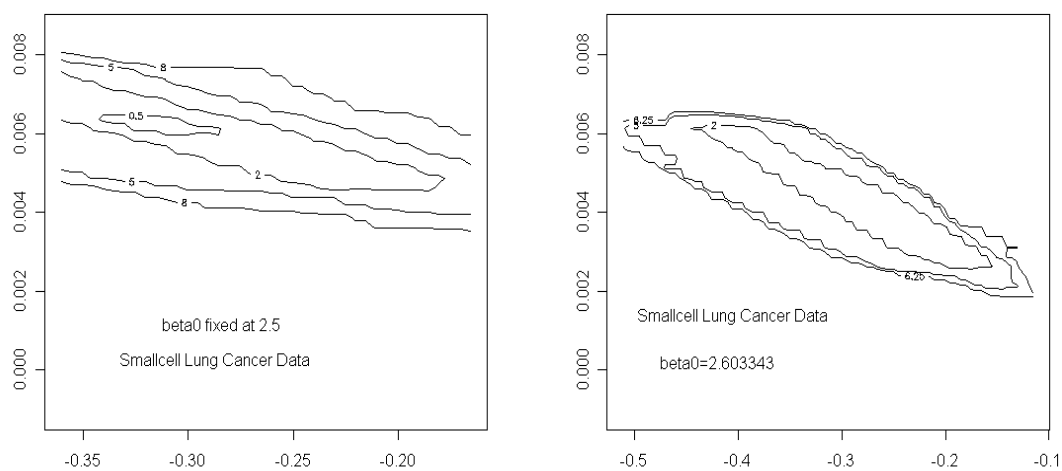
$$\hat{\beta}_0 = 3.028, \ \hat{\beta}_1 = -0.163, \ \text{ and } \ \hat{\beta}_2 = -0.004 \qquad \text{(Ying et al., 1995)}$$

$$\hat{\beta}_0 = 2.693, \ \hat{\beta}_1 = -0.146, \ \text{ and } \ \hat{\beta}_2 = 0.001 \qquad \text{(Huang et al., 2005)}$$

Huang et al. (2005) did not always treat the largest Y observation as uncensored. This resulted in weights that sum to less than one. We recommend to always treat the largest Y observation as uncensored so that the weights always sum to one. The median regression estimates then become

$$\hat{\beta}_0 = 2.603, \ \hat{\beta}_1 = -0.263, \ \text{ and } \ \hat{\beta}_2 = 0.004 \qquad \text{(with last weight).}$$

Figure 5: Confidence Region for (β_1, β_2) in the small-cell lung cancer data. The original parameter space is 3-dimensional, therefore only two two-dimensional cuts at $\beta_0 = 2.5$ and $\beta_0 = 2.603$ are displayed. The contour lines correspond to different critical values.



REFERENCES

Buchinsky, M., and Hahn (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, **66, 3**, 653–672.

Buckley, J.J., and James, I.R. (1979), "Linear Regression With Censored Data," *Biometrika*, **66**, 429–36.

Chernozhukov, V., and Hong, H. (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, **115**, 293–346.

Cox, D.R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society B*, **34, 2**, 187–202.

Freedman, D.A. (1981), "Bootstrapping Regression Models," *Annals of Statistics*, **9**, 1218–1228.

Gross, S., and Lai, T.L. (1996), "Nonparametric Estimation and Regression Analysis With Left Truncated and Right Censored Data," *Journal of the American Statistical Association*, **91**, 1166–1180.

Honoré, B. , Khan, S., and Powell, J. (2002). "Quantile Regression Under Random Censoring," *Journal of Econometrics*, **109**, 67–105.

Huang, J., Ma, S. and Xie, H. (2005), "Least Absolute Deviations Estimation for the Accelerated Failure Time Model," Technical Report 350, University of Iowa, Dept. of Statistics and Actuarial Science.

Ihaka, R., and Gentleman, R. (1996), "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, **5, 3**, 299–314.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), "Rank-Based Inference for the Accelerated Failure Time Model," *Biometrika*, **90, 2**, 341–353.

Kalbfleisch, J. D., and Prentice, R.L. (2002), *The Statistical Analysis of Failure Time Data, Second Edition*, New York: Wiley.

Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, **53, 282**, 457–481.

Koul, H., Susarla, V., and Van Ryzin, J. (1981), "Regression Analysis With Randomly Right-Censored Data," *Annals of Statistics*, **9**, 1276–1288.

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, New York: Marcel Dekker, pp. 151–165.

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

Lai, T. L., and Ying, Z. L. (1991), "Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis With Censored Data," *Annals of Statistics*, **19, 3**, 1370–1402.

Leurgans, S. (1987), "Linear Models, Random Censoring and Synthetic Data," *Biometrika*,

74, 2, 301–309.

Li, G., and Wang, Qi-Hua (2003), "Empirical Likelihood Regression Analysis for Right Censored Data," *Statistica Sinica*, **13**, 51–68.

Maksymiuk, A. W., Jett, J. R., Earle, J. D., Su, J. Q., Diegert, F. A., Mailliard, J. A., Kardinal, C. G., Krook, J. E., Veeder, M. H., Wiesenfeld, M., Tschetter, L. K., and Levitt, R. (1994), "Sequencing and Schedule Effects of Cisplatin Plus Etoposide in Small Cell Lung Cancer, Results of a North Central Cancer Treatment Group Randomized Clinical Trial", *Journal of Clinical Oncology*, **12**, 70–76.

Murphy, S., and van der Vaart, A. (1997), "Semi-Parametric Likelihood Ratio Inference," *Annals of Statistics* **25**, 1471–1509.

Owen, A. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, **75**, 237–249.

Owen, A. (1990), "Empirical Likelihood Ratio Confidence Regions," *Annals of Statistics*, **18**, 90–120.

Owen, A. (1991), "Empirical Likelihood for Linear Models," *Annals of Statistics*, **19, 4**, 1725-1747.

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

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, **98**, 1001–1012.

Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, **25**, 303–325.

Powell, J. L. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, **32**, 143–155.

Qin, G. S., and Jing, B. Y. (2001), "Empirical Likelihood for Censored Linear Regression," *Scandinavian Journal of Statistics*, **28**, 661–673.

Reid, N. (1994), "A Conversation With Sir David Cox," *Statistical Science*, **9**, 439–455.

Rotnitzky A, Robins J. M. (2005), "Inverse Probability Weighted Estimation in Survival Analysis," in: *The Encyclopedia of Biostatistics, 2nd edition*, Ed. P. Armitage and T. Coulton, New York: Wiley.

Stute, W. (1993), "Consistent Estimation Under Random Censorship When Covariables are

Present," *Journal of Multivariate Analysis*, **45**, 89–103.

Stute, W. (1996), "Distributional Convergence Under Random Censorship When Covariables are Present," *Scandinavian Journal of Statistics* **23**, 461–471.

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

Thomas, D. R., and Grunkemeier, G. L. (1975), "Confidence Interval Estimation of Survival Probabilities for Censored Data," *Journal of the American Statistical Association*, **70**, 865–871.

Wang, Q.H., and Li, G. (2002), "Empirical Likelihood Semiparametric Regression Analysis Under Random Censorship," *Journal of Multivariate Analysis*, **83**, 469–486.

Wei, L. J. (1992), "The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis," *Statistics in Medicine*, **11**, 1871–1879.

Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Hazard and Survival Functions," *Journal of the American Statistical Association*, **94**, 137–145.

Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival Analysis with Median Regression Models," *Journal of the American Statistical Association* **90, 429**, 178–184.

Zheng, Z. (1984), "Regression with Randomly Censored Data," unpublished Ph.D. dissertation, Columbia University, Dept. of Statistics.

Zhou, M. (1992), "M-Estimation in Censored Linear Models," *Biometrika*, **79, 4**, 837–841.

Zhou, M. (2005a), "Empirical Likelihood Ratio With Arbitrarily Censored/Truncated Data by a Modified EM algorithm," *Journal of Computational and Graphical Statistics*, **14**, 643–656.

Zhou, M. (2005b), "Empirical Likelihood Analysis of the Rank Estimator for the Censored Accelerated Failure Time Model," *Biometrika*, **92**, 492–498.

Zhou, M., and Li. G. (2004), "Empirical Likelihood Analysis of the Buckley-James Estimator," Technical Report 393, University of Kentucky, Dept. of Statistics.