KAROLINSKA INSTITUTET
Division of Biostatistics

Working paper

April 2021

LAPLACE REGRESSION: A ROBUST AND COMPUTATIONALLY EFFICIENT ESTIMATOR FOR CENSORED QUANTILES

Paolo Frumento and Matteo Bottai

LAPLACE REGRESSION: A ROBUST AND COMPUTATIONALLY EFFICIENT ESTIMATOR FOR CENSORED QUANTILES

Paolo Frumento and Matteo Bottai

Karolinska Institutet, Institute of Environmental Medicine, Unit of Biostatistics

Abstract:

Quantile regression is an increasingly popular approach to statistical modeling. Numerous methods have been proposed to extend quantile regression to censored data. Maximizing the log-likelihood of a Laplace distribution yields an efficient yet biased estimator and presents computational advantages over the other methods. We discuss the bias of Laplace regression estimators and present results from a comprehensive simulation study, showing that the bias is generally negligible and the mean squared error is smaller than that of the other methods. Possible explanations of this empirical evidence are discussed. Analysis of big data represents an important application of the Laplace regression estimator.

Key words and phrases: Censored quantile regression; asymmetric Laplace distribution; big data.

1 Introduction

Quantile regression (Koenker, 2005) permits estimating conditional quantiles of an outcome. Typically, a linear predictor $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p$ is used to describe the effect of a q-dimensional vector $\boldsymbol{x}$ of covariates on the p-th quantile of the response variable of interest. The vector $\boldsymbol{\beta}_p$ of regression coefficients is quantile-specific and no global assumptions are maintained, making of quantile regression a distribution-free approach.

Applying quantile regression to censored data is of great interest in various applied fields. For example, quantiles of survival time are often preferred to hazard ratios, due to their simpler interpretation. Throughout the paper, we denote by T a response variable of interest, and by C a censoring variable such that one can only observe $Y = \min(T, C)$ and $\delta = I(T \leq C)$. We denote by $f_T(t \mid \boldsymbol{x})$, $F_T(t \mid \boldsymbol{x})$, and $Q_T(p \mid \boldsymbol{x})$ the conditional probability density function (PDF), cumulative distribution function (CDF), and quantile function (QF) of T. The distribution of C given $\boldsymbol{x}$ is defined analogously. We assume that (i) T and C are independent, given $\boldsymbol{x}$; and (ii) for a given $p \in (0, 1)$,

$$Q_T(p \mid \boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p \tag{1}$$

is the conditional p-th quantile of T.

Various methods have been proposed to estimate $\boldsymbol{\beta}_p$ based on a censored random sample $(y_i, \delta_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$. Portnoy (2003) discussed a recursive reweighting algorithm that generalizes the Kaplan-Meier estimator using the redistribution-of-mass concept introduced by Efron (1967); Peng and Huang (2008) introduced a martingale-based approach that generalizes the Nelson-Aalen estimator. Both methods estimate the true outcome distribution F_T under a global linearity assumption that requires all lower-order quantiles to be linear. To overcome global linearity, that is often considered too restrictive, Wang and

Wang (2009) proposed a two-step estimator. The first step requires estimating F_T nonparametrically, while the second step is a weighted quantile regression based on a weighting function $w(F_T)$. Other two-step estimators are discussed by Leng and Tong (2013) and Frumento and Bottai (2016).

In their paper, Bottai and Zhang (2010) explored the use of the log-likelihood of the asymmetric Laplace distribution. This method is computationally simpler than the other aforementioned approaches, and does not require estimating F_T. Moreover, it can be easily generalized to handle truncation, include frailty terms, and fit nonlinear quantiles. Although Laplace regression is generally biased, empirical evidence demonstrates that the bias is usually very small. This result is counterintuitive, and the need for further investigation motivated this paper. The most relevant application of Laplace regression is the analysis of big data, where other methods may prove extremely time-consuming.

The paper is structured as follows. In Section 2, we briefly recap the Laplace regression framework, and describe its properties in Sections 3. In Section 4 we present a wide simulation study that measures its bias and compares it with that of Portnoy's (2003) estimator, which represents the most widely used method for censored quantile regression. Other estimators were shown to be similar to Portnoy's (Koenker, 2008; Frumento and Bottai, 2016), and are not considered in this paper. In section 5 we discuss computation times and argue that, as of today, Laplace regression may represent the only feasible method to analyze big data.

2 Laplace regression

2.1 The asymmetric Laplace distribution

We consider the following conditional PDF and CDF:

$$f_p(t \mid \boldsymbol{x}, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = \frac{p(1-p)}{\sigma_p(\boldsymbol{x})} \exp\left\{ \frac{(\omega_p - p)(t - \boldsymbol{x}^T\boldsymbol{\beta}_p)}{\sigma_p(\boldsymbol{x})} \right\} \tag{2}$$

$$F_p(t \mid \boldsymbol{x}, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = 1 - \omega_p + (p - 1 + \omega_p) \exp\left\{ \frac{(\omega_p - p)(t - \boldsymbol{x}^T\boldsymbol{\beta}_p)}{\sigma_p(\boldsymbol{x})} \right\}$$

where $\omega_p = I(t \leq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p)$. This distribution is an asymmetric Laplace (AL) distribution in which parameters are functions of covariates. Consistently with model (1), the location parameter is $\boldsymbol{x}^T\boldsymbol{\beta}_p$ and represents the conditional p-th quantile, as $F_p(\boldsymbol{x}^T\boldsymbol{\beta}_p \mid \boldsymbol{x}, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = p$. The scale parameter $\sigma_p(\boldsymbol{x}) > 0$ may depend on $\boldsymbol{x}$ through an unknown vector $\boldsymbol{\eta}_p$ such that $\sigma_p(\boldsymbol{x}) = \sigma(\boldsymbol{x} \mid \boldsymbol{\eta}_p)$. Throughout the paper, we will use the notation AL(p) for this conditional distribution. Note that p is an asymmetry parameter and represents the order of the quantile. Examples of the PDF of the standard AL distribution are illustrated in Figure 1.

2.2 Equivalence between Laplace regression and ordinary quantile regression

We denote by $(\boldsymbol{x}_i, t_i)$ a random sample from the joint distribution of $(\boldsymbol{x}, T)$, $i = 1, \ldots, n$. The log-likelihood of the AL(p) model is the following:

$$l_n(\boldsymbol{\beta}_p, \sigma_p) = n^{-1} \sum_{i=1}^{n} -\log \sigma_p(\boldsymbol{x}_i) + (\omega_{i,p} - p)(t_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_p)/\sigma_p(\boldsymbol{x}_i). \tag{3}$$

If σ_p is scalar, maximizing (3) with respect to $\boldsymbol{\beta}_p$ is equivalent to minimizing the loss function of ordinary quantile regression, given by

$$n^{-1} \sum_{i=1}^{n} (p - \omega_{i,p})(t_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_p),$$

and yields the same consistent estimator of $\boldsymbol{\beta}_p$ irrespectively of the true outcome distribution. If σ_p is allowed to depend on covariates, observations are given a non-constant weight $\sigma_p(\boldsymbol{x}_i)^{-1}$, which is proportional to the AL(p) density at $t_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_p$. This also yields a consistent estimator of $\boldsymbol{\beta}_p$, and frequently leads to a gain in efficiency. On the other hand, it requires carrying out joint estimation of $\boldsymbol{\beta}_p$ and $\sigma_p(\boldsymbol{x})$, which is unnecessary when σ_p is scalar.

2.3 Censored Laplace regression

In their paper, Bottai and Zhang (2010) introduced the idea of estimating censored quantile regression coefficients by fitting the AL distribution to the data, extending the equivalence between Laplace regression and quantile estimation. The log-likelihood of the AL(p) model for censored data is

$$l_n(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = \tag{4}$$

$$n^{-1}\sum_{i=1}^{n} \delta_i \log f_p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x}_i)) + (1-\delta_i)\log \bar{F}_p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x}_i))$$

where $y_i = \min(t_i, c_i)$, $\delta_i = I(t_i \leq c_i)$, and $\bar{F}_p(\cdot) = 1 - F_p(\cdot)$. Maximizing (4), however, yields a biased estimator of $\boldsymbol{\beta}_p$ except in trivial cases (see Section 3). Since the model is misspecified (as $F_T \neq F_p$), a misspecified CDF is used for the contribution of censored observations in the likelihood function.

Unexpectedly, the bias has been shown to be very small in various simulation scenarios. A heuristic interpretation has been provided, based on a maximum entropy principle in which a fundamental role is played by the scale parameter σ_p, that represents the scale parameter of an Exponential distribution governing the weighted residuals $(p-\omega_{i,p})(t_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}_p)$. We refer to Bottai and Zhang (2010) for details.

3 The score function and its expectation

The first derivatives of log-likelihood (4) with respect to its arguments are

$$S_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = n^{-1} \sum_{i=1}^{n} \frac{\boldsymbol{x}_i}{\sigma_p(\boldsymbol{x}_i)} \times \tag{5}$$

$$\times \Big\{ p - \omega_{i,p} + \frac{\omega_{i,p}(1-\delta_i)(1-p)}{1 - F_p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x}_i))} \Big\}$$

and

$$S_{\sigma_p}(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = n^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_p(\boldsymbol{x}_i)} \times \tag{6}$$

$$\times \Big[\frac{y_i - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_p}{\sigma_p(\boldsymbol{x}_i)} \Big\{ p - \omega_{i,p} + \frac{\omega_{i,p}(1-\delta_i)(1-p)}{1 - F_p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x}_i))} \Big\} - \delta_i \Big],$$

respectively. The expected values of $S_{\boldsymbol{\beta}_p}$ and S_{σ_p} are

$$\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = E_{\boldsymbol{x}} \Big[\frac{\boldsymbol{x}}{\sigma_p(\boldsymbol{x})} \Big\{ p - 1 + \bar{F}_T(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\beta}_p \mid \boldsymbol{x}) \bar{F}_C(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\beta}_p \mid \boldsymbol{x}) + \tag{7}$$

$$+ (1-p) \int_{-\infty}^{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\beta}_p} \frac{\bar{F}_T(t \mid \boldsymbol{x})}{\bar{F}_p(t \mid \boldsymbol{x}, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x}))} \mathrm{d}F_C(t \mid \boldsymbol{x}) \Big\} \Big]$$

and

$$\bar{S}_{\sigma_p}(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) = E_{\boldsymbol{x}} \Big[\frac{1}{\sigma_p(\boldsymbol{x})^2} \Big\{ pE[Y \mid \boldsymbol{x}] - \int_{-\infty}^{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\beta}_p} t f_Y(t \mid \boldsymbol{x}) \mathrm{d}t + \tag{8}$$

$$+ (1-p) \int_{-\infty}^{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{\beta}_p} \frac{t \bar{F}_T(t \mid \boldsymbol{x})}{\bar{F}_p(t \mid \boldsymbol{x}, \boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x}))} \mathrm{d}F_C(t \mid \boldsymbol{x}) - \sigma_p(\boldsymbol{x}) E[\delta \mid \boldsymbol{x}] \Big\} \Big],$$

respectively. In the above formulas, $E_{\boldsymbol{x}}$ denotes expectation over the distribution of $\boldsymbol{x}$. In equation (8) we omitted a quantity that is zero when $\bar{S}_{\boldsymbol{\beta}_p} = 0$, while $f_Y(y \mid \boldsymbol{x}) = f_T(y \mid \boldsymbol{x}) \bar{F}_C(y \mid \boldsymbol{x}) + f_C(y \mid \boldsymbol{x}) \bar{F}_T(y \mid \boldsymbol{x})$ is the conditional distribution of $Y = \min(T, C)$, and $E[\delta \mid \boldsymbol{x}] = \int_{-\infty}^{\infty} f_T(t \mid \boldsymbol{x}) \bar{F}_C(t \mid \boldsymbol{x}) \mathrm{d}t$ is the expected conditional probability that $T \leq C$. Equations (7) and (8) can be used to investigate the asymptotic behavior of the Laplace regression estimator.

We denote by $\boldsymbol{\beta}_p^0$ the true population parameter satisfying $Q_T(p \mid \boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{\beta}_p^0$. As shown by Bottai and Zhang (2010), $\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x}))$ is trivially zero in two

cases, namely (a) $F_C(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p^0 \mid \boldsymbol{x}) = 0$, i.e., no censoring before the quantile; and (b) $F_T(t \mid \boldsymbol{x}) = F_p(t \mid \boldsymbol{x}, \boldsymbol{\beta}_p^0, \sigma_p^0(\boldsymbol{x}))$ for some σ_p^0, i.e., the AL(p) distribution is the true model. In all other situations, censored Laplace regression yields a biased estimator of $\boldsymbol{\beta}_p$.

We define $\hat{\boldsymbol{\beta}}_p$ and $\hat{\sigma}_p$ to be the solutions to

$$\begin{pmatrix} \bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) \\ \bar{S}_{\sigma_p}(\boldsymbol{\beta}_p, \sigma_p(\boldsymbol{x})) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{9}$$

i.e., the asymptotic values of the maximum likelihood estimators. Further, we define σ_p^* such that

$$\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p^*(\boldsymbol{x})) = 0, \tag{10}$$

i.e., the ideal value of σ_p that would make the Laplace regression estimator of $\boldsymbol{\beta}_p$ unbiased. We demonstrate that, for any given $\boldsymbol{x}$, $\sigma_p^*(\boldsymbol{x})$ exists and is unique. Elementary algebra gives that

$$\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x})) = \tag{11}$$
$$(1-p)E_{\boldsymbol{x}}\Big[\frac{\boldsymbol{x}}{\sigma_p(\boldsymbol{x})}\Big\{\int_{-\infty}^{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p^0} \frac{\bar{F}_T(t \mid \boldsymbol{x})}{\bar{F}_p(t \mid \boldsymbol{x}, \boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x}))} \mathrm{d}F_C(t \mid \boldsymbol{x}) - F_C(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p^0 \mid \boldsymbol{x})\Big\}\Big],$$

which is a continuous function of $\sigma_p(\boldsymbol{x})$, and that

$$\lim_{\sigma_p(\boldsymbol{x})\to 0^+} \bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x})) = -\infty, \quad \lim_{\sigma_p(\boldsymbol{x})\to +\infty} \bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x})) = 0^+.$$

Moreover, it can be easily shown that $\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x}))$ has a single change of sign, which permits concluding that there is a unique $\sigma_p^*(\boldsymbol{x})$ satisfying (10). The typical behavior of $\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p(\boldsymbol{x}))$ is exemplified in Figure 2.

The presence of bias is due to the fact that, because of model misspecification, $\hat{\sigma}_p \neq \sigma_p^*$ and thus $\hat{\boldsymbol{\beta}}_p \neq \boldsymbol{\beta}_p^0$. In principle, unbiasedness could be achieved by maximizing (4), the censored AL(p) log-likelihood, with respect to $\boldsymbol{\beta}_p$ only, letting $\sigma_p(\boldsymbol{x}) = \sigma_p^*(\boldsymbol{x})$. Evaluating $\sigma_p^*(\boldsymbol{x})$, however, is not feasible, because equation (11) involves the unknown quantities $F_T(t \mid \boldsymbol{x})$, $F_C(c \mid \boldsymbol{x})$, and $\boldsymbol{\beta}_p^0$ itself.

Although $\hat{\sigma}_p$ does not estimate any population parameter, it measures the dispersion of the data around the quantile and is subject to a maximum entropy interpretation. More importantly, and quite unexpectedly, maximum likelihood estimators of $\boldsymbol{\beta}_p$ proved very accurate. An illustrative example is reported in Figure 3, where unconditional quantiles of highly skewed and multimodal distributions are shown to be almost perfectly estimated with severe censoring, despite the fact that the true distribution is drastically different form the AL.

Empirical evidence suggests that $\hat{\sigma}_p$ and σ_p^* are likely to be quite different, and typically $\hat{\sigma}_p < \sigma_p^*$. The absolute bias, however, increases very slowly as $\hat{\sigma}_p$ departs from σ_p^*, which clarifies how, but does not explain why, censored Laplace estimators are robust to model misspecification. Extensive simulation results are presented in the next section.

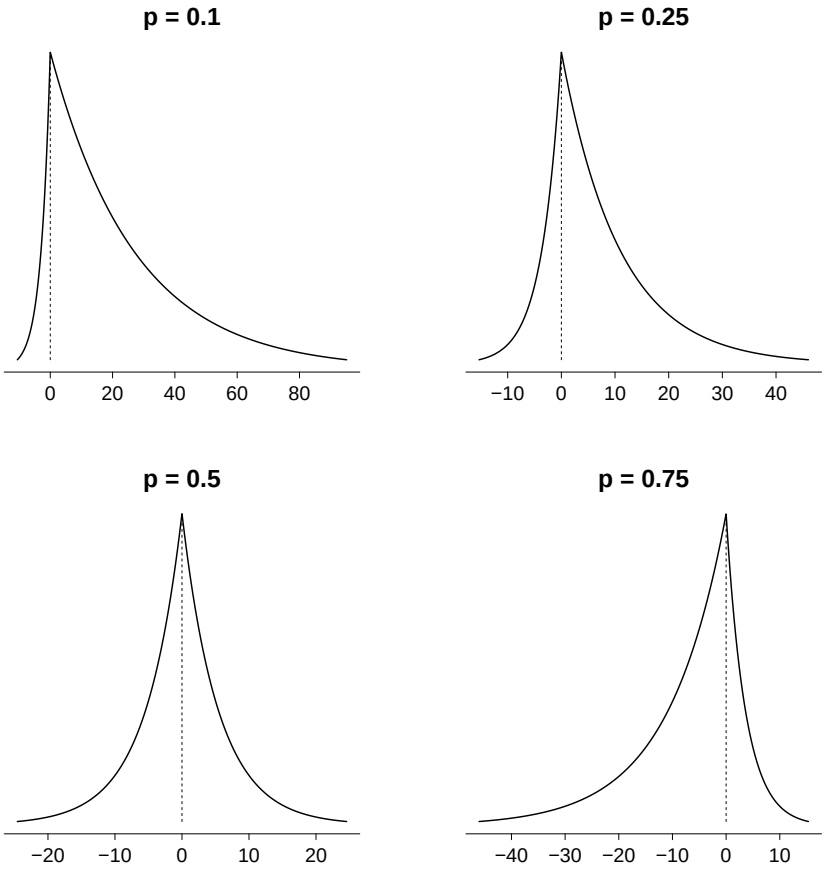

Figure 1: Shape of different asymmetric Laplace distributions with location equal to 0 and scale equal to 1.

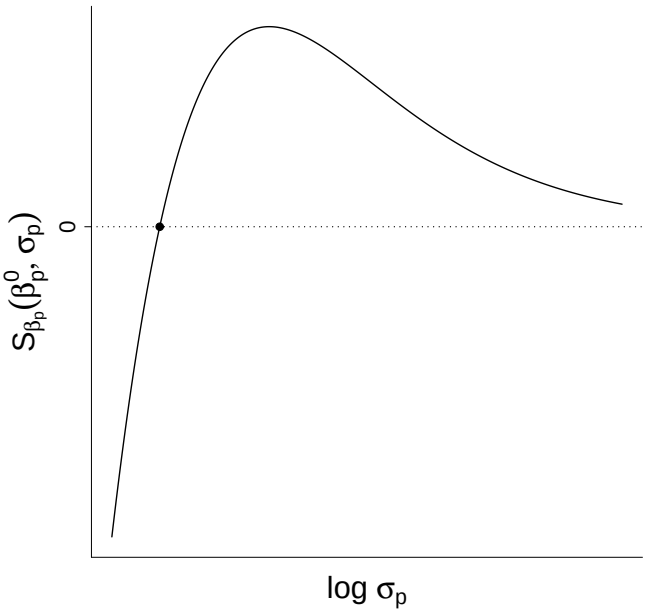

Figure 2: Shape of $\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p^0, \sigma_p)$. The function goes to $-\infty$ as σ_p tends to 0, and to 0^+ as σ_p tends to $+\infty$. At σ_p^*, the function crosses the zero, i.e., solving $\bar{S}_{\boldsymbol{\beta}_p}(\boldsymbol{\beta}_p, \sigma_p^*) = 0$ yields an unbiased estimator of $\boldsymbol{\beta}_p$.

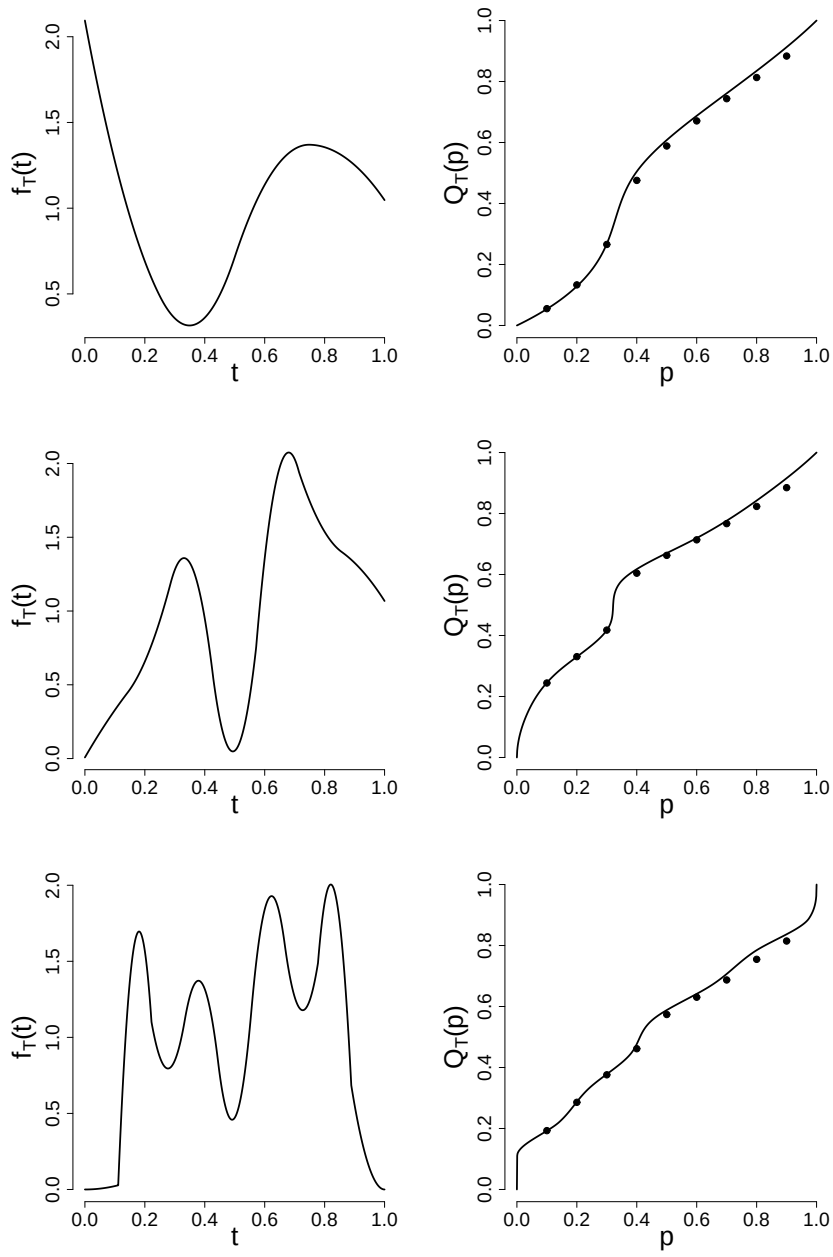

Figure 3: PDF of multimodal, highly skewed distributions (left) and the corresponding QF (right). The dots represent the deciles $\hat{\beta}_{0.1}, \ldots, \hat{\beta}_{0.9}$, obtained solving (9) when T and C have the same distribution, such that the expected proportion of censoring is 0.5.

4 Simulation results

We conducted an extensive simulation study to evaluate the performance of censored Laplace regression and compared it with Portnoy's (2003) estimator, which represents the most commonly used method and is implemented in the `quantreg` R package. To implement AL regression, we modeled the scale parameter as $\log(\sigma_p(\boldsymbol{x} \mid \boldsymbol{\eta}_p)) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\eta}_p$ and used a gradient-based algorithm (Bottai, Orsini and Geraci, 2014) to optimize the log-likelihood (4) with respect to $(\boldsymbol{\beta}_p, \boldsymbol{\eta}_p)$.

The following procedure was used to randomly generate a variety of scenarios, corresponding to different joint distributions of $(\boldsymbol{x}, T, C)$.

- *Covariates.* Six covariates were used: (x_1, x_2, x_3) binary, and (x_4, x_5, x_6) taking on integer values between 1 and 10. For each scenario, a different joint distribution of $\boldsymbol{x} = [1, x_1, \ldots, x_6]^{\mathrm{T}}$ was defined by randomly assigning a probability weight, say w, to each of the $2^3 \times 10^3 = 8000$ covariate patterns. Weights were generated as $w = w_1 \times w_2$, where w_1 was binary with $P(w_1 = 1) = 0.01$, and $w_2 \sim Exp(1)$.

- *Response variable.* The conditional QF of T was defined as $Q_T(p \mid \boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}(p)$, where $\boldsymbol{\beta}(p) = [\beta_0(p), \beta_1(p), \ldots, \beta_6(p)]^{\mathrm{T}}$ were obtained as the interpolating splines (Hyman, 1983) between $p = (0, 1/k, 2/k, \ldots, 1)$ and a set of monotonically increasing values $\boldsymbol{b} = (b_0, b_1, b_2, \ldots, b_k)$. The value of k was randomly selected between 3 and 6, while $\boldsymbol{b}$ was generated as follows: first, b_0 and b_k were drawn from a $U(-5, 5)$ distribution; then, $b_1, \ldots, b_{k-1}$ were drawn from a $U(b_0, b_k)$ distribution.

- *Censoring variable.* The censoring variable was defined to have a $U(0, \theta(\boldsymbol{x}))$ distribution, with $\theta(\boldsymbol{x})$ such that the probability of censoring was a prespecified value $\alpha(\boldsymbol{x})$. For each scenario and each covariate pattern, $\alpha(\boldsymbol{x})$ was

drawn from a $U(0.20, 0.30)$ distribution, leading to an average censoring of 0.25.

In total, $B = 10,000$ scenarios were generated. For each scenario, $R = 250$ simulated datasets were used to evaluate the following quantities:

$$\text{bias}(\hat{\boldsymbol{\beta}}_p) = E_{\boldsymbol{x}}\left[\mid p - F_T(\boldsymbol{x}^{\mathrm{T}} E[\hat{\boldsymbol{\beta}}_p]) \mid\right], \tag{12}$$

$$\text{mse}(\hat{\boldsymbol{\beta}}_p) = E_{\boldsymbol{x}}\left[\{p - F_T(\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_p)\}^2\right]. \tag{13}$$

The first quantity measures the absolute bias of an estimator $\hat{\boldsymbol{\beta}}_p$ of $\boldsymbol{\beta}_p$. If the estimator is unbiased, then $E[\hat{\boldsymbol{\beta}}_p] = \boldsymbol{\beta}_p$ and $\text{bias}(\hat{\boldsymbol{\beta}}_p) = 0$. Intuitively, if $\text{bias}(\hat{\boldsymbol{\beta}}_p) = \epsilon$, the quantile being estimated is between $p - \epsilon$ and $p + \epsilon$. The second quantity measures the mean squared error (MSE) as the dispersion of $F_T(\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_p)$ around $p = F_T(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_p)$. The above measures of bias and MSE are averaged with respect to the distribution of $\boldsymbol{x}$ and are unaffected by the scale of T, C, and $\boldsymbol{x}$.

Results are displayed in Tables 1 and 2 and Figures 4 and 5. The bias of AL regression was only slightly larger than that of Portnoy's estimator, and was generally negligible. For example, with $n = 1000$, the bias at the 6^{th} decile was smaller than 0.02 in about 95% of scenarios, and excedeed 0.03 in only 1% of cases. However, while the bias of Portnoy's estimator was fairly the same at all quantiles, that of AL regression increased sharply at large p. Yet, the absolute bias at the 8^{th} decile was generally smaller than 0.03, which might be considered acceptable in most applications. Empirical evidence suggests that one should not estimate quantiles larger than $1 - p_c$, where p_c is the proportion of censored data.

Remarkably, the AL estimator exhibits empirical consistency, as the absolute bias seems to decrease with the sample size. Moreover, its MSE appeared to be nearly always smaller than that of Portnoy's estimator.

To have a term of comparison, we also applied ordinary quantile regression ignoring censoring. With this approach, we found a significant bias even at small quantiles. For instance, with $n = 1000$, the bias at the 2^{nd} decile was larger than 0.05 in more than 70% of scenarios. At larger quantiles, the bias excedeed 0.10 in most scenarios, and was commonly above 0.15. This demonstrated that the relatively small bias observed for AL regression could not be attributed to the fact that censoring was ignorable.

Table 1: Summary statistics of bias

		$p = 0.2$			$p = 0.4$			$p = 0.6$			$p = 0.8$		
		AL	POR	QR	AL	POR	QR	AL	POR	QR	AL	POR	QR
$n = 250$	1^{st} quartile	.01	.01	.04	.01	.01	.08	.01	.01	.11	.02	.01	.10
	median	.02	.01	.05	.01	.01	.10	.02	.01	.12	.03	.01	.11
	3^{rd} quartile	.02	.01	.07	.02	.01	.11	.03	.01	.14	.04	.01	.12
	centile 0.90	.02	.02	.08	.02	.02	.12	.04	.02	.15	.05	.02	.13
	centile 0.95	.03	.02	.08	.03	.02	.13	.04	.02	.15	.05	.02	.14
	centile 0.99	.03	.03	.09	.04	.03	.15	.05	.03	.17	.06	.03	.15
$n = 500$	1^{st} quartile	.00	.00	.05	.00	.00	.09	.01	.00	.11	.02	.00	.10
	median	.01	.01	.06	.01	.01	.10	.01	.01	.12	.02	.01	.11
	3^{rd} quartile	.01	.01	.07	.01	.01	.11	.02	.01	.13	.03	.01	.12
	centile 0.90	.01	.01	.08	.02	.01	.12	.03	.01	.14	.03	.01	.13
	centile 0.95	.01	.01	.09	.02	.01	.13	.03	.01	.15	.04	.01	.13
	centile 0.99	.02	.02	.09	.03	.02	.15	.04	.02	.16	.04	.02	.14
$n = 1000$	1^{st} quartile	.00	.00	.05	.00	.00	.09	.01	.00	.12	.01	.00	.10
	median	.00	.00	.06	.01	.00	.10	.01	.00	.12	.02	.00	.11
	3^{rd} quartile	.00	.00	.07	.01	.00	.11	.02	.00	.13	.02	.00	.11
	centile 0.90	.01	.01	.08	.01	.01	.12	.02	.01	.14	.03	.01	.12
	centile 0.95	.01	.01	.09	.02	.01	.13	.02	.01	.15	.03	.01	.13
	centile 0.99	.01	.01	.10	.02	.01	.14	.03	.01	.16	.03	.01	.14

Descriptive statistics of the absolute bias (equation 12) of asymmetric Laplace (AL), Portnoy's (POR) and standard quantile regression (QR) estimators of $\boldsymbol{\beta}_p$, across $10,000$ simulation scenarios, at $p = (0.2, 0.4, 0.6, 0.8)$ and three different sample sizes, $n = 250, 500, 1000$.

Table 2: Summary statistics of root MSE

		$p = 0.2$			$p = 0.4$			$p = 0.6$			$p = 0.8$		
		AL	POR	QR	AL	POR	QR	AL	POR	QR	AL	POR	QR
$n = 250$	1^{st} quartile	.07	.07	.07	.08	.09	.11	.09	.09	.14	.08	.08	.14
	median	.07	.07	.08	.08	.09	.12	.09	.10	.15	.09	.09	.14
	3^{rd} quartile	.07	.08	.09	.08	.09	.13	.09	.10	.16	.09	.09	.15
	centile 0.90	.08	.08	.09	.09	.10	.14	.09	.10	.16	.10	.10	.15
	centile 0.95	.08	.09	.10	.09	.10	.15	.10	.11	.17	.11	.12	.16
	centile 0.99	.09	.10	.10	.09	.10	.16	.10	.12	.18	.12	.14	.16
	P(MSE $>$ MSE_{AL})		.89	.74		.99	1.0		.96	1.0		.57	1.0
$n = 500$	1^{st} quartile	.05	.05	.06	.06	.06	.10	.06	.06	.13	.06	.06	.12
	median	.05	.05	.07	.06	.07	.11	.06	.07	.14	.06	.06	.13
	3^{rd} quartile	.05	.05	.08	.06	.07	.12	.07	.07	.14	.06	.06	.13
	centile 0.90	.05	.06	.09	.06	.07	.13	.07	.07	.15	.07	.07	.14
	centile 0.95	.05	.06	.09	.06	.07	.14	.07	.07	.16	.07	.07	.14
	centile 0.99	.06	.07	.10	.07	.07	.15	.07	.08	.16	.08	.09	.15
	P(MSE $>$ MSE_{AL})		.97	.97		.98	1.0		.94	1.0		.56	1.0
$n = 1000$	1^{st} quartile	.03	.04	.06	.04	.05	.10	.04	.05	.12	.04	.04	.11
	median	.03	.04	.07	.04	.05	.11	.05	.05	.13	.04	.04	.12
	3^{rd} quartile	.03	.04	.08	.04	.05	.12	.05	.05	.14	.05	.04	.12
	centile 0.90	.04	.04	.08	.04	.05	.13	.05	.05	.15	.05	.05	.13
	centile 0.95	.04	.04	.09	.05	.05	.13	.05	.05	.15	.05	.05	.13
	centile 0.99	.04	.04	.10	.05	.05	.15	.05	.05	.16	.05	.06	.14
	P(MSE $>$ MSE_{AL})		.97	1.0		.98	1.0		.89	1.0		.48	1.0

Descriptive statistics of the root mean squared error (equation 13). We also report the proportion of scenarios in whichthe MSE of AL regression was smaller than that of other estimators.

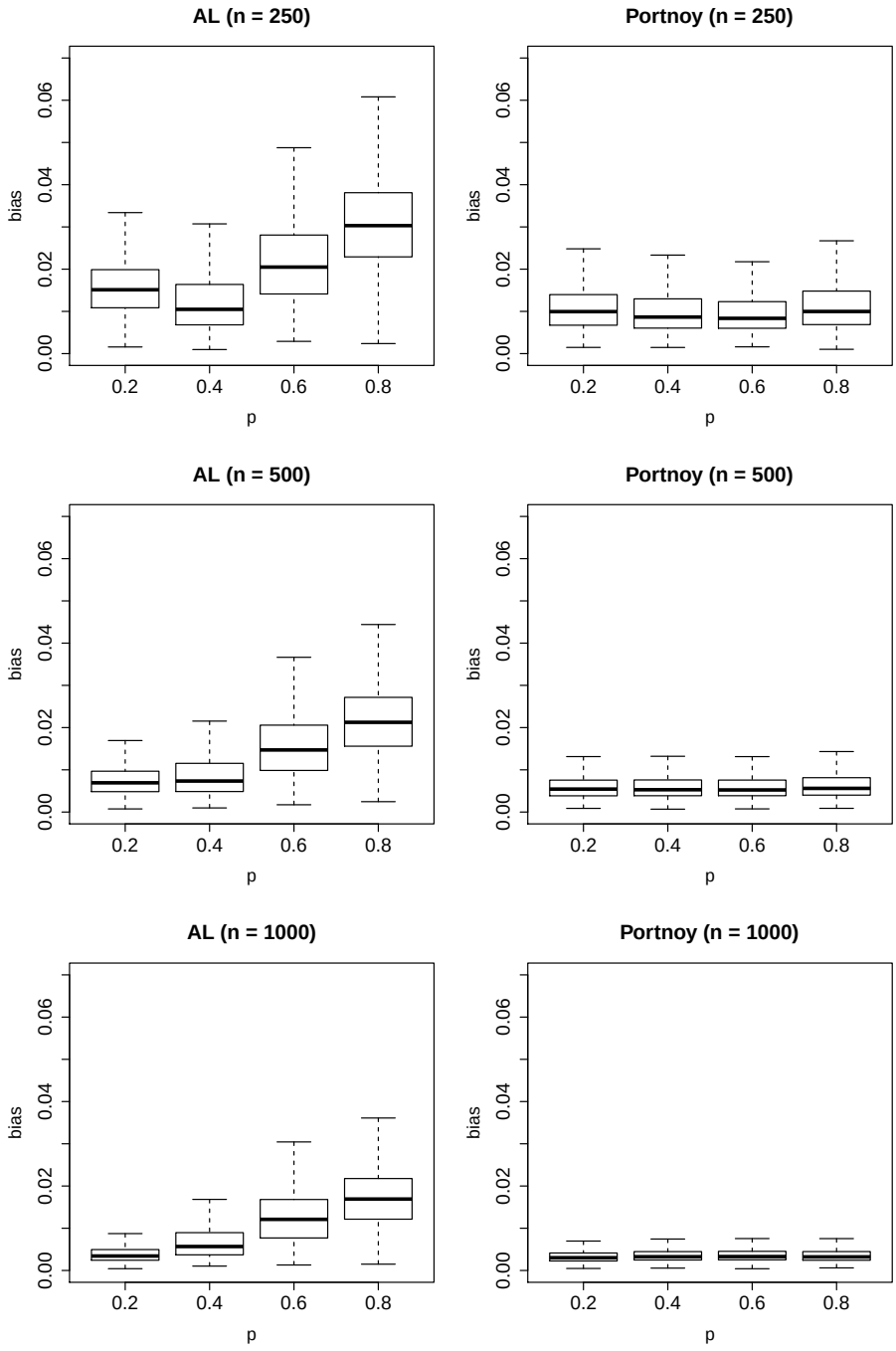

Figure 4: Boxplot of the bias (equation 12) of AL and Portnoy's estimators of $\boldsymbol{\beta}_p$, at different values of p, for $n = 250, 500$, and 1000).

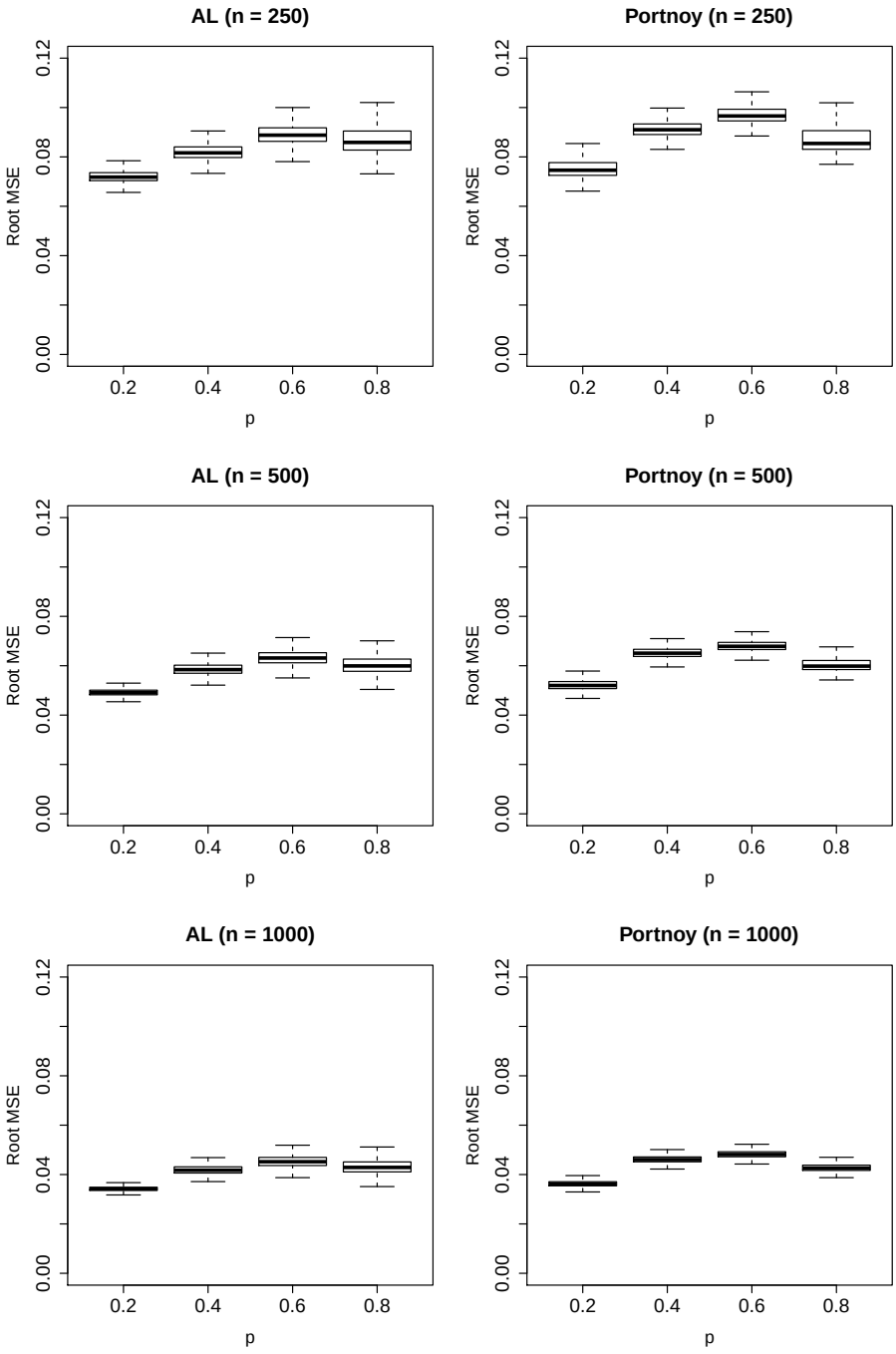

Figure 5: Boxplot of the root MSE (equation 13) of AL and Portnoy's estimators of $\boldsymbol{\beta}_p$, at different values of p, for $n = 250, 500$, and 1000.

5 Using Laplace regression for big data

While computation times of AL regression grow linearly along with the sample size, those of other methods grow at an approximately quadratic rate. Additionally, inference on estimators like Portnoy's is bootstrap-based, as no simple closed-form asymptotic covariance matrix is available. Instead, inference on AL regression can be performed by treating equations (5) and (6) as estimating equations, and using standard methods-of-moment asymptotic theory.

We considered a scenario like those described in Section 4, with six covariates plus intercept. For different values of n, Figure 6 shows the average computation time to fit the model once, using a 64-bit operating system with dual core processor 2.80/2.93 GHz, 4,00 GB RAM.

With $n = 50,000$, AL regression runs in about 0.5 seconds, while Portnoy's method takes, on average, almost 4 minutes. With $n = (100, 250, 500) \times 10^3$ (not shown in Figure 6), average computation times to fit AL regression are 1.3, 3.7, and 8.1 seconds, respectively. Based on extrapolation, we estimated that Portnoy's method will run in about 16 minutes with $n = 100,000$; 1h40' with $n = 250,000$; and 6h50' with $n = 500,000$. Note that the code for AL regression is written in plain R language, while Portnoy's estimator, as implemented in the `quantreg` R package, uses an efficient Fortran routine.

Imagine a plausible research situation, and suppose to estimate the deciles ($p = 0.1, 0.2, \ldots, 0.9$) of a censored outcome, and to formulate 20 different regression models. Using AL, each quantile is estimated separately and the fitting routine is called $9 \times 20 = 180$ times. With Portnoy's method, all quantiles are estimated at once, but bootstrap is needed to compute standard errors. Using $R = 100$ bootstrap replications, the fitting routine is called $20 \times 100 = 2,000$ times. Assuming a sample size $n = 500,000$, which is not uncommon in epidemi-

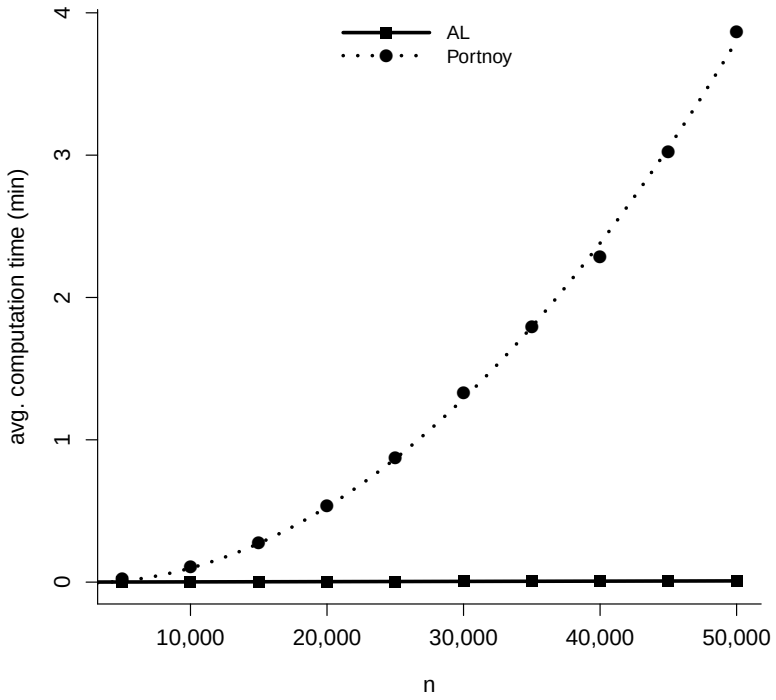

Figure 6: Average computation times for Portnoy's and Laplace regression estimators, at different values of n. The lines represent a quadratic interpolation.

ological studies, the analysis can be completed in about 24 minutes using Laplace regression, or in one and a half year using Portnoy's method.

Peng and Huang's (2008) (also implemented in `quantreg`) and Wang and Wang's (2009) methods are even slower than Portnoy's. Frumento and Bottai's (2016) estimator, which is implemented in the `ctqr` package, is relatively fast and does not require bootstrap. However, it is not nearly as fast as the described Laplace regression, which stands as the only computationally convenient method to fit censored quantile regression with large datasets.

6 Conclusions

The interest in censored Laplace regression is motivated by three main reasons. First, a biased estimator is found in practice to be almost perfectly unbiased, which makes it intriguing to investigate its properties. Second, compared with the existing methods, Laplace regression is much simpler computationally. In particular, no estimation of F_T is needed, thus avoiding complicated and time-consuming procedures; and no bootstrap is required to perform inference. This makes it suitable to be applied to big data, which represents a desirable feature in many applied situations. Third, differently from the other mentioned approaches, Laplace regression can be easily extended to handle truncated data, include frailty terms, and model non linear quantiles.

We discussed the role of the scale parameter of the AL distribution, and demonstrated that non-trivial conditions for unbiasedness exist. Such conditions, however, do not corresponds to those imposed by likelihood maximization.

In a wide simulation study, we showed that maximizing the likelihood of the Laplace distribution yields very reliable estimators of quantiles, despite model misspecification. This is consistent with previous findings and justifies using this method in real-data applications. This form of robustness is not shared by other distributions: for example, fitting a Normal model on censored data would not yield a good estimator of the mean, unless the true distribution is Normal or very close to it.

Specifying how covariates enter $\sigma_p(\boldsymbol{x})$ is very important in practice. A poor modeling of $\sigma_p(\boldsymbol{x})$ may exacerbate model misspecification and cause additional bias. On the other hand, parametrizing $\sigma_p(\boldsymbol{x})$ as a flexible function of covariates permits achieving a good fit of the data, making model misspecification less relevant.

An R code for Laplace regression is provided upon request to the Authors. A Stata command is also available (Bottai and Orsini, 2013).

References

Bottai, M. and Orsini, N. (2013). "A command for Laplace regression", *The Stata Journal* **13**(2), 302–314.

Bottai, M. and Zhang, J. (2010). "Laplace regression with censored data", *Biometrical Journal* **52**(4), 487–503.

Bottai, M., Orsini, N., and Geraci, M. (2014). "A gradient search maximization algorithm for the asymmetric Laplace likelihood", *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655.2014.908879.

Efron, B. (1967). "The two sample problem with censored data", *Proceedings of the 5th Berkeley Symposium*, Vol. 4, 831–853. Berkeley: University of California Press.

Frumento, P., and Bottai, M. (2016). "An estimating equation for censored and truncated quantile regression", *Computational Statistics and Data Analysis*, doi: 10.1016/j.csda.2016.08.015. [In press].

Hyman, J. M. (1983). "Accurate monotonicity preserving cubic interpolation", *SIAM Journal on Scientific and Statistical Computing* **4**,645–654.

Koenker, R. (2005). *Quantile regression*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

Koenker, R. (2008). "Censored quantile regression redux", *Journal of Statistical Software*, **27**(6).

Koenker, R. (2015). *quantreg: Quantile Regression*, R package version 5.11. http://CRAN.R-project.org/package=quantreg.

Leng, C., and Tong, X. (2013). "A Quantile Regression Estimator for Censored Data", *Bernoulli* **19**(1), 344–361.

Peng, L. and Huang, Y. (2008). "Survival analysis with quantile regression models", *Journal of the American Statistical Association* **103**(482), 637–649.

Portnoy, S. (2003). "Censored regression quantiles", *Journal of the American Statistical Association* **98**(464), 1001–1012.

Wang, H. J. and Wang, L. (2009). "Locally weighted censored quantile regression", *Journal of the American Statistical Association* **104**(487), 1117–1128.

Address for correspondence:

Karolinska Institutet, Institute of Environmental Medicine, Unit of Biostatistics

Nobels väg 13, 171 77 Stockholm, Sweden.

E-mail: paolo.frumento@ki.se