A bootstrap estimator for the Student-t regression model

Donald M. Pianto
Departmento de Estatística,Universidade de Brasília

April 30, 2010

Abstract

The Student-t regression model suffers from monotone likelihood. This means that the likelihood achieves its maximum value at infinite values of one or more of the parameters, in this case the unknown degrees of freedom. This leads to problems when one uses iterative algorithms to locate the solutions to the non-linear equations generated by the likelihood. Fonseca et al. (2008) deal with this problem by using the Jeffreys priors. We implement a bootstrap estimator which is based on resampling the data until samples without monotone likelihood are encountered. Results from this analysis will be presented.
Keywords: monotone likelihood; bootstrap; t errors

1 Introduction

In Pianto and Cribari Neto (2010) the authors introduce a bootstrap estimator to deal with monotone likelihood in a model for speckled imagery developed by Frery et al. (1997). Monotone likelihood occurs when the maximum of the likelihood is attained when one or more of the parameters goes to ∞. The bootstrap estimator compared favorably with an estimator based on the Jeffreys prior and is comparable with the estimator in Firth (1993).

Fonseca et al. (2008) demonstrate that the regression model with Student-t errors also suffers from monotone likelihood. In this paper, we propose to implement a boostrap estimator and the Firth estimator for this model.

2 Regression Model with t-errors

Consider an n-vector of observations $y = (y_1, \ldots, y_n)'$. The model we study is given by

$$y = X\beta + \epsilon, \tag{1}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is the error vector where the components are independent and identically distributed according to the Student-t distribution with location zero, scale σ and degrees of freedom ν. $X = [x_1, \ldots, x_n]'$ is the $n \times k$ matrix of explanatory variables (assumed to be of full rank). The parameter space is given by $\theta = (\beta, \sigma, \nu) \in \mathbb{R}^k \times (0, \infty)^2$.

The likelihood is given by

$$L(\beta, \sigma, \nu; y, X) = \frac{\Gamma(\frac{\nu+1}{2})^n}{\Gamma(\frac{\nu}{2})^n (\pi\nu)^{n/2} \sigma^n} \prod_{i=1}^{n} \left[1 + \frac{1}{\nu} \left(\frac{y_i - x_i'\beta}{\sigma} \right)^2 \right]^{-(\nu+1)/2}. \tag{2}$$

The reduced log-likelihood is given by

$$\begin{aligned} \ell(\beta, \sigma, \nu; y, X) \quad &= \quad n \left\{ \log(\Gamma(\frac{\nu+1}{2})) - \log(\Gamma(\frac{\nu}{2})) - \log(\nu)/2 - \log(\sigma) \right\} \\ &\quad -\frac{\nu+1}{2} \sum_{i=1}^{n} \log \left[1 + \frac{1}{\nu} \left(\frac{y_i - x_i'\beta}{\sigma} \right)^2 \right]. \end{aligned} \tag{3}$$

In Fernandez and Steel (1999) the authors show the following. With $s(\beta)$ equal to the number of observations for which $y_i = x_i'\beta$, they define $\tilde{\beta} = \text{argmax}_\beta s(\beta)$ and $d_0 = s(\tilde{\beta})/(n - s(\tilde{\beta}))$. Note that $s(\tilde{\beta}) \geq k$. When $\nu < d_0$ the likelihood goes to infinity when $\sigma \to 0$. Hence, we restrict the parameter space to $\theta = (\beta, \sigma, \nu) \in \mathbb{R}^k \times (0, \infty) \times (d_0, \infty)$.

If $\hat{\beta}$ and $\hat{\sigma}$ are obtained from MLE assuming normal errors and we define $\hat{z}_i = (y_i - x_i'\hat{\beta})/\hat{\sigma}$ then Theorem 1 of Fonseca et al. (2008) states that if $\sum_{i=1}^n (\hat{z}_i^2 - 1)^2 < 2n$ then the likelihood is monotone.

3 A resampling solution for monotone likelihood

Loughin (1998) discourages the use of bootstrap resampling methods for bias estimation and correction when monotone likelihood has a significant probability of occurring in the resamples. He argues that the results obtained are only valid conditional on the MLE estimates being finite and he demonstrates a severe underestimation of the bias when monotone likelihood occurs frequently.

However, Cribari-Neto et al. (2002) successfully implement just such bootstrap bias corrections. Their results remain conditional on the MLE estimates being finite, however, their bias correction is quite good (especially the "better" bootstrap bias estimate based on Efron (1990)).

We briefly discuss Efron's (1990) "better" bias estimates, and Cribari-Neto et al.'s (2002) adaptation of Efron's estimator.

Let $X \overset{\text{iid}}{\sim} f_\theta(x)$ be a random sample of size N from a distribution F_θ. The non-parametric bootstrap approximates F by $\hat{F}$, the empirical distribution function based on the data. One generates samples from $\hat{F}$ by sampling with replacement from the data. A non-parametric bootstrap bias estimate is formed by averaging the bootstrap parameter estimate, $\hat{\theta}_i$, over many samples from $\hat{F}$ and subtracting the original estimate. A bootstrap bias corrected estimate (BBC) is then formed by subtracting this estimated bias from the original estimate:

$$\hat{\theta}_{\text{BBC}} = \hat{\theta} - \left(\frac{1}{B} \sum_{i=1}^{B} \hat{\theta}_i - \hat{\theta} \right). \tag{4}$$

Efron's (1990) suggestion requires a little notation. In each of the B bootstrap resamples above, the sample can be described by the weight that each observation receives in the new empirical distribution function. For the original sample, each observation received weight $1/N$. This can be succinctly recorded in a vector $P^0 = 1/N(1, \ldots, 1)$. For the i-th bootstrap sample this vector becomes

$$P^i = \frac{1}{N} \left(\#\{X_1\}_i, \ldots, \#\{X_N\}_i \right),$$

where $\#\{X_j\}_i$ represents the number of times X_j occurs in the i-th bootstrap sample. The vector is called the resampling vector. Efron's idea is based on the possibility of writing the parameter estimate as a closed function of the data using the vector P^0. For example, the estimate of the mean can be written $\bar{X} = T(P^0) = P^0 \cdot x$. An estimate of the second moment could be written $\bar{X^2} = T(P^0) = P^0 \cdot (x^2)$.

In order to accelerate the convergence of the bias estimate such that fewer bootstrap repetitions are required, Efron suggests that when calculating the estimate of the bias that one subtracts the parameter estimate resulting from using

$$P^* = \frac{1}{B}\sum_{i=1}^{B} P_i,$$

$T(P^*)$. The new better bootstrap bias correction (BBBC) estimate would then be

$$\hat{\theta}_{\text{BBBC}} = \hat{\theta} - \left(\frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i - T(P^*)\right). \tag{5}$$

It makes intuitive sense that this would yield faster convergence to an accurate bias estimate as the number of bootstrap samples (B) increases, since for small values of B the obtained samples could be significantly different from the original sample. In this case the true bias and the difference between the bootstrap samples and the original sample would get confounded in the bias estimate. We note that as $B \to \infty$ for fixed N, $P^* \to P^0$.

As Cribari-Neto et al. (2002) note, most MLEs do not have closed forms. To implement the BBBC they write the estimating equations as a function of P^0 and then use the estimate obtained by replacing P^0 with P^* in the estimating equations to correct the bias.

One can write the Student-t log-likelihood as a function of P^0 to obtain

$$\begin{aligned}\ell(P^0, \theta) &= N\left\{\log(\Gamma(\frac{\nu+1}{2})) - \log(\Gamma(\frac{\nu}{2})) - \log(\nu)/2 - \log(\sigma)\right\} \\ &\quad -\frac{\nu+1}{2} NP^0 \cdot \log\left[1 + \frac{1}{\nu}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right].\end{aligned}$$

where $T(P^0) = (\hat{\theta})$. One can then calculate $T(P^*)$ obtained by replacing P^0 by P^* in Equation (6) and insert it into Equation (5) to generate their BBBC estimate.

We suggest an estimator based on bootstrap resampling which works even in the presence of monotone likelihood. We first define the estimator and then comment on its properties. Fonseca et al. (2008) defined a divergence criterion under which the Student-t regression model likelihood is monotone. Our suggestion is to perform a non-parametric bootstrap of the data, keeping only those bootstrap samples where the divergence criterion is not satisfied until a certain pre-determined number of non-diverging bootstrap samples are obtained. Then our estimate is given by $T(P^*)$ as calculated in Equation (6). Whereas this estimator requires much resampling and checking of the divergence criterion, it only requires one non-linear maximization.

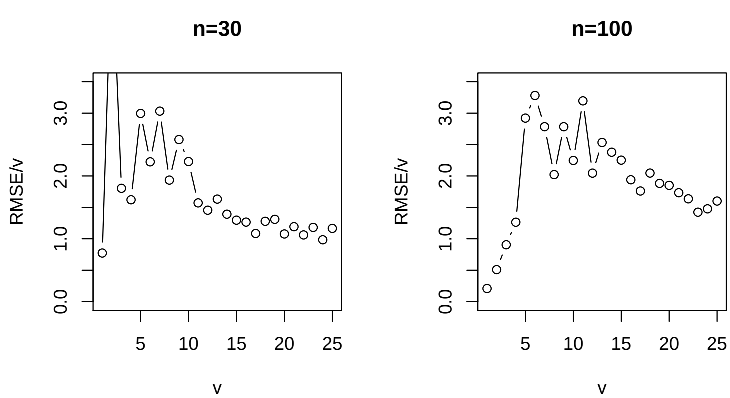
For our proposed estimator to be useful it must be consistent and it must exist and be finite for the great majority of samples. The following are the possible problems that could occur with the estimator: (i) no bootstrap samples may exist which do not satisfy the divergence criterion; (ii) even if we have B valid bootstrap samples, the pseudo-sample corresponding to P^* may satisfy the divergence criterion; (iii) the estimator may not be consistent (may not converge in probability to the true parameter value). In the following paragraphs we shall address each of the problems listed above. We will study these properties.

4 Monte Carlo Results

For comparability, we follow the Monte Carlo experiment in Fonseca et al. (2008). We take $k = 5$, $x_i^{'} = (1, x_{1i}, x_{2i}, x_{3i}, x_{4i})$, $\sigma^2 = 1.5$ and $\beta' = (2, 1, 0.3, 0.9, 1)$ and assume independent Gaussian distributions for each component of of x_i with marginal variances equal to 1. The Monte Carlo experiment is repeated for the double (n, ν) with $n \in \{30, 100\}$ and $\nu \in \{1, 2, \ldots, 25\}$.

Our preliminary results can be found in Figure 1. We note that our results exhibit the same behavior as the estimators implemented in Fonseca et al. (2008) although the root mean squared error for their estimators is smaller. We also intend to provide coverage rates for this estimator.

Figure 1: Relative Root Mean Squared Error of the bootstrap estimator of the degrees of freedom



5 Conclusion

We will implement Firth's correction (Firth, 1993) and compare it with the bootstrap estimator and the estimators in Fonseca et al. (2008). We will also compare the estimators on real data to asses their performance.

References

Cribari-Neto, F., Frery, A. C., Silva, M. F., 2002. Improved estimation of clutter properties in speckled imagery. Computational Statistics & Data Analysis 40, 801–824.

Efron, B., 1990. More efficient bootstrap computations. Journal of the American Statistical Association 85, 79–89.

Fernandez, C., Steel, M. F. J., 1999. Multivariate Student-t regression models: Pitfalls and inference. Biometrika 86 (1), 153–167.

Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38.

Fonseca, T. C. O., Ferreira, M. A. R., Migon, H. S., 2008. Objective Bayesian analysis for the Student-t regression model. Biometrika 95 (2), 325–333.

Frery, A. C., Müller, H.-J., Yanasse, C., Sant'Anna, S. J. S., 1997. A model for extremely heterogeneous clutter. IEEE Transactions on Geoscience and Remote Sensing 35, 648–659.

Loughin, T. M., 1998. On the bootstrap and monotone likelihood in the Cox proportional hazards regression model. Lifetime Data Analysis 4, 393–403.

Pianto, D. M., Cribari Neto, F., 2010. Dealing with monotone likelihood in a model for speckled data, preprint.

Sartori, N., 2006. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. Journal of Statistical Planning and Inference 136, 4259–4275.

Vasconcellos, K. L., Cordeiro, G. M., 2000. Bias corrected estimates in multivariate student t regression models. Communications in Statistics 29, 797–822.