Valid properties of truncated Student-$t$ regression model with applications in analysis of censored data

Chi ZHANG$^{a,\#}$, Guo-Liang TIAN$^{b,\#}$, Yibo ZHAI$^{b}$, and Yu FEI$^{c,*}$

$^a$College of Economics, Shenzhen University, Shenzhen 518060, Guangdong Province, P. R. China

$^b$Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, P. R. China

$^c$School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, Yunnan Province, P. R. China

$\#$Equally contributed

$^*Corresponding author’s email: feiyu@ynufe.edu.cn

Abstract. Kim (2008) introduced an incorrect stochastic representation (SR) for the truncated Student-$t$ (Tt) random variable. By pointing out that the gamma mixture based on a truncated normal distribution actually cannot result in a true Tt distribution, in this paper, we first propose three correct SRs and then recalculate the corresponding moments of the Tt distribution. Different from those derived by following the invalid SR of Kim (2008), the correct moments of the Tt distribution play a crucial role in parameter estimations. Based on the third SR proposed and the correct expressions of truncated moments, expectation–maximization (EM) algorithms are developed for calculating the maximum likelihood estimates of parameters in the Tt distribution. Extensions to a Tt regression model and a $t$ interval–censored regression model are provided as well. Simulated experiments are conducted to evaluate the performance of the proposed methods. Finally, two real data analyses corroborate the theoretical results.

Keywords: EM algorithm; Interval–censored regression model; Stochastic representation; Truncated Student-$t$ distribution; Truncated Student-$t$ regression model.
1. Introduction

In both frequentist and Bayesian statistics, Student-$t$ distribution is usually used as the sampling distribution of certain test statistics under the assumption of normality and a robust alternative to the normal distribution in the analysis of continuous observations with heavy tails or outliers (Johnson et al., 1995, Ch.28). However, its mean is undefined if the degree of freedom $\nu \leq 1$; its variance is undefined if $0 < \nu \leq 1$ and is infinite if $1 < \nu \leq 2$. Especially when $\nu = 1$, the $t$ distribution reduces to the Cauchy distribution whose mean, variance, skewness and kurtosis are all undefined. These disadvantages definitely limit its application range to certain extent since, in general, the first two moments not only are utilized for providing descriptive information on the distribution, but also are employed for estimating the parameters of the distribution in the method of moments.

The main reason for such undefined mean and/or variance is that the support of the $t$ variate is the whole real line. In practice, it is impossible that a realization of the $t$ random variable (r.v.) takes the value of $-\infty$ or $\infty$. In fact, the observed data restricted within certain intervals are frequently encountered in survival analysis, dosage–response studies, biological assays, army selection and other fields. Therefore, to overcome the aforementioned drawbacks associated with the $t$-distribution, one way is to adopt the truncated Student-$t$ ($Tt$) distribution as an alternative to the $t$ distribution since the moments of the doubly $Tt$ distribution always exist and are finite.

Kim (2008) tries to derive the moments of the univariate $Tt$ distribution. The extensions of his work include the discussion of (i) deriving moments in the truncated multivariate $t$ distribution (Ho et al., 2012; Arismendi, 2013), (ii) sampling from $Tt$ distributions (Ho et al., 2012), and (iii) parameter estimation in Student-$t$ censored regression models (Massuia et al., 2015) and the related mixture models (Lachos et al., 2019). These works have been frequently cited, however, their foundation, the derivation of the moments of the univariate $Tt$ distribution, involves a flaw. In Kim (2008), an incorrect stochastic representation (SR) for the r.v. following the standard $Tt$ distribution based on a gamma r.v. and a conditional truncated normal distribution was introduced and was used for deriving the truncated mo-
ments, which is also adopted by Genç (2013). However, the SR cannot result in a true Tt
distribution, as shown in §2.1.1 of the current paper. This incorrectness was inherent in the
aforementioned extension works as well. As a result, some of the obtained conclusions are
unconvincing and a series of modifications should be addressed. Other works (e.g., Garay
et al., 2017; Lachos et al., 2017 and Zeller et al., 2019) mainly used (multivariate) t models
or their mixtures to analyze censored data without imposing SR on the Tt distribution.
Therefore, the first aim of this paper is to propose three valid SRs for the univariate Tt
variate and to provide correct results on moments of the Tt distribution in §2.

Generally speaking, expectation–maximization (EM) algorithms (Dempster et al., 1977)
and minorization–maximization (MM) algorithms (Becker et al., 1997; Lange et al., 2000;
Hunter & Lange, 2004; Tian, Huang & Xu, 2019) are effective methods for calculating the
maximum likelihood estimates (MLEs) of parameters of interest. The simplicity in concept
and stability in convergence make both algorithms to be quite popular in statistical comput-
ing. Massuia et al. (2015) employed an EM-type algorithm to solve the estimation problem
in Student-t censored regression models, which is closely related with the truncated t distri-
bution. Therefore, motivated by the third SR proposed in this paper, the second aim of this
paper is to develop EM algorithms for calculating the MLEs of parameters in both the Tt
distribution and the Tt regression model in §3.

Other contributed works on truncated t distribution include: Nadarajah & Kotz (2006,
2007) provided R programs for computing the related statistic characteristics of truncated t
distributions, such as mean, variance, and cumulative probability. Furthermore, Nadarajah
& Kotz (2008) derived the explicit expressions of the moments and considered the parameter
estimations for the truncated t distribution, in which the moments are expressed as the
Gauss hypergeometric function and the estimation are implemented by both the method of
moments and the method of maximum likelihood.

The rest of the paper is organized as follows. In §4, the Tt regression model is generalized
to the t interval–censored regression model. In §5, simulation studies are conducted to
compare the estimation performances under different assumptions. In §6, two real data sets
are analyzed through the proposed methods. Concluding remarks are given in §7.
2. Truncated Student-\(t\) distribution

The probability density function (pdf) of Student-\(t\) distribution with location parameter \(\mu \in (-\infty, \infty)\), scale parameter \(\sigma > 0\) and \(\nu > 0\) degrees of freedom is

\[
t(x|\mu, \sigma^2, \nu) = \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)\Gamma(1/2)\sqrt{\nu\sigma}} \left[1 + \frac{1}{\nu} \left(\frac{x - \mu}{\sigma}\right)^2\right]^{-(\nu+1)/2}, \quad x \in \mathbb{R} \hat{=} (-\infty, \infty),
\]

where the symbol “\(\hat{=}\)” means equal by definition. We write it by \(X \sim t(\mu, \sigma^2, \nu)\). In particular, when \(\mu = 0\) and \(\sigma = 1\), it reduces to the standard Student-\(t\) distribution with \(\nu\) degrees of freedom, denoted by \(t(\nu)\).

The cumulative distribution function (cdf) of \(t(\nu)\) is

\[
F_\nu(x) = \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)\Gamma(1/2)} \int_{-\infty}^{x} (\nu + t^2)^{-(\nu+1)/2} dt,
\]

which can be calculated by the built-in R function \(pt(x, df, lower.tail = TRUE, log.p = FALSE)\). When \(\nu = 1\), the \(t(1)\) distribution reduces to the Cauchy distribution. Particularly, we have

\[
F_\nu(x) = \begin{cases} 
\frac{1}{2} + \frac{\arctan(x)}{\pi}, & \text{if } \nu = 1, \\
\frac{1}{2} + \frac{x}{2\sqrt{2 + x^2}}, & \text{if } \nu = 2.
\end{cases}
\]

**Definition 1** (Truncated Student-\(t\) distribution). A continuous r.v. \(Y\) is said to follow Student-\(t\) distribution truncated on the interval \([a, b]\), denoted by \(Y \sim \text{Tt}(\mu, \sigma^2, \nu; [a, b])\) and called as truncated Student-\(t\) (Tt) distribution or Type I Tt distribution, if its pdf is

\[
f_\nu(y) = \frac{1}{c} \cdot t(y|\mu, \sigma^2, \nu) \cdot I(a \leq y \leq b), \tag{2.1}
\]

where the normalizing constant is given by

\[
c = F_\nu(\beta) - F_\nu(\alpha), \quad \beta = \frac{b - \mu}{\sigma}, \quad \alpha = \frac{a - \mu}{\sigma}, \tag{2.2}
\]

and \(I(\cdot)\) is the indicator function.

If both \(a\) and \(b\) are finite, then \(Y\) is said to follow the doubly Tt distribution. Next, \(\text{Tt}(\mu, \sigma^2, \nu; [a, \infty))\) and \(\text{Tt}(\mu, \sigma^2, \nu; (-\infty, b])\) are referred to as the left and right Tt distributions, respectively. Naturally, \(\text{Tt}(\mu, \sigma^2, \nu; (-\infty, \infty))\) is the classical t distribution \(t(\mu, \sigma^2, \nu)\). In this section, we provide three valid SRs and correct formulae for the moments of the Tt distribution based on (2.1).
2.1 Correct stochastic representations

Let \( X \sim t(\mu, \sigma^2, \nu) \), the most intuitive SR of \( Y \sim T_t(\mu, \sigma^2, \nu; [a, b]) \) is

\[
Y \overset{d}{=} X | (a \leq X \leq b).
\] (2.3)

It is not conspicuous to motivate an EM algorithm from this SR for calculating the MLEs of \( \{\mu, \sigma^2, \nu\} \). Hence, in order to introduce another two SRs, we first present the concept of the complementary r.v. (Tian et al., 2018). A r.v. \( Y_c \) is called the complementary r.v. of \( Y \), if its pdf is

\[
f_{Y_c}(y) = \frac{\Gamma((\nu + 1)/2)}{(1 - c) \cdot \Gamma(\nu/2) \Gamma(1/2) \sqrt{\nu\sigma}} \left[ 1 + \frac{1}{\nu} \left( \frac{y - \mu}{\sigma} \right)^2 \right]^{-(\nu+1)/2} \cdot [I(y < a) + I(y > b)],
\]

where \( c \) is the constant defined by (2.2); i.e., it has a similar distribution on a complementary interval of \( Y \). We write \( Y_c \sim T_t(\mu, \sigma^2, \nu; [a, b]^c) \).

As a result, the r.v. \( X \sim t(\mu, \sigma^2, \nu) \) can be stochastically represented by

\[
X \overset{d}{=} ZY + (1 - Z)Y_c,
\] (2.4)

where \( Z \sim \text{Bernoulli}(c) \), \( Y \sim T_t(\mu, \sigma^2, \nu; [a, b]) \), \( Y_c \sim T_t(\mu, \sigma^2, \nu; [a, b]^c) \) and \( \{Z, Y, Y_c\} \) are mutually independent. The correctness of the SR (2.4) is easy to verify. Alternatively, another SR for the r.v. \( X \sim t(\mu, \sigma^2, \nu) \) can be expressed as

\[
X \overset{d}{=} Z_0Y + Z_1Y_L + Z_2Y_R = (Y^*)^\top Z^*,
\] (2.5)

where the random vector \( Y^* = (Y, Y_L, Y_R)^\top \) with three components \( Y \sim T_t(\mu, \sigma^2, \nu; [a, b]) \), \( Y_L \sim T_t(\mu, \sigma^2, \nu; (-\infty, a]) \) and \( Y_R \sim T_t(\mu, \sigma^2, \nu; (b, +\infty)) \), the random vector

\[
Z^* = (Z_0, Z_1, Z_2)^\top \sim \text{Multinomial}_3(1; c, c_1, c_2)
\]

with \( c_1 = F_\nu(\alpha) \), \( c_2 = 1 - F_\nu(\beta) \) with \( \{c, \alpha, \beta\} \) are defined in (2.2), and \( \{Z^*, Y, Y_L, Y_R\} \) are mutually independent. More importantly, the third SR (2.5) can motivate an EM algorithm for calculating the MLEs of the parameters \( \{\mu, \sigma^2, \nu\} \).
Remark 1 (Invalid SR in Kim (2008) and other literature) The pdf of $\eta \sim \text{Gamma}(a, b)$ is defined as $b^a \eta^{a-1} e^{-\eta b} / \Gamma(a)$, and it is easy to verify that $E[\log(\eta)] = \frac{\Gamma'(a)}{\Gamma(a)} - \log(b)$. In Kim (2008), the r.v. $T \sim Tt(0, 1; [\alpha, \beta])$ is stochastically represented as

$$T = \eta^{-1/2}Z,$$  \hspace{1cm} (2.6)

where $\eta \sim \text{Gamma}(\nu/2, \nu/2)$ and $Z$ follows the standard normal distribution truncated on the interval $[\eta^{1/2}\alpha, \eta^{1/2}\beta]$, denoted by $Z(\eta) \sim TN(0, 1; [\eta^{1/2}\alpha, \eta^{1/2}\beta])$. However, the SR (2.6) is actually invalid. In fact, the pdf of $T$ is derived as

$$f_T(t) = \int_0^\infty \frac{\eta^2 \exp(-\eta^2/2)}{[\Phi(\eta^{1/2}\beta) - \Phi(\eta^{1/2}\alpha)]\sqrt{2\pi}} dH(\eta), \quad t \in [\alpha, \beta],$$  \hspace{1cm} (2.7)

where $\Phi(\cdot)$ is the cdf of $N(0, 1)$ and $H(\cdot)$ is the cdf of $\eta$. Clearly, the right-hand side of (2.7) is not a $Tt$ density. Similar fallacies have been found in Ho et al. (2012) and Arismendi (2013) when deriving moments of the multivariate $Tt$ distribution.

Definition 2 (Type II truncated Student-$t$ distribution). Let the pdf of the r.v. $T$ be defined by (2.7). Let $W = \mu + \sigma T$, then we say $W$ following the Type II Tt distribution, denoted by $W \sim Tt(\Pi)(\mu, \sigma^2, \nu; [a, b])$, where $a = \mu + \sigma \alpha$ and $b = \mu + \sigma \beta$.

In other words, Kim (2008) wrongly treated the Type II Tt distribution as the Type I Tt distribution.

2.2 Correct results about moments

The aim of this subsection is to provide correct moments of the Tt r.v. $Y \sim Tt(\mu, \sigma^2, \nu; [a, b])$ since Kim (2008) only derived the moments of the Type II Tt distribution based on the SR (2.6). Let $Y_0 \sim Tt(0, 1, \nu; [\alpha, \beta])$ with $\alpha$ and $\beta$ defined in (2.2), we can write

$$Y = \mu + \sigma Y_0,$$  \hspace{1cm} (2.8)

where we only consider the case of $|a|, |b| < \infty$ such that $|\alpha|, |\beta| < \infty$.  

6
2.2.1 Moments of the standard Tt distribution

The moment generating function of $Y_0$ is given by

$$M_{Y_0}(t) = E(e^{tY_0}) = \frac{\Gamma((\nu + 1)/2)\nu^{\nu/2}}{c \cdot \Gamma(\nu/2)\Gamma(1/2)} \int_\alpha^\beta e^{ty} \cdot (\nu + y^2)^{-\nu/2} \, dy$$

with $c$ given by (2.2). Note that the moments of $Y_0$ can be calculated by

$$E(Y_0^k) = M_{Y_0}^{(k)}(t)\bigg|_{t=0} = \frac{d^kM_{Y_0}(t)}{dt^k} \bigg|_{t=0}$$

for $k > 0$. Before we summarize the first four moments of $Y_0$, we denote

$$G_\nu(\ell) = \frac{\Gamma((\nu - \ell)/2)\nu^{\nu/2}}{2c \cdot \Gamma(\nu/2)\Gamma(1/2)}, \quad \ell = 1, 3.$$

**Lemma 1** If $Y_0 \sim \text{Tt}(0, 1; \nu; [\alpha, \beta])$, then

$$E(Y_0) = \begin{cases} \log(1 + \beta^2) - \log(1 + \alpha^2) & \text{if } \nu = 1, \\ \frac{\alpha}{(\nu + \alpha^2)^{\nu/2}} - \frac{\beta}{(\nu + \beta^2)^{\nu/2}} + \int_\alpha^\beta (\nu + y^2)^{-\nu/2} \, dy & \text{if } \nu \neq 1. \end{cases}$$

**Lemma 2** If $Y_0 \sim \text{Tt}(0, 1, \nu; [\alpha, \beta])$, then

$$E(Y_0^2) = \begin{cases} \frac{\beta - \alpha}{\arctan(\beta) - \arctan(\alpha)} - 1 & \text{if } \nu = 1, \\ G_\nu(1) \left[ \frac{\alpha}{(\nu + \alpha^2)^{\nu/2}} - \frac{\beta}{(\nu + \beta^2)^{\nu/2}} + \int_\alpha^\beta (\nu + y^2)^{-\nu/2} \, dy \right] & \text{if } \nu \neq 1. \end{cases}$$

In particular, when $\nu \geq 2$, $E(Y_0^2)$ has the following explicit expressions:

$$E(Y_0^2) = \begin{cases} 2\log\left( \frac{\beta + \sqrt{2 + \beta^2}}{\alpha + \sqrt{2 + \alpha^2}} \right) / \left( \frac{\beta}{\sqrt{2 + \beta^2}} - \frac{\alpha}{\sqrt{2 + \alpha^2}} \right) - 2, & \text{if } \nu = 2, \\ G_\nu(1) \left[ \frac{\alpha}{(\nu + \alpha^2)^{\nu/2}} - \frac{\beta}{(\nu + \beta^2)^{\nu/2}} \right] + \frac{\nu [F_{\nu-2}(\beta_1) - F_{\nu-2}(\alpha_1)]}{(\nu - 2)c}, & \text{if } \nu > 2, \end{cases}$$

where

$$\alpha_1 = \sqrt{\frac{\nu - 2}{\nu}} \quad \text{and} \quad \beta_1 = \beta \sqrt{\frac{\nu - 2}{\nu}}.$$
Lemma 3 If $Y_0 \sim \text{Tt}(0, 1, \nu; [\alpha, \beta])$, then

$$E(Y_0^3) = \begin{cases} \frac{\beta^2 - \alpha^2 - \log(1 + \beta^2) + \log(1 + \alpha^2)}{2 \left[ \arctan(\beta) - \arctan(\alpha) \right]}, & \text{if } \nu = 1, \\ G_{\nu}(1) \left[ \frac{\alpha^2}{3 + \alpha^2} - \frac{\beta^2}{3 + \beta^2} + \log \left( \frac{3 + \beta^2}{3 + \alpha^2} \right) \right], & \text{if } \nu = 3, \\ G_{\nu}(1) \left[ \alpha^2(\nu + \alpha^2)^{-\nu-1/2} - \beta^2(\nu + \beta^2)^{-\nu-1/2} \right] \\ + G_{\nu}(3) \left[ (\nu + \alpha^2)^{-\nu-3/2} - (\nu + \beta^2)^{-\nu-3/2} \right], & \text{if } \nu \neq 1 \text{ and } \nu \neq 3. \end{cases}$$

Lemma 4 Let $Y_0 \sim \text{Tt}(0, 1, \nu; [\alpha, \beta])$, we consider three cases when $E(Y_0^4)$ is calculated.

When $\nu = 1$,

$$E(Y_0^4) = \frac{\beta^3}{3} - \frac{\alpha^3}{3} - \beta + \alpha \frac{\arctan(\beta) - \arctan(\alpha)}{\arctan(\beta) - \arctan(\alpha)} + 1.$$  

When $\nu = 3$,

$$E(Y_0^4) = G_{\nu}(1) \left\{ \frac{\alpha^3}{3 + \alpha^2} - \frac{\beta^3}{3 + \beta^2} + 3(\beta - \alpha) + 3\sqrt{3} \left[ \arctan \left( \frac{\alpha}{\sqrt{3}} \right) - \arctan \left( \frac{\beta}{\sqrt{3}} \right) \right] \right\}.$$  

When $\nu \neq 1$ and $\nu \neq 3$,

$$E(Y_0^4) = G_{\nu}(1) \left[ \frac{\alpha^3}{3 + \alpha^2} - \frac{\beta^3}{3 + \beta^2} + 3(\beta - \alpha) + 3\sqrt{3} \left[ \arctan \left( \frac{\alpha}{\sqrt{3}} \right) - \arctan \left( \frac{\beta}{\sqrt{3}} \right) \right] \right] + 3 \int_{\alpha}^{\beta} y^2(\nu + y^2)^{-\nu-3/2} dy.$$  

In particular, when $\nu \geq 4$, $E(Y_0^4)$ has the following explicit expressions:

$$E(Y_0^4) = \begin{cases} \frac{4}{c} \left[ \alpha^3(4 + \alpha^2)^{-3/2} - \beta^3(4 + \beta^2)^{-3/2} \right] \\ + \frac{12}{c} \left[ \frac{\alpha}{\sqrt{4 + \alpha^2}} - \frac{\beta}{\sqrt{4 + \beta^2}} + \log \left( \frac{\sqrt{4 + \beta^2} + \beta}{\sqrt{4 + \alpha^2} + \alpha} \right) \right], & \text{if } \nu = 4, \\ G_{\nu}(1) \left[ \frac{\alpha^3}{3 + \alpha^2} - \frac{\beta^3}{3 + \beta^2} + 3(\beta - \alpha) + 3\sqrt{3} \left[ \arctan \left( \frac{\alpha}{\sqrt{3}} \right) - \arctan \left( \frac{\beta}{\sqrt{3}} \right) \right] \right] \\ + \frac{3}{2} G_{\nu}(3) \left[ \alpha(\nu + \alpha^2)^{-\nu-3/2} - \beta(\nu + \beta^2)^{-\nu-3/2} \right] \\ + 3\nu^2 \frac{F_{\nu-4}(\beta_2) - F_{\nu-4}(\alpha_2)}{(\nu - 2)(\nu - 4)c}, & \text{if } \nu > 4, \end{cases}.$$
where
\[ \alpha_2 = \alpha \sqrt{\frac{\nu - 4}{\nu}} \quad \text{and} \quad \beta_2 = \beta \sqrt{\frac{\nu - 4}{\nu}}. \]

The detailed derivations of moments in the above lemmas are put in Appendix A.1.

### 2.2.2 Some limiting results

Let \( Y_0 \sim \text{Tt}(0, 1; (\alpha, \beta)) \). When \( \alpha \to -\infty \) and \( \beta \to \infty \), we have \( Y_0 \sim t(\nu) \). By using the limiting results that
\[ \lim_{\alpha \to -\infty} \alpha^k(\nu + \alpha^2)^{-(\nu - \ell)/2} = 0 \quad \text{and} \quad \lim_{\beta \to \infty} \beta^k(\nu + \beta^2)^{-(\nu - \ell)/2} = 0 \]
for \( \nu > k + \ell \), \( k = 0, 1, 2, 3 \), \( \ell = 1, 3 \) with the technical proof given in Appendix A.2, from Lemmas 1–4, we obtain
\[ E(Y_0) = 0, \quad \text{for } \nu > 1, \]
\[ E(Y_0^2) = \frac{\nu}{\nu - 2}, \quad \text{for } \nu > 2, \]
\[ E(Y_0^3) = 0, \quad \text{for } \nu > 3, \]
\[ E(Y_0^4) = \frac{3\nu^2}{(\nu - 2)(\nu - 4)}, \quad \text{for } \nu > 4. \]

Besides, the corresponding skewness and kurtosis are given by
\[ \gamma = \text{Skewness}(Y_0) = 0 \quad \text{for } \nu > 3 \quad \text{and} \quad \kappa = \text{Kurtosis}(Y_0) = \frac{3(\nu - 2)}{\nu - 4} \quad \text{for } \nu > 4. \]

These results concur with those of the \( t(\nu) \) distribution.

### 2.2.3 Moments of the Tt distribution

For \( k > 0 \), the \( k \)-th non-central moment of \( Y \) in (2.8) can be recursively obtained as
\[ \mu_k = E(Y^k) = \sum_{j=0}^{k} \binom{k}{j} \mu^{k-j} \sigma^j E(Y_0^j). \]

It is clear that the expectation, variance, skewness, and kurtosis are all tractable for the Tt distribution. It is noteworthy that for \( t(\nu) \) distributions, \( \mu_k \) either does not exist or is infinite when \( \nu \leq k \); however, the moments of the doubly Tt distributions do not have such a restriction.
The second-order moment of $Y_0 \sim Tt(0, 1, \nu; [\alpha, \beta])$ exists even for $\nu < 2$ since the function $g_1(y) = (\nu + y^2)^{-\frac{\nu}{2}}$ is bounded within the interval $[\alpha, \beta]$ when $|\alpha|, |\beta| < \infty$, which can be obtained appealing to numerical methods. Similarly, since $g_2(y) = (\nu + y^2)^{-\frac{\nu}{2} - 1}$ is also bounded within the finite interval $[\alpha, \beta]$, the fourth-order moment of $Y_0$ exists even for $\nu < 4$.

Lemmas 1–4 list the first four moments of $Y_0$ for $\nu = 1$; that is, the moments of the truncated standard Cauchy distribution. These results are consistent with those in Corollary 2 of Nadarajah & Kotz (2008), however, their first-order moment is wrong, it should be divided by 2.

The following result previously presented by Massuia et al. (2015) will be used in the implementation of the EM algorithm in Section 3.

**Lemma 5** Let $Y \sim Tt(\mu, \sigma^2, \nu; [a, b])$, $d(\mu, \sigma^2, Y) \equiv (Y - \mu)/\sigma$ and $r > 0$. Then for $k = 0, 1, 2$, we have

$$
E \left\{ \left[ \frac{\nu + 1}{\nu + d^2(\mu, \sigma^2, Y)} \right]^r Y^k \right\} = \frac{(\nu + 1)^r \Gamma \left( \frac{\nu + 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{c \nu^{\nu/2} \Gamma \left( \frac{\nu + 2r}{2} \right)} E(W^k) \left[ F_{\nu+2r} \left( \frac{b - \mu}{\sigma^*} \right) - F_{\nu+2r} \left( \frac{a - \mu}{\sigma^*} \right) \right],
$$

where $c$ is defined by (2.2), $W \sim Tt(\mu, \sigma^{*2}, \nu + 2r; [a, b])$ and $\sigma^{*2} = \nu \sigma^2 / (\nu + 2r)$.

### 2.3 Generation of random variables from the Tt distribution

When both $a$ and $b$ are finite, to generate random variables from $Y \sim Tt(\mu, \sigma^2, \nu; [a, b])$ with density $f_Y(y)$ specified by (2.1), we first select an appropriate grid points $\{y_j\}_{j=1}^m$ that cover the interval $[a, b]$, and then approximate the density $f_Y(y)$ by a discrete distribution at $\{y_j\}_{j=1}^m$ with probabilities $p_j = f_Y(y_j)/\sum_{j=1}^m f_Y(y_{j'})$ for $j = 1, \ldots, m$. In other words, $Y$ approximately follows a finite discrete distribution, denoted by $Y \sim FD_{\text{Discrete}_m}(\{y_j\}, \{p_j\})$.

The built-in R function `sample(y, m, prob=p, replace=T/F)` can be used in the grid method.

However, when $a = -\infty$ or $b = \infty$, the grid method cannot be applied. For example, we consider the case of $b = \infty$. When $a < \mu$, based on (2.3), we continuously generate random samples from $X \sim t(\mu, \sigma^2, \nu)$ until a sample satisfying $X \geq a$ occurs. In the worst situation,
the efficiency of this method is 50%. However, when \( a > \mu \), especially when \( a \gg \mu \), this method is very inefficient.

Motivated by Ho et al. (2012), we could adopt the slice sampling scheme to generate samples from the general Tt distribution. From (2.1), the density function of \( Y \) is

\[
f_Y(y) \propto \left[1 + \frac{(y - \mu)^2}{\nu \sigma^2}\right]^{-(\nu+1)/2} \cdot I(a \leq y \leq b).
\]

We introduce an auxiliary variable \( R \) such that the joint density of \((Y, R)\) is

\[
f_{Y,R}(y, r) \propto I(0 < r < \frac{1}{\nu} \left[1 + \frac{(y - \mu)^2}{\nu \sigma^2}\right]^{-(\nu+1)/2} \cdot I(a \leq y \leq b).
\]

Then the conditional distributions are given by

\[
R|Y = y \sim U\left(0, \frac{1}{\nu} \left[1 + \frac{(y - \mu)^2}{\nu \sigma^2}\right]^{-(\nu+1)/2}\right) \quad \text{and} \quad Y|R = r \sim U\left(\max\{a, \mu - \sigma \kappa_r^{1/2}\}, \min\{b, \mu + \sigma \kappa_r^{1/2}\}\right),
\]

respectively, where \( \kappa_r = \nu\left(r^2 - \frac{2}{\nu+1} - 1\right) \). Thus, the above Gibbs sampling method can be used to draw dependent samples from the Tt distribution, extending the slice sampling for the standard case in Ho et al. (2012) to the general case with \( \mu \) and \( \sigma^2 \).

3. MLEs of parameters in the truncated \( t \) distribution

The SR (2.4) or (2.5) can motivate an EM algorithm to calculate the MLEs of the parameters in the Tt distribution. For the sake of convenience, we focus on the SR (2.5).

3.1 An EM algorithm

3.1.1 Introduction of latent variables

Let \( Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Tt}(\mu, \sigma^2, \nu; [a, b]) \) and \( y_1, \ldots, y_n \) denote their realizations. The observed data are then denoted by \( Y_{\text{obs}} = \{y_i\}_{i=1}^n \). Based on the SR (2.5), for each \( Y_i \) we first independently introduce the latent vectors \( Z_i^* = (Z_{0i}, Z_{1i}, Z_{2i})^\top \sim \text{Multinomial}_3(1;c,c_1,c_2) \), the latent variables \( Y_{L,i} \sim \text{Tt}(\mu, \sigma^2, \nu; (-\infty, a)) \) and \( Y_{R,i} \sim \text{Tt}(\mu, \sigma^2, \nu; (b, \infty)) \) to yield the complete-data r.v.’s

\[
X_i \overset{d}{=} (Y_i^*)^\top Z_i^*, \quad i = 1, \ldots, n,
\]
where \( \{X_i\}_{i=1}^n \) \( \overset{\text{iid}}{\sim} t(\mu, \sigma^2, \nu) \), \( Y_i^* = (Y_i, Y_{i,j}, Y_{R,i})^\top \) and \( \{Z_i^*\}_{i=1}^n, \{Y_i\}_{i=1}^n, \{Y_{L,i}\}_{i=1}^n, \{Y_{R,i}\}_{i=1}^n \) are mutually independent. Second, note that \( t(\mu, \sigma^2, \nu) \) distribution is a Gamma mixture of the normal distribution, we introduce the latent variables of the second layer: \( \{\tau_i\}_{i=1}^n \overset{\text{iid}}{\sim} \text{Gamma}(\nu/2, \nu/2) \) and \( X_i|\tau_i \sim N(\mu, \tau_i^{-1}\sigma^2) \) for \( i = 1, \ldots, n \).

### 3.1.2 M-step

Let \( z_i^* = (z_{0i}, z_{1i}, z_{2i})^\top, y_{L,i}, y_{R,i} \) be the realizations of the latent vectors/variables \( Z_i^*, Y_{L,i}, Y_{R,i} \), then the missing data are denoted by \( Y_{\text{mis}} = \{z_i^*, y_{L,i}, y_{R,i}, \tau_i\}_{i=1}^n \), so that the complete data are \( Y_{\text{com}} = Y_{\text{obs}} \cup Y_{\text{mis}} = \{x_i, \tau_i\}_{i=1}^n \) with \( x_i = z_{0i}y_i + z_{1i}y_{L,i} + z_{2i}y_{R,i} \). Let \( \theta = (\mu, \sigma^2, \nu)^\top \), the complete-data likelihood function of \( \theta \) is given by

\[
L(\theta|Y_{\text{com}}) = \prod_{i=1}^n \text{Gamma}(\tau_i|\nu/2, \nu/2) \cdot N(x_i|\mu, \tau_i^{-1}\sigma^2)
\]

\[
\propto \prod_{i=1}^n \left(\frac{\nu}{2}\right)^{\frac{\nu+1}{2}} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \tau_i^{\nu+1-1} \exp\left(-\frac{\tau_i\nu}{2}\right) \cdot \frac{1}{\sqrt{\sigma^2}} \exp\left[-\frac{\tau_i(x_i - \mu)^2}{2\sigma^2}\right].
\]

The corresponding complete-data log-likelihood function of \( \theta \) is denoted by \( \ell(\theta|Y_{\text{com}}) = \log[L(\theta|Y_{\text{com}})] \), and suppose that the conditional predictive distribution \( f(Y_{\text{mis}}|Y_{\text{obs}}, \theta) \) is available. Let \( \theta^{(t)} = (\mu^{(t)}, \sigma^{2(t)}, \nu^{(t)})^\top \) be the \( t \)-th approximation of the MLEs \( \hat{\theta} \) of \( \theta \). The Q-function in the EM algorithm is defined by

\[
Q(\theta|\theta^{(t)}) = E\left[\ell(\theta|Y_{\text{com}})|Y_{\text{obs}}, \theta^{(t)}\right]
\]

\[
= \int \ell(\theta|Y_{\text{obs}}, Y_{\text{mis}}) \times f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}) \, dY_{\text{mis}}.
\]

In the M-step of the EM algorithm, we need to maximize \( Q \) with respect to \( \theta \) to obtain

\[
\theta^{(t+1)} = \arg\max_{\theta \in \Theta} Q(\theta|\theta^{(t)}).
\]

First, by partially differentiating the Q-function, we have

\[
\frac{\partial Q(\theta|\theta^{(t)})}{\partial \theta} = \int \frac{\partial \ell(\theta|Y_{\text{obs}}, Y_{\text{mis}})}{\partial \theta} \times f(Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}) \, dY_{\text{mis}}.
\]

Note that \( x_i = z_{0i}y_i + z_{1i}y_{L,i} + z_{2i}y_{R,i} \), from the independence among \( \{Z_i^* = (Z_{0i}, Z_{1i}, Z_{2i})^\top, Y_{L,i}, Y_{R,i}\} \) and the conditional distribution of \( X_i \) given \( Y_i = y_i \) in (3.4), we can express the
complete-data MLEs of parameters in terms of \( \{x_i\}_{i=1}^n \). By setting \( \partial \ell(\theta|Y_{\text{obs}}, Y_{\text{mis}})/\partial \theta = 0 \), we immediately obtain the complete-data MLEs of \((\mu, \sigma^2)\) given by

\[
\hat{\mu} = \frac{\sum_{i=1}^n \tau_i x_i}{\sum_{i=1}^n \tau_i}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n \tau_i (x_i - \hat{\mu})^2}{n},
\]

(3.2)

and the complete-data MLE of \( \nu \) being the solution to the following equation

\[
\log \left( \frac{\nu}{2} \right) + 1 - \frac{\Gamma'(\nu/2)}{\Gamma(\nu/2)} + \frac{1}{n} \sum_{i=1}^n (\log \tau_i - \tau_i) = 0.
\]

(3.3)

Since (3.2) and (3.3) only involve functions of \( \tau_i x_i^k \) and \( \log(\tau_i) \) on latent variables, in §3.1.3, these terms should be replaced by their conditional expectations given observed data \( y_i \)'s.

### 3.1.3 E-step

The E-step is to calculate the conditional expectations \( E(\tau_i X_{i}^k | Y_i = y_i, \theta) \) for \( k = 0, 1, 2 \) and \( E(\log(\tau_i) | Y_i = y_i, \theta) \). First, note that the conditional density of \( \tau_i \) given \((X_i, Y_i) = (x_i, y_i)\) and \( \theta \) is proportional to \( \tau_i^{-(\nu+1)/2} \exp\{-\tau_i [\nu + (x_i - \mu)^2/\sigma^2]/2\} \), implying that

\[
\tau_i | (x_i, y_i, \theta) \sim \text{Gamma} \left( \frac{\nu + 1}{2}, \frac{\nu + (x_i - \mu)^2/\sigma^2}{2} \right).
\]

Second, from (2.5) it follows that

\[
X_i | (Y_i = y_i) \overset{d}{=} \begin{cases} 
\text{Degenerate}(y_i), & \text{with probability } c, \\
Y_{L,i}, & \text{with probability } c_1, \\
Y_{R,i}, & \text{with probability } c_2,
\end{cases}
\]

(3.4)

where \( \text{Degenerate}(y) \) denotes the degenerate distribution putting all mass at the point \( y \).

Then, for \( k = 0, 1, 2 \), we have

\[
E \left( \tau_i X_{i}^k | Y_i = y_i, \theta^{(t)} \right) = E_{X_i} \left[ E \left( \tau_i X_{i}^k | X_i, Y_i = y_i, \theta^{(t)} \right) \right]
\]

\[= E \left[ \frac{(\nu^{(t)} + 1)X_{i}^k}{\nu^{(t)} + d^2(\theta^{(t)}_{-3}, X_i)} \bigg| Y_i = y_i, \theta^{(t)} \right]
\]

\[= \left(3.4\right) \frac{c^{(t)}(\nu^{(t)} + 1)y_{i}^k}{\nu^{(t)} + d^2(\theta^{(t)}_{-3}, Y_i)} + c_1^{(t)} E \left[ \frac{(\nu^{(t)} + 1)Y_{L,i}^k}{\nu^{(t)} + d^2(\theta^{(t)}_{-3}, Y_{L,i})} \bigg| \theta^{(t)} \right]
\]

\[+ c_2^{(t)} E \left[ \frac{(\nu^{(t)} + 1)Y_{R,i}^k}{\nu^{(t)} + d^2(\theta^{(t)}_{-3}, Y_{R,i})} \bigg| \theta^{(t)} \right],
\]

13
where $\theta_{-3} = (\mu, \sigma^2)\top$, $d^2(\theta_{-3}, X_i) = [d(\mu, \sigma^2, X_i)]^2 = (X_i - \mu)^2 / \sigma^2$,

$$c^{(t)} = F_\nu \left( \frac{b - \mu^{(t)}}{\sigma^{(t)}} \right) - F_\nu \left( \frac{a - \mu^{(t)}}{\sigma^{(t)}} \right), \quad c_1^{(t)} = F_\nu \left( \frac{a - \mu^{(t)}}{\sigma^{(t)}} \right), \quad c_2^{(t)} = 1 - c^{(t)} - c_1^{(t)}.$$ 

By using Lemma 5, for $r > 0$, we have

$$\mathcal{E}^{r k}_L (\theta) \doteq c_1 E \left\{ \left[ \frac{\nu + 1}{\nu + d^2(\theta_{-3}, Y_{L,i})} \right]^{r} Y_{L,i}^{k} \mid \theta \right\} = \frac{\left( \nu + 1 \right)^r \Gamma \left( \frac{\nu + 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{\nu^r \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{\nu + 2r + 1}{2} \right)} E(W_{Lr}^k) F_{\nu + 2r} \left( \frac{a - \mu}{\sigma^*} \right),$$

$$\mathcal{E}^{r k}_R (\theta) \doteq c_2 E \left\{ \left[ \frac{\nu + 1}{\nu + d^2(\theta_{-3}, Y_{R,i})} \right]^{r} Y_{R,i}^{k} \mid \theta \right\} = \frac{\left( \nu + 1 \right)^r \Gamma \left( \frac{\nu + 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{\nu^r \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{\nu + 2r + 1}{2} \right)} E(W_{Rr}^k) \left[ 1 - F_{\nu + 2r} \left( \frac{b - \mu}{\sigma^*} \right) \right],$$

where $W_{Lr} \sim Tt(\mu, \nu \sigma^2 / (\nu + 2r), \nu + 2r; (-\infty, a))$ and $W_{Rr} \sim Tt(\mu, \nu \sigma^2 / (\nu + 2r), \nu + 2r; (b, \infty))$. Furthermore, we do linear transformations

$$W_{Lr} = \mu + \sqrt{\frac{\nu}{\nu + 2r}} \sigma W_{Lr}^* \quad \text{and} \quad W_{Rr} = \mu + \sqrt{\frac{\nu}{\nu + 2r}} \sigma W_{Rr}^*,$$

where $W_{Lr}^* \sim Tt(0, 1, \nu + 2r; (-\infty, a^*))$, $W_{Rr}^* \sim Tt(0, 1, \nu + 2r; (b^*, \infty))$,

$$a^* = \sqrt{\frac{\nu + 2r}{\nu}} \cdot \frac{a - \mu}{\sigma} \quad \text{and} \quad b^* = \sqrt{\frac{\nu + 2r}{\nu}} \cdot \frac{b - \mu}{\sigma}.$$

By using Lemmas 1–2, we obtain the first two moments of $W_{Lr}^*$ and $W_{Rr}^*$ as

$$E(W_{Lr}^*) = -\frac{\Gamma \left( \frac{\nu + 2r - 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{2F_{\nu + 2r}(a^*) \Gamma \left( \frac{\nu + 2r}{2} \right) \Gamma \left( \frac{1}{2} \right)} (\nu + 2r + a^*)^{-(\nu + 2r - 1)/2},$$

$$E(W_{Lr}^{*2}) = -\frac{\Gamma \left( \frac{\nu + 2r - 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{2F_{\nu + 2r}(a^*) \Gamma \left( \frac{\nu + 2r}{2} \right) \Gamma \left( \frac{1}{2} \right)} a^* (\nu + 2r + a^*)^{-(\nu + 2r - 1)/2} + \frac{(\nu + 2r) F_{\nu + 2r - 2} \left( \frac{a - \mu}{\sigma} \right)}{(\nu + 2r - 2) F_{\nu + 2r}(a^*)},$$

$$E(W_{Rr}^*) = -\frac{\Gamma \left( \frac{\nu + 2r - 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{2 \left[ 1 - F_{\nu + 2r}(b^*) \right] \Gamma \left( \frac{\nu + 2r}{2} \right) \Gamma \left( \frac{1}{2} \right)} (\nu + 2r + b^*)^{-(\nu + 2r - 1)/2},$$

$$E(W_{Rr}^{*2}) = -\frac{\Gamma \left( \frac{\nu + 2r - 1}{2} \right) \Gamma \left( \frac{\nu + 2r}{2} \right)}{2 \left[ 1 - F_{\nu + 2r}(b^*) \right] \Gamma \left( \frac{\nu + 2r}{2} \right) \Gamma \left( \frac{1}{2} \right)} b^* (\nu + 2r + b^*)^{-(\nu + 2r - 1)/2} + \frac{(\nu + 2r) \left[ 1 - F_{\nu + 2r - 2} \left( \frac{b - \mu}{\sigma} \right) \right]}{(\nu + 2r - 2) \left[ 1 - F_{\nu + 2r}(b^*) \right]}.$$
Finally, we have

\[
E \left( \log \tau_i \Bigg| Y_i = y_i, \theta^{(t)} \right) \\
= E_{X_i|Y_i} \left[ E \left( \log \tau_i \Bigg| X_i, Y_i = y_i, \theta^{(t)} \right) \right] \\
= \frac{\Gamma'\left(\nu^{(t)}+1\right)}{\Gamma\left(\nu^{(t)}+1\right)} + \log 2 - E_{X_i|Y_i} \left[ \log \left( \nu^{(t)} + d^2(\theta^{(t)}_{-3}, X_i) \right) \Bigg| Y_i = y_i, \theta^{(t)} \right].
\]

### 3.1.4 The EM iteration

The EM iterations on \((\mu, \sigma^2)\) are given by

\[
\mu^{(t+1)} = \frac{(1/n) \sum_{i=1}^{n} \mathcal{E}_i^{11}(\theta^{(t)}) + \mathcal{E}_i^{11}(\theta^{(t)}) + \mathcal{E}_i^{11}(\theta^{(t)})}{(1/n) \sum_{i=1}^{n} \mathcal{E}_i^{10}(\theta^{(t)}) + \mathcal{E}_i^{10}(\theta^{(t)}) + \mathcal{E}_i^{10}(\theta^{(t)})},
\]

\[
\sigma^{2(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathcal{E}_i^{12}(\theta^{(t)}) - 2\mu^{(t+1)} \mathcal{E}_i^{11}(\theta^{(t)}) + \left[ \mu^{(t+1)} \right]^2 \mathcal{E}_i^{10}(\theta^{(t)}) \right\} + \mathcal{E}_i^{12}(\theta^{(t)}) + \mathcal{E}_i^{12}(\theta^{(t)}) - 2\mu^{(t+1)} \left[ \mathcal{E}_i^{11}(\theta^{(t)}) + \mathcal{E}_i^{11}(\theta^{(t)}) \right] + \left[ \mu^{(t+1)} \right]^2 \left[ \mathcal{E}_i^{10}(\theta^{(t)}) + \mathcal{E}_i^{10}(\theta^{(t)}) \right],
\]

and \(\nu^{(t+1)}\) is the solution to the following equation

\[
\log(\nu) + 1 - \frac{\Gamma'\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} + \frac{\Gamma'\left(\nu^{(t)}+1\right)}{\Gamma\left(\nu^{(t)}+1\right)} - \frac{1}{n} \sum_{i=1}^{n} b_i(\theta^{(t)}) - \frac{1}{n} \sum_{i=1}^{n} \mathcal{E}_i^{10}(\theta^{(t)}) - \mathcal{E}_i^{10}(\theta^{(t)}) - \mathcal{E}_i^{10}(\theta^{(t)}) = 0, \quad (3.5)
\]

where

\[
\mathcal{E}_i^{rk}(\theta^{(t)}) = c^{(t)} \left[ \frac{\nu^{(t)} + 1}{\nu^{(t)} + d^2(\theta^{(t)}_{-3}, y_i)} \right]^r (y_i \mathcal{Y}_i),
\]

\[
b_i(\theta^{(t)}) = E_{X_i|Y_i} \left\{ \log \left[ \nu^{(t)} + d^2(\theta^{(t)}_{-3}, X_i) \right] \right\} Y_i = y_i, \theta^{(t)} \}
\]

\[
= c^{(t)} \log \left[ \nu^{(t)} + d^2(\theta^{(t)}_{-3}, y_i) \right] + \left( \int_{a}^{b} + \int_{a}^{\infty} \right) \log \left[ \nu^{(t)} + d^2(\theta^{(t)}_{-3}, x_i) \right] t(x_i|\mu^{(t)}, \sigma^{2(t)}, \nu^{(t)}) \, dx_i,
\]

for \(k = 0, 1, 2\) and \(i = 1, \ldots, n\). The integrals could be implemented by using the “integrate” function in R. It should be mentioned that in practical analysis we usually take the sample mean and sample variance as the initial values \((\mu^{(0)}, \sigma^{2(0)})\) for parameters \((\mu, \sigma^2)\), while the initial value \(\nu^{(0)}\) for \(\nu\) is chosen as an arbitrary constant, say, two for simplicity.
Besides, the built-in \texttt{R} function “\texttt{optim}” can provide the Hessian matrix. The square root of the diagonal elements of the inverse negative Hessian matrix can be used as the estimated standard deviations of the MLEs of the parameters, and then the asymptotic Wald CIs of the parameters are obtained.

### 3.2 The Tt regression model

When a set of covariates is available, the simplest Tt regression model can be formulated as

\[ Y_i \overset{\text{ind}}{\sim} \text{Tt}(w_i^\top \beta, \sigma^2, \nu; [a, b]), \quad i = 1, \ldots, n, \tag{3.6} \]

where \( Y_i \) is the response variable for subject \( i \), \( w_i = (1, w_{i1}, \ldots, w_{ip})^\top \) is the covariate vector of interest, \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^\top \) is the corresponding vector of the regression coefficients, and \( \sigma^2 \) is assumed to be the common dispersion parameter across all subjects and is not related to the covariates. Let \( y_i \) be the realization of \( Y_i \) and denote the observed data by

\[ Y_{\text{obs}} = \{y_i\}_{i=1}^n. \]

Let \( \theta = (\beta^\top, \sigma^2, \nu)^\top \), the observed-data log-likelihood of \( \theta \) is then given by

\[ \ell(\theta|Y_{\text{obs}}) = \sum_{i=1}^n \log \left[ \text{Tt}(y_i|w_i^\top \beta, \sigma^2, \nu; [a, b]) \right]. \tag{3.7} \]

Based on the SR (2.5), we independently introduce the latent vectors \( Z_i = (Z_i, Z_{1i}, Z_{2i})^\top \sim \text{Multinomial}_3(1; c_i, c_{1i}, c_{2i}) \), where

\[ c_i = F_\nu((b - w_i^\top \beta)/\sigma) - F_\nu((a - w_i^\top \beta)/\sigma), \quad c_{1i} = F_\nu((a - w_i^\top \beta)/\sigma) \]

and \( c_{2i} = 1 - c_i - c_{1i} \), and the latent variables

\[ Y_{L,i} \overset{\text{ind}}{\sim} \text{Tt}(w_i^\top \beta, \sigma^2, \nu; (-\infty, a)) \quad \text{and} \quad Y_{R,i} \overset{\text{ind}}{\sim} \text{Tt}(w_i^\top \beta, \sigma^2, \nu; (b, \infty)). \]

The augmented data are \( \{y_i, z_i, y_{L,i}, y_{R,i}\}_{i=1}^n \), where unobserved \( z_i = (z_i, z_{1i}, z_{2i})^\top \), \( y_{L,i}, y_{R,i} \) are the realizations of the corresponding latent vectors/variables. We denote the complete data by \( Y_{\text{com}} = \{x_i\}_{i=1}^n \) with \( x_i = z_i y_i + z_{1i} y_{L,i} + z_{2i} y_{R,i} \).

Then the complete-data MLEs of \((\beta, \sigma^2)\) are given by

\[
\begin{align*}
\hat{\beta} &= \left[ \sum_{i=1}^n \frac{w_i w_i^\top}{\nu + (x_i - w_i^\top \beta)^2/\sigma^2} \right]^{-1} \left[ \sum_{i=1}^n \frac{w_i x_i}{\nu + (x_i - w_i^\top \beta)^2/\sigma^2} \right], \\
\hat{\sigma}^2 &= \frac{\nu + 1}{n} \sum_{i=1}^n \frac{(x_i - w_i^\top \beta)^2}{\nu + (x_i - w_i^\top \beta)^2/\sigma^2}. 
\end{align*}
\]
The MLE of $\nu$ is the solution to the equation
\[
\frac{\Gamma'(\nu + \frac{1}{2})}{\Gamma(\nu + \frac{1}{2})} - \frac{\Gamma'(\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})} + 1 + \log \nu - \frac{1}{n} \sum_{i=1}^{n} \left\{ \log \left( \nu + \frac{(x_i - \mathbf{w}_i \beta)^2}{\sigma^2} \right) + \frac{\nu + 1}{\nu + (x_i - \mathbf{w}_i \beta)^2/\sigma^2} \right\} = 0.
\]

The E-step is to replace the latent variables by their conditional expectations, which are quite similar with the case without covariates. The initial values for $(\beta, \sigma^2)$ are often chosen as the OLS estimates given by
\[
\beta^{(0)} = (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{y} \quad \text{and} \quad \sigma^{2(0)} = \frac{(\mathbf{y} - \mathbf{W} \beta^{(0)})^\top (\mathbf{y} - \mathbf{W} \beta^{(0)})}{n},
\]
where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)^\top$ is the $n \times (p + 1)$ matrix of covariates and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top$ is the response vector. The initial value for $\nu$ is the same as that in the Tt distribution.

For both Section 3.1 and Section 3.2, the iterative process is stopped if the absolute value of the difference between two successive evaluations of the observed-data log-likelihood function $\ell(\theta | Y_{\text{obs}})$ is smaller than some predetermined value; that is, $|\ell(\theta^{(t)} | Y_{\text{obs}}) - \ell(\theta^{(t-1)} | Y_{\text{obs}})| \leq \lambda$, where $\lambda$ is said to be the precision.

### 3.3 Bootstrap confidence intervals

The asymptotic Wald confidence intervals for parameters are reliable only for large sample sizes and may become useless if either a boundary for some parameter is beyond its nature limit, say the lower bound for $\sigma^2$ is less than 0, or the lower bound for $\nu$ is less than 0. For small to moderate sample size, the bootstrap method is a useful tool to find a bootstrap CI for an arbitrary function of $\theta = (\mu, \sigma^2, \nu)^\top$, say, $\vartheta = h(\theta)$. With the MLEs $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\nu})^\top$ obtained by using the proposed EM algorithm, we can generate $\{Y_i^{(g)}\}_{i=1}^{n} \overset{iid}{\sim} \text{Tt}(\hat{\mu}, \hat{\sigma}^2, \hat{\nu}; [a, b])$. With the bootstrap sample $Y_{\text{obs}}^{*} = \{y_i^{(g)}\}_{i=1}^{n}$, a bootstrap replication of $\hat{\theta}$ is similarly obtained, denoted by $\hat{\theta}_g$, then we have $\hat{\vartheta}_g = h(\hat{\theta}_g)$. Independently repeat this process for $g$ from 1 to $G$ to obtain $G$ bootstrap replications $\{\hat{\vartheta}_g\}_{g=1}^{G}$. Consequently, the standard error, $\text{se}(\hat{\vartheta})$, of $\hat{\vartheta}$ can be estimated by the sample standard deviation of the $G$ replications, i.e.,
\[
\hat{\text{se}}(\hat{\vartheta}) = \left\{ \frac{1}{G-1} \sum_{g=1}^{G} \left[ \hat{\vartheta}_g - (\hat{\vartheta}_1^{*} + \cdots + \hat{\vartheta}_G^{*})/G \right]^2 \right\}^{1/2}.
\]
If \( \{ \hat{\theta}_g \}_{g=1}^G \) is approximately normally distributed, the \((1 - \alpha)100\%\) simple bootstrap CI (SCI) for \( \theta \) is

\[
[\hat{\theta} - z_{\alpha/2} \cdot \text{se}(\hat{\theta}), \hat{\theta} + z_{\alpha/2} \cdot \text{se}(\hat{\theta})].
\] (3.9)

Alternatively, if \( \{ \hat{\theta}_g \}_{g=1}^G \) is non-normally distributed, the \((1 - \alpha)100\%\) bootstrap percentile CI (PCI) for \( \theta \) is defined as

\[
[\hat{\theta}_L, \hat{\theta}_U],
\] (3.10)

where \( \hat{\theta}_L \) and \( \hat{\theta}_U \) are the \( 100(\alpha/2) \) and \( 100(1 - \alpha/2) \) percentiles of \( \{ \hat{\theta}_g \}_{g=1}^G \), respectively.

### 4. Student-\( t \) interval censored regression model

Another application of the truncated \( t \) distribution is to combine it with the interval–censored regression model. The original Student-\( t \) linear regression model is of the form

\[
Y_i = x_i^\top \beta + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} t(0, \sigma^2, \nu), \quad i = 1, \ldots, n,
\]

where \( Y_i \) is the response variable for subject \( i \), \( x_i = (1, x_{i1}, \ldots, x_{ip})^\top \) is the covariate vector, \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^\top \) is the vector of regression coefficients, and the errors are independently and identically Student-\( t \) distributed. Then, we have \( Y_i \overset{\text{ind}}{\sim} t(x_i^\top \beta, \sigma^2, \nu) \) for \( i = 1, \ldots, n \).

When incorporating the censoring mechanism, for the \( i \)-th item, there is a censoring interval \([\tau_i, \kappa_i]\), where \( \tau_i \) and \( \kappa_i \) denote the left and right end-points, respectively. The value of \( Y_i \) is observable if \( Y_i \notin [\tau_i, \kappa_i] \), otherwise it is not observable. Define \( \delta_i = I(Y_i \notin [\tau_i, \kappa_i]) \), then \( \delta_i = 1 \) implies that the observation is not censored such that the value of \( Y_i \) is observed; while \( \delta_i = 0 \) indicates that only the censoring interval \([\tau_i, \kappa_i]\) is observed. Then the observations are of the form

\[
(Y_{\text{obs},i}, \delta_i) = \begin{cases} 
(y_i, 1), & \text{if } Y_i \notin [\tau_i, \kappa_i], \\
([\tau_i, \kappa_i], 0), & \text{if } Y_i \in [\tau_i, \kappa_i],
\end{cases}
\] (4.1)

for \( i = 1, \ldots, n \). We call this the \( t \) interval–censored regression (\( t \)-ICR) model. When \( \kappa_i = \infty \), the corresponding observation is right-censored; when \( \tau_i = -\infty \), it is left-censored.

According to the \( t \)-ICR model specified by (4.1), suppose that there are \( m \) censored values, \( n - m \) uncensored observations. Without loss of generality, the observed data are denoted
by \(Y_{\text{obs}} = \{[\tau_1, \kappa_1], \ldots, [\tau_m, \kappa_m], y_{m+1}, \ldots, y_n\} = \{Y_{\text{obs}}\}_{i=1}^n\). The log-likelihood function for \(\theta = (\beta^T, \sigma^2, \nu)^T\) is

\[
\ell(\theta|Y_{\text{obs}}) = \sum_{i=1}^m \log \left[ F_\nu\left(\frac{k_i - x_i^T\beta}{\sigma}\right) - F_\nu\left(\frac{\tau_i - x_i^T\beta}{\sigma}\right) \right] + \sum_{i=m+1}^n \log \left[ t(y_i|x_i^T\beta, \sigma^2, \nu) \right].
\]

We introduce latent variables \(Y_i \sim t(x_i^T\beta, \sigma^2, \nu)\) and \(y_i\) denotes its realization for \(i = 1, \ldots, m\). Thus, the complete data are \(Y_{\text{com}} = \{y_i\}_{i=1}^n\), where \(\{y_i\}_{i=m+1}^n\) are in fact observed.

Now, the complete-data log-likelihood is written as

\[
\ell(\theta|Y_{\text{com}}) = \sum_{i=1}^n \log \left[ t(y_i|x_i^T\beta, \sigma^2, \nu) \right]
\]

\[
= n \log \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)} - \frac{n}{2} \log(\pi \nu) - \frac{n}{2} \log \sigma^2 - \frac{\nu + 1}{2} \sum_{i=1}^n \log \left[ 1 + \frac{(y_i - x_i^T\beta)^2}{\nu \sigma^2} \right].
\]

Then the complete-data MLEs are given by

\[
\begin{align*}
\hat{\beta} & = \left[ \sum_{i=1}^n \frac{x_i x_i^T}{\nu + (y_i - x_i^T \beta)^2/\sigma^2} \right]^{-1} \left[ \sum_{i=1}^n \frac{x_i y_i}{\nu + (y_i - x_i^T \beta)^2/\sigma^2} \right], \\
\hat{\sigma}^2 & = \frac{\nu + 1}{n} \sum_{i=1}^n \frac{(y_i - x_i^T \beta)^2}{\nu + (y_i - x_i^T \beta)^2/\sigma^2}.
\end{align*}
\]

The MLE of \(\nu\) is the solution to the equation

\[
\frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)} - \frac{\Gamma(\nu/2)}{\Gamma(\nu/2)} + 1 + \log \nu - \frac{1}{n} \sum_{i=1}^n \log \left[ \nu + \frac{(y_i - x_i^T \beta)^2}{\sigma^2} \right] - \frac{1}{n} \sum_{i=1}^n \frac{\nu + 1}{\nu + (y_i - x_i^T \beta)^2/\sigma^2} = 0.
\]

Note that \(y_1, \ldots, y_m\) are unobserved values that should be replaced by their conditional expectations given the observed data. For the \(i\)-th censored observation, we have

\[
Y_i|Y_{\text{obs}} \overset{d}{=} Y_i|(Y_i \in [\tau_i, \kappa_i]) \sim T_t(x_i^T \beta, \sigma^2, \nu; [\tau_i, \kappa_i]),
\]

indicating that the conditional distributions of the latent variables are all \(T_t\) distributions.

Then the following expectations

\[
E_{ki}(\theta^{(t)}) = E \left[ \frac{Y_i}{\nu^{(t)} + d^2(x_i^T \beta^{(t)}, \sigma^2, y_i)} | Y_{\text{obs}}, \theta^{(t)} \right], \quad k = 0, 1, 2 \quad \text{and} \quad i = 1, \ldots, m,
\]

19
can be obtained by using Lemma 5. For expression simplicity, we adopt the same notation for uncensored observations as
\[
\mathcal{E}_{ki}(\theta^{(t)}) = \frac{y^k_i}{\nu^{(t)} + d^2(x_i^\top \beta^{(t)}, \sigma^{2(t)}, y_i)}; \quad k = 0, 1, 2 \quad \text{and} \quad i = m + 1, \ldots, n.
\]

Next we denote
\[
b_i(\theta^{(t)}) = E\left\{\log \left[\nu^{(t)} + d^2(x_i^\top \beta^{(t)}, \sigma^{2(t)}, Y_i)\right] \left| Y_{\text{obs}}, \theta^{(t)}\right\} \right.
\]
\[= \int_{r_i}^{\nu_i} \log \left[\nu^{(t)} + d^2(x_i^\top \beta^{(t)}, \sigma^{2(t)}, y_i)\right] \frac{t(y_i|x_i^\top \beta^{(t)}, \sigma^{2(t)}, \nu^{(t)})}{F_{\nu^{(t)}}(\nu_i^{(t)}) - F_{\nu^{(t)}}(r_i^{(t)})} \, dy_i,
\]
where
\[
\kappa_i^{(t)} = \frac{\kappa_i}{\sigma^{(t)}} \quad \text{and} \quad \tau_i^{(t)} = \frac{\tau_i - x_i^\top \beta^{(t)}}{\sigma^{(t)}},
\]
for \(i = 1, \ldots, m\); while
\[
b_i(\theta^{(t)}) = \log \left[\nu^{(t)} + d^2(x_i^\top \beta^{(t)}, \sigma^{2(t)}, y_i)\right], \quad i = m + 1, \ldots, n.
\]

The EM algorithm to obtain the estimates of the parameters is summarized as
\[
\begin{align*}
\beta^{(t+1)} &= \left[\sum_{i=1}^{n} \mathcal{E}_{0i}(\theta^{(t)}) x_i x_i^\top\right]^{-1} \left[\sum_{i=1}^{n} \mathcal{E}_{1i}(\theta^{(t)}) x_i\right], \\
\sigma^{2(t+1)} &= \frac{\nu + 1}{n} \sum_{i=1}^{n} \left[\mathcal{E}_{2i}(\theta^{(t)}) - 2\mathcal{E}_{1i}(\theta^{(t)}) x_i^\top \beta^{(t+1)} + \mathcal{E}_{0i}(\theta^{(t)})(x_i^\top \beta^{(t+1)})^2\right],
\end{align*}
\]
and update \(\nu^{(t+1)}\) as the solution to the following equation with respect to \(\nu\)
\[
\frac{\Gamma'(\frac{\nu+1}{2})}{\Gamma'(\frac{\nu}{2})} - \frac{\Gamma'(\frac{\nu}{2})}{\Gamma'(\frac{\nu+1}{2})} + 1 + \log \nu - \frac{1}{n} \sum_{i=1}^{n} b_i(\theta^{(t)}) - \frac{1}{n} \sum_{i=1}^{n} (\nu + 1) \mathcal{E}_{0i}(\theta^{(t+1)}) = 0.
\]

Although for the \(t\)-ICR model, we cannot have the OLS estimates for parameters since some responses are censored. Instead, we could construct a pseudo OLS estimate for \(\beta\) by setting
\[
y^* = (y_1^*, \ldots, y_m^*, y_{m+1}, \ldots, y_n)^\top
\]
where
\[
y_i^* = \begin{cases} 
\tau_i, & \text{if } \kappa_i = \infty, \\
\kappa_i, & \text{if } \tau_i = -\infty, \\
(\tau_i + \kappa_i)/2, & \text{otherwise},
\end{cases}
\]
for \(i = 1, \ldots, m\) and replacing \(y\) in (3.8) by \(y^*\) to assign initial values for \((\beta, \sigma^2)\). The way to choose the initial value for \(\nu\) still remains the same. The stopping rule of the above process is the same as that stated at the end of Section 3.2.
5. Numerical experiments

5.1 Experiment 1

As we know that the $t$ distribution has heavier tails than the normal distribution, a similar characteristic also exists on the truncated $t$ distribution when comparing with the truncated normal (Tn) distribution. In this simulation study, we investigate the influence on parameter estimates when the inappropriate Tn distribution is adopted.

We consider the case with covariates, i.e., the Tt regression model given by (3.6). The experiment setting is as follows: The truncated points are chosen to be $[a, b] = [0, 5]$, the true parameters are set as $\beta^\top = (\beta_0, \beta_1) = (2, 1)$, $\sigma^2 = 1$ and $\nu = 5$. The total sample size $n$ is considered as 100(50)300, where $a(s)b$ means from $a$ to $b$ with step size being $s$. The design matrix of the covariates $W$ is constructed in the following way:

$$ W = [\mathbf{1}_n, \mathbf{1}_{[n/10]} \otimes \mathbf{t}], \quad \text{with } \mathbf{t}^\top = (0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0), $$

where the operator “$\otimes$” denotes the Kronecker product. The $i$-th row of $W$, $\mathbf{w}_i^\top$, is the designed covariates for subject $i$. With these parameter configurations, the data sets are generated as

$$ Y_i \stackrel{\text{ind}}{\sim} \text{Tt}(\mathbf{w}_i^\top \beta, \sigma^2, \nu; [a, b]), \quad i = 1, \ldots, n. $$

The corresponding observed values are denoted by $\{y_i\}_{i=1}^n$, we fit the generated data by both truncated $t$ regression (TtR) and truncated normal regression (TnR) models. Then under each model, the bootstrap method is adopted to obtain a replication of the estimators and two types of bootstrap CIs, SCI and PCI, specified by (3.9) and (3.10), respectively, based on 500 bootstrap samples. Finally, the above process is repeated 500 times to compute the standard deviation (Std) and the mean squared error (MSE) of the estimators, and the coverage probability (CP) of two types of CIs. Meanwhile, the mean of standard deviations obtained by the Hessian matrix, known as the empirical standard errors, is also provided for comparisons. The simulation results are summarized in Table 1.

The results reveal that when we fit truncated $t$ samples by TtR model or inappropriate TnR model, the estimation accuracies from the TnR model are much worse than those from
Table 1. Simulation results based on simulated Tt samples (Doubly truncated case)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Model</th>
<th>Para.</th>
<th>Mean</th>
<th>SE\textsubscript{emp}</th>
<th>Std\textsuperscript{B}</th>
<th>MSE</th>
<th>CP\textsuperscript{†}</th>
<th>CP\textsuperscript{‡}</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>TtR</td>
<td>β\textsubscript{0}</td>
<td>1.996253</td>
<td>0.263958</td>
<td>0.265512</td>
<td>0.070511</td>
<td>0.946</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>1.008362</td>
<td>0.229684</td>
<td>0.236870</td>
<td>0.056177</td>
<td>0.962</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.019119</td>
<td>0.249758</td>
<td>0.278737</td>
<td>0.078060</td>
<td>0.922</td>
<td>0.928</td>
</tr>
<tr>
<td></td>
<td>TnR</td>
<td>β\textsubscript{0}</td>
<td>2.010097</td>
<td>0.268558</td>
<td>0.265750</td>
<td>0.070725</td>
<td>0.946</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>0.981775</td>
<td>0.240180</td>
<td>0.242463</td>
<td>0.059121</td>
<td>0.946</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.217234</td>
<td>0.259367</td>
<td>0.316790</td>
<td>0.147546</td>
<td>0.940</td>
<td>0.872</td>
</tr>
<tr>
<td>150</td>
<td>TtR</td>
<td>β\textsubscript{0}</td>
<td>1.991909</td>
<td>0.218494</td>
<td>0.215660</td>
<td>0.046575</td>
<td>0.958</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>1.009151</td>
<td>0.189527</td>
<td>0.186367</td>
<td>0.034816</td>
<td>0.946</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.021569</td>
<td>0.208646</td>
<td>0.212757</td>
<td>0.045731</td>
<td>0.940</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>TnR</td>
<td>β\textsubscript{0}</td>
<td>2.003273</td>
<td>0.220371</td>
<td>0.217863</td>
<td>0.047471</td>
<td>0.952</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>0.984381</td>
<td>0.196738</td>
<td>0.191823</td>
<td>0.037040</td>
<td>0.940</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.226411</td>
<td>0.213351</td>
<td>0.236930</td>
<td>0.107398</td>
<td>0.946</td>
<td>0.792</td>
</tr>
<tr>
<td>200</td>
<td>TtR</td>
<td>β\textsubscript{0}</td>
<td>1.981297</td>
<td>0.187909</td>
<td>0.184490</td>
<td>0.034387</td>
<td>0.950</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>1.016170</td>
<td>0.162303</td>
<td>0.155919</td>
<td>0.024572</td>
<td>0.972</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.001678</td>
<td>0.177457</td>
<td>0.175717</td>
<td>0.030879</td>
<td>0.946</td>
<td>0.792</td>
</tr>
<tr>
<td></td>
<td>TnR</td>
<td>β\textsubscript{0}</td>
<td>2.001821</td>
<td>0.166750</td>
<td>0.171067</td>
<td>0.029267</td>
<td>0.948</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>0.999890</td>
<td>0.144068</td>
<td>0.150267</td>
<td>0.022610</td>
<td>0.942</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.004731</td>
<td>0.156253</td>
<td>0.164657</td>
<td>0.027134</td>
<td>0.958</td>
<td>0.952</td>
</tr>
<tr>
<td>250</td>
<td>TtR</td>
<td>β\textsubscript{0}</td>
<td>2.007388</td>
<td>0.168974</td>
<td>0.172018</td>
<td>0.029698</td>
<td>0.954</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>0.982978</td>
<td>0.150440</td>
<td>0.155308</td>
<td>0.024410</td>
<td>0.932</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.214332</td>
<td>0.160510</td>
<td>0.179244</td>
<td>0.077981</td>
<td>0.832</td>
<td>0.738</td>
</tr>
<tr>
<td></td>
<td>TnR</td>
<td>β\textsubscript{0}</td>
<td>2.006465</td>
<td>0.153268</td>
<td>0.146474</td>
<td>0.021496</td>
<td>0.957</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>0.999890</td>
<td>0.132222</td>
<td>0.124782</td>
<td>0.015571</td>
<td>0.966</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.003731</td>
<td>0.144595</td>
<td>0.151109</td>
<td>0.022848</td>
<td>0.949</td>
<td>0.946</td>
</tr>
<tr>
<td>300</td>
<td>TtR</td>
<td>β\textsubscript{0}</td>
<td>2.017566</td>
<td>0.154899</td>
<td>0.146717</td>
<td>0.021834</td>
<td>0.959</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β\textsubscript{1}</td>
<td>0.975692</td>
<td>0.137804</td>
<td>0.129609</td>
<td>0.017389</td>
<td>0.959</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>σ\textsuperscript{2}</td>
<td>1.212526</td>
<td>0.148088</td>
<td>0.165044</td>
<td>0.072407</td>
<td>0.806</td>
<td>0.671</td>
</tr>
</tbody>
</table>

SE\textsubscript{emp}: empirical standard error obtained by the Hessian matrix; Std\textsuperscript{B}: standard deviation of the bootstrap samples; MSE: the sum of the variance and the squared bias of the estimators; CP\textsuperscript{†}: the coverage proportion of SCI; CP\textsuperscript{‡}: the coverage proportion of PCI.
the TtR model, especially the estimates on scale parameter $\sigma^2$ with estimated values around 1.2 which are far from its true value. The standard deviations of the estimators given by TnR models from both the empirical and the bootstrap methods are always larger than those given by TtR models, so are the MSEs in the majority of cases. The empirical standard errors and the standard deviations of the estimates are close to each other, and both of them as well as as the MSEs decrease with the sample size. The coverage probabilities of two bootstrap CIs under two assumptions for parameters $\beta_0$ and $\beta_1$ are quite satisfactory in different values of sample size. For the coverage probability of interval estimators for parameter $\sigma^2$, when the wrong assumption of Tn distribution is taken, the SCI performs well when sample size is less than 200, while it dramatically drops with the enlarged sample size, and the PCI has quite a bad behavior in all situations. The two interval estimators for $\sigma^2$ from the Tt distribution always give stable coverage probabilities around 0.95. Figure 1 shows the distinct coverage probabilities of two types of interval estimators for parameter $\sigma^2$ under the framework of Tt and Tn distributions.

5.2 Experiment 2

Subsequently, we investigate the influences on parameter inferences when data are contaminated. For the purpose of comparing the performances of the TtR and TnR models, we choose the percentage of change on estimation of the parameters as the comparison index, which is defined as

$$\eta = \left| \frac{\hat{\gamma}' - \hat{\gamma}}{\hat{\gamma}} \right| \times 100\%,$$

the relative change on estimation of the parameter $\gamma$, where $\hat{\gamma}$ is the MLE of $\gamma$ based on the original data set and $\hat{\gamma}'$ denotes the MLE of $\gamma$ after some observations are contaminated.

To evaluate the effects on parameter estimations in both doubly truncated and one-sided truncated $t$ regression models, we consider the following parameter configurations:

1. Doubly truncated TtR: $[a, b] = [0, 5], \beta^T = (\beta_0, \beta_1) = (2, 1), \sigma^2 = 1$ and $\nu = 5$.
2. One-sided truncated TtR: $[a, b] = (-\infty, 2], \beta^T = (\beta_0, \beta_1) = (-1, 3), \sigma^2 = 2$ and $\nu = 6$.
3. Doubly truncated TnR: $[a, b] = [0, 5], \beta^T = (\beta_0, \beta_1) = (2, 1)$ and $\sigma^2 = 1$. 

23
Figure 1: Coverage probabilities of interval estimators for $\sigma^2$ from the TtR (solid line) and TnR (dashed line) models. Left plot: the simple bootstrap CI of $\sigma^2$. Right plot: the bootstrap percentile CI of $\sigma^2$.

(4) One-sided truncated TnR: $[a, b] = (-\infty, 2]$, $\beta^T = (\beta_0, \beta_1) = (-1, 3)$ and $\sigma^2 = 2$.

The covariates $w_i$’s are constructed in the same way as previously. The sample size is fixed as $n = 200$. For each case, we generate samples from the corresponding TtR/TnR model, or equivalently $Y_i \sim \text{ind} \: Tt(w_i^T \beta, \sigma^2, \nu; [a, b]) / Y_i \sim \text{ind} \: \text{TN}(w_i^T \beta, \sigma^2; [a, b])$ for $i = 1, \ldots, n$, and fit the observed data $\{y_i\}_{i=1}^n$ by both TtR and TnR models, recording the MLEs of the parameters $(\beta_0, \beta_1, \sigma^2)$ from the two models respectively with the precision of $\delta = 10^{-6}$. Then we assume the minimum of the data was contaminant by adding $\Delta$ units, where $\Delta = 0.5(0.5)5$, and assess them again using the two models. Repeat the above process 1000 times, the percentage of change on estimation of the parameters is taken as the average value of the relative change

$1$ The pdf of $Y \sim \text{TN}(\mu, \sigma^2; [a, b])$ is $\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] / [\Phi(b_0) - \Phi(a_0)]$ with $b_0 = \frac{b-\mu}{\sigma}$ and $a_0 = \frac{a-\mu}{\sigma}$. 24
for each parameter throughout the whole process. The results are shown in Tables 2–5.

**Table 2.** Percentage of change on estimation of the parameters (Doubly truncated TtR)

<table>
<thead>
<tr>
<th>Δ</th>
<th>β₀</th>
<th>β₁</th>
<th>σ²</th>
<th>β₀</th>
<th>β₁</th>
<th>σ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0013</td>
<td>0.0032</td>
<td>0.0103</td>
<td>0.0056</td>
<td>0.0113</td>
<td>0.0206</td>
</tr>
<tr>
<td>1</td>
<td>0.0036</td>
<td>0.0079</td>
<td>0.0220</td>
<td>0.0106</td>
<td>0.0207</td>
<td>0.0363</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0073</td>
<td>0.0142</td>
<td>0.0337</td>
<td>0.0152</td>
<td>0.0285</td>
<td>0.0473</td>
</tr>
<tr>
<td>2</td>
<td>0.0121</td>
<td>0.0212</td>
<td>0.0427</td>
<td>0.0193</td>
<td>0.0348</td>
<td>0.0538</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0170</td>
<td>0.0273</td>
<td>0.0461</td>
<td>0.0231</td>
<td>0.0398</td>
<td>0.0558</td>
</tr>
<tr>
<td>3</td>
<td>0.0210</td>
<td>0.0310</td>
<td>0.0426</td>
<td>0.0265</td>
<td>0.0437</td>
<td>0.0534</td>
</tr>
<tr>
<td>3.5</td>
<td>0.0235</td>
<td>0.0329</td>
<td>0.0334</td>
<td>0.0297</td>
<td>0.0465</td>
<td>0.0464</td>
</tr>
<tr>
<td>4</td>
<td>0.0247</td>
<td>0.0341</td>
<td>0.0219</td>
<td>0.0324</td>
<td>0.0483</td>
<td>0.0354</td>
</tr>
<tr>
<td>4.5</td>
<td>0.0249</td>
<td>0.0345</td>
<td>0.0124</td>
<td>0.0350</td>
<td>0.0496</td>
<td>0.0247</td>
</tr>
<tr>
<td>5</td>
<td>0.0246</td>
<td>0.0341</td>
<td>0.0092</td>
<td>0.0372</td>
<td>0.0504</td>
<td>0.0254</td>
</tr>
</tbody>
</table>

**Table 3.** Percentage of change on estimation of the parameters (One-sided truncated TtR)

<table>
<thead>
<tr>
<th>Δ</th>
<th>β₀</th>
<th>β₁</th>
<th>σ²</th>
<th>β₀</th>
<th>β₁</th>
<th>σ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0102</td>
<td>0.0092</td>
<td>0.0083</td>
<td>0.0931</td>
<td>0.0285</td>
<td>0.0344</td>
</tr>
<tr>
<td>1</td>
<td>0.0144</td>
<td>0.0136</td>
<td>0.0156</td>
<td>0.1717</td>
<td>0.0513</td>
<td>0.0627</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0189</td>
<td>0.0180</td>
<td>0.0233</td>
<td>0.2449</td>
<td>0.0701</td>
<td>0.0863</td>
</tr>
<tr>
<td>2</td>
<td>0.0253</td>
<td>0.0246</td>
<td>0.0315</td>
<td>0.3108</td>
<td>0.0878</td>
<td>0.1065</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0313</td>
<td>0.0301</td>
<td>0.0396</td>
<td>0.3741</td>
<td>0.1010</td>
<td>0.1226</td>
</tr>
<tr>
<td>3</td>
<td>0.0401</td>
<td>0.0368</td>
<td>0.0472</td>
<td>0.4338</td>
<td>0.1101</td>
<td>0.1348</td>
</tr>
<tr>
<td>3.5</td>
<td>0.0517</td>
<td>0.0439</td>
<td>0.0540</td>
<td>0.4899</td>
<td>0.1184</td>
<td>0.1445</td>
</tr>
<tr>
<td>4</td>
<td>0.0628</td>
<td>0.0492</td>
<td>0.0584</td>
<td>0.5452</td>
<td>0.1238</td>
<td>0.1512</td>
</tr>
<tr>
<td>4.5</td>
<td>0.0758</td>
<td>0.0542</td>
<td>0.0612</td>
<td>0.5984</td>
<td>0.1277</td>
<td>0.1550</td>
</tr>
<tr>
<td>5</td>
<td>0.0880</td>
<td>0.0579</td>
<td>0.0616</td>
<td>0.6522</td>
<td>0.1294</td>
<td>0.1557</td>
</tr>
</tbody>
</table>

Figures 2–3 display the trends of the relative changes of each parameter against different values of Δ under both TtR and TnR models in doubly truncated and one-sided truncated situations, with the true model being the TtR and TnR, respectively. We observe that the
percentage of changes on estimations of the parameters from the TtR model are always smaller than that from the TnR model when varying $\Delta$ from 0 to 5. Especially in the one-sided truncated case, the changes on estimations of all parameters are extremely significant once the data have been contaminated. Thus, the TtR model is less affected by data contamination.

Table 4. Percentage of change on estimation of the parameters (Doubly truncated TnR)

<table>
<thead>
<tr>
<th>$\Delta$</th>
<th>TtR $\beta_0$</th>
<th>TtR $\beta_1$</th>
<th>TtR $\sigma^2$</th>
<th>TnR $\beta_0$</th>
<th>TnR $\beta_1$</th>
<th>TnR $\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0063</td>
<td>0.0062</td>
<td>0.0124</td>
<td>0.0227</td>
</tr>
<tr>
<td>1</td>
<td>0.0013</td>
<td>0.0018</td>
<td>0.0160</td>
<td>0.0120</td>
<td>0.0229</td>
<td>0.0398</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0033</td>
<td>0.0060</td>
<td>0.0288</td>
<td>0.0172</td>
<td>0.0316</td>
<td>0.0512</td>
</tr>
<tr>
<td>2</td>
<td>0.0098</td>
<td>0.0158</td>
<td>0.0396</td>
<td>0.0220</td>
<td>0.0387</td>
<td>0.0571</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0178</td>
<td>0.0261</td>
<td>0.0401</td>
<td>0.0263</td>
<td>0.0443</td>
<td>0.0577</td>
</tr>
<tr>
<td>3</td>
<td>0.0224</td>
<td>0.0307</td>
<td>0.0303</td>
<td>0.0303</td>
<td>0.0484</td>
<td>0.0531</td>
</tr>
<tr>
<td>3.5</td>
<td>0.0232</td>
<td>0.0305</td>
<td>0.0181</td>
<td>0.0338</td>
<td>0.0512</td>
<td>0.0432</td>
</tr>
<tr>
<td>4</td>
<td>0.0223</td>
<td>0.0289</td>
<td>0.0089</td>
<td>0.0370</td>
<td>0.0524</td>
<td>0.0292</td>
</tr>
<tr>
<td>4.5</td>
<td>0.0211</td>
<td>0.0270</td>
<td>0.0044</td>
<td>0.0398</td>
<td>0.0522</td>
<td>0.0211</td>
</tr>
<tr>
<td>5</td>
<td>0.0198</td>
<td>0.0252</td>
<td>0.0043</td>
<td>0.0423</td>
<td>0.0510</td>
<td>0.0324</td>
</tr>
</tbody>
</table>

Table 5. Percentage of change on estimation of the parameters (One-sided truncated TnR)

<table>
<thead>
<tr>
<th>$\Delta$</th>
<th>TtR $\beta_0$</th>
<th>TtR $\beta_1$</th>
<th>TtR $\sigma^2$</th>
<th>TnR $\beta_0$</th>
<th>TnR $\beta_1$</th>
<th>TnR $\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.0185</td>
<td>0.0027</td>
<td>0.0028</td>
<td>0.0873</td>
<td>0.0314</td>
<td>0.0434</td>
</tr>
<tr>
<td>1</td>
<td>0.0394</td>
<td>0.0055</td>
<td>0.0073</td>
<td>0.1811</td>
<td>0.0561</td>
<td>0.0770</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0608</td>
<td>0.0082</td>
<td>0.0145</td>
<td>0.2524</td>
<td>0.0759</td>
<td>0.1019</td>
</tr>
<tr>
<td>2</td>
<td>0.0751</td>
<td>0.0103</td>
<td>0.0247</td>
<td>0.3075</td>
<td>0.0920</td>
<td>0.1196</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0846</td>
<td>0.0145</td>
<td>0.0371</td>
<td>0.3498</td>
<td>0.1036</td>
<td>0.1297</td>
</tr>
<tr>
<td>3</td>
<td>0.1174</td>
<td>0.0248</td>
<td>0.0472</td>
<td>0.3818</td>
<td>0.1111</td>
<td>0.1324</td>
</tr>
<tr>
<td>3.5</td>
<td>0.1879</td>
<td>0.0380</td>
<td>0.0497</td>
<td>0.4052</td>
<td>0.1148</td>
<td>0.1278</td>
</tr>
<tr>
<td>4</td>
<td>0.2693</td>
<td>0.0488</td>
<td>0.0442</td>
<td>0.4220</td>
<td>0.1150</td>
<td>0.1167</td>
</tr>
<tr>
<td>4.5</td>
<td>0.3163</td>
<td>0.0533</td>
<td>0.0344</td>
<td>0.4320</td>
<td>0.1104</td>
<td>0.1003</td>
</tr>
<tr>
<td>5</td>
<td>0.3439</td>
<td>0.0544</td>
<td>0.0247</td>
<td>0.4372</td>
<td>0.1028</td>
<td>0.0868</td>
</tr>
</tbody>
</table>
Figure 2: Relative changes on the maximum likelihood estimations of $\beta$ and $\sigma^2$ from the TtR (solid line) and ThR (dashed line) models for different contaminations $\Delta$ in the doubly truncated and one-sided truncated cases based on the generated data from the TtR model.
Figure 3: Relative changes on the maximum likelihood estimations of $\beta$ and $\sigma^2$ from the TtR (solid line) and TnR (dashed line) models for different contaminations $\Delta$ in the doubly truncated and one-sided truncated cases based on the generated data from the TnR model.
6. Applications

6.1 Clinical trial data

We use the data provided by Shih & Weisberg (1986), which are collected from a clinical trial on 34 male patients as listed in Table 6. The outcome variable is the endogenous creatinine (CR) clearance and there exist three measured covariates which are body weight (WT) in kilogram, serum creatinine (SC) concentration in mg/deciliter, and AGE in years. From the data set, we see that only 28 of 34 patients have full records as two have missing WT and four have missing SC. A typical model recommended in many pharmacokinetics textbooks for modeling CR as a function of WT, SC, and AGE is of the form

\[ E[\log(\text{CR})] = \beta_0 + \beta_1 \log(\text{WT}) + \beta_2 \log(\text{SC}) + \beta_3 \log(140 - \text{AGE}), \]

where the logarithm of variable CR is taken as the response variable.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>CR</th>
<th>WT</th>
<th>SC</th>
<th>AGE</th>
<th>Patient ID</th>
<th>CR</th>
<th>WT</th>
<th>SC</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>132.0</td>
<td>71.0</td>
<td>0.71253</td>
<td>38</td>
<td>18</td>
<td>130.0</td>
<td>85.0</td>
<td>1.11969</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>53.0</td>
<td>69.0</td>
<td>1.48161</td>
<td>78</td>
<td>19</td>
<td>94.0</td>
<td>68.0</td>
<td>1.37982</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>50.0</td>
<td>85.0</td>
<td>2.20545</td>
<td>69</td>
<td>20</td>
<td>130.0</td>
<td>65.0</td>
<td>1.11969</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>82.0</td>
<td>100.0</td>
<td>1.42505</td>
<td>70</td>
<td>21</td>
<td>59.0</td>
<td>53.0</td>
<td>0.97266</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>110.0</td>
<td>59.0</td>
<td>0.67860</td>
<td>45</td>
<td>22</td>
<td>38.0</td>
<td>50.0</td>
<td>1.60602</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>100.0</td>
<td>73.0</td>
<td>0.75777</td>
<td>65</td>
<td>23</td>
<td>65.0</td>
<td>74.0</td>
<td>1.58339</td>
<td>66</td>
</tr>
<tr>
<td>7</td>
<td>68.0</td>
<td>63.0</td>
<td>1.11969</td>
<td>76</td>
<td>24</td>
<td>85.0</td>
<td>67.0</td>
<td>1.40244</td>
<td>31</td>
</tr>
<tr>
<td>8</td>
<td>92.0</td>
<td>81.0</td>
<td>0.91611</td>
<td>61</td>
<td>25</td>
<td>140.0</td>
<td>80.0</td>
<td>0.67860</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>60.0</td>
<td>74.0</td>
<td>1.54947</td>
<td>68</td>
<td>26</td>
<td>80.0</td>
<td>67.0</td>
<td>1.19886</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>94.0</td>
<td>87.0</td>
<td>0.93873</td>
<td>64</td>
<td>27</td>
<td>4.30</td>
<td>68.0</td>
<td>7.60001</td>
<td>81</td>
</tr>
<tr>
<td>11</td>
<td>105.0</td>
<td>79.0</td>
<td>0.99528</td>
<td>66</td>
<td>28</td>
<td>43.2</td>
<td>72.2</td>
<td>2.10001</td>
<td>43</td>
</tr>
<tr>
<td>12</td>
<td>98.0</td>
<td>93.0</td>
<td>1.07445</td>
<td>49</td>
<td>29</td>
<td>75.0</td>
<td>NA</td>
<td>1.35719</td>
<td>78</td>
</tr>
<tr>
<td>13</td>
<td>112.0</td>
<td>60.0</td>
<td>0.70122</td>
<td>43</td>
<td>30</td>
<td>41.0</td>
<td>NA</td>
<td>1.05183</td>
<td>38</td>
</tr>
<tr>
<td>14</td>
<td>125.0</td>
<td>70.0</td>
<td>0.71253</td>
<td>42</td>
<td>31</td>
<td>120.0</td>
<td>107.0</td>
<td>NA</td>
<td>62</td>
</tr>
<tr>
<td>15</td>
<td>108.0</td>
<td>83.0</td>
<td>0.99528</td>
<td>66</td>
<td>32</td>
<td>52.0</td>
<td>75.0</td>
<td>NA</td>
<td>70</td>
</tr>
<tr>
<td>16</td>
<td>30.0</td>
<td>70.0</td>
<td>2.52212</td>
<td>78</td>
<td>33</td>
<td>73.0</td>
<td>62.0</td>
<td>NA</td>
<td>63</td>
</tr>
<tr>
<td>17</td>
<td>111.0</td>
<td>73.0</td>
<td>1.13100</td>
<td>35</td>
<td>34</td>
<td>57.0</td>
<td>52.0</td>
<td>NA</td>
<td>68</td>
</tr>
</tbody>
</table>

NOTE: ‘NA’ indicates the value is not available.
As in medical measurement, the creatinine levels depend on many factors, such as weight, age, gender. In general, the measurement of creatinine is usually within some normal ranges, especially it is upper bounded. For the purpose of illustration, we consider to fit the response variable denoted by $Y = \log(CR)$ for the first 28 subjects using the following right-truncated $t$ regression model:

$$Y_i \sim \text{Tt}(w_i^\top \beta, \sigma^2, \nu; (-\infty, b]), \quad i = 1, \ldots, 28,$$

where $w_i^\top = (1, \log(WT_i), \log(SC_i), \log(140 - AGE_i))^\top$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$ and $b$ is chosen to be six, which is greater than the maximum of $\log(CR)$ among all observations. The MLEs of all parameters are presented in Table 7, which are obtained via the EM algorithm after 107 iterations with initial values being $\beta^{(0)} = (-2.9740, 0.9937, -1.0849, 0.7330)^\top$, $\sigma^{2(0)} = 0.0379$ from OLS estimates and $\nu^{(0)} = 2$, the precision is $10^{-7}$. As the sample size $n = 28$ is really small, the bootstrap method is adopted to construct the confidence interval for parameters. By generating 6000 bootstrap samples, the standard deviations and two types of bootstrap CIs, i.e., SCI and PCI, are computed, respectively. All results are summarized in Table 7. Meanwhile, for comparative purposes, the estimation results by TnR model are also provided in Table 7. Based on the log-likelihood and AIC value, TtR model gives a better fit on this data set than TnR model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>StdB</th>
<th>SCI</th>
<th>PCI</th>
<th>TnR MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-3.0026</td>
<td>0.8995</td>
<td>$[-4.7286, -1.2024]$</td>
<td>$[-4.3432, -1.4637]$</td>
<td>-2.9741</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.8797</td>
<td>0.1711</td>
<td>[ 0.5338, 1.2044 ]</td>
<td>[ 0.4550, 1.0477 ]</td>
<td>0.9937</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.7178</td>
<td>0.0397</td>
<td>$[-0.8039, -0.6482]$</td>
<td>$[-0.7847, -0.6481]$</td>
<td>-1.0849</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.8446</td>
<td>0.1694</td>
<td>[ 0.5186, 1.1825 ]</td>
<td>[ 0.4616, 1.0828 ]</td>
<td>0.7330</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0104</td>
<td>0.0029</td>
<td>[ 0.0006, 0.0121 ]</td>
<td>[ 0.0011, 0.0121 ]</td>
<td>0.0378</td>
</tr>
<tr>
<td>$\nu$</td>
<td>2.0067</td>
<td>0.8498</td>
<td>[ 1.425, 3.4737 ]</td>
<td>[ 0.6897, 3.2416 ]</td>
<td>—</td>
</tr>
<tr>
<td>log-$L$</td>
<td>9.7103</td>
<td></td>
<td></td>
<td></td>
<td>6.0946</td>
</tr>
<tr>
<td>AIC</td>
<td>-7.4206</td>
<td></td>
<td></td>
<td></td>
<td>-2.1891</td>
</tr>
</tbody>
</table>

NOTE: StdB = the standard deviation of bootstrap samples.
6.2 Insulation life data with censoring times

We analyze the data set used by Massuia et al. (2015) to illustrate the correct results for Student-t censored linear regression model. The data set is from Tan et al. (2010), which concentrates on accelerated life tests on electrical insulation in 40 motorettes. Ten motorettes were tested at each of the four temperatures, and testing was terminated at different times at each temperature level. The data are summarized in Table 8, where \( y_i = \log_{10}(\text{failure time}) \) and \( T_i = 1000/(\text{temperature} + 273.2) \). Notice that the first \( m = 17 \) observations are uncensored data, and the remaining \( n - m = 23 \) are censored.

<table>
<thead>
<tr>
<th>i</th>
<th>( y_i )</th>
<th>( T_i )</th>
<th>i</th>
<th>( y_i )</th>
<th>( T_i )</th>
<th>i</th>
<th>( y_i )</th>
<th>( T_i )</th>
<th>i</th>
<th>( y_i )</th>
<th>( T_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.2464</td>
<td>2.2563</td>
<td>13</td>
<td>2.6107</td>
<td>2.0276</td>
<td>18</td>
<td>3.9066</td>
<td>2.3629</td>
<td>30</td>
<td>3.7362</td>
<td>2.2563</td>
</tr>
<tr>
<td>2</td>
<td>3.4428</td>
<td>2.2563</td>
<td>14</td>
<td>2.6107</td>
<td>2.0276</td>
<td>19</td>
<td>3.9066</td>
<td>2.3629</td>
<td>31</td>
<td>3.2253</td>
<td>2.1589</td>
</tr>
<tr>
<td>3</td>
<td>3.5371</td>
<td>2.2563</td>
<td>15</td>
<td>2.7024</td>
<td>2.0276</td>
<td>20</td>
<td>3.9066</td>
<td>2.3629</td>
<td>32</td>
<td>3.2253</td>
<td>2.1589</td>
</tr>
<tr>
<td>4</td>
<td>3.5492</td>
<td>2.2563</td>
<td>16</td>
<td>2.7024</td>
<td>2.0276</td>
<td>21</td>
<td>3.9066</td>
<td>2.3629</td>
<td>33</td>
<td>3.2253</td>
<td>2.1589</td>
</tr>
<tr>
<td>5</td>
<td>3.5775</td>
<td>2.2563</td>
<td>17</td>
<td>2.7024</td>
<td>2.0276</td>
<td>22</td>
<td>3.9066</td>
<td>2.3629</td>
<td>34</td>
<td>3.2253</td>
<td>2.1589</td>
</tr>
<tr>
<td>6</td>
<td>3.6866</td>
<td>2.2563</td>
<td>18</td>
<td>3.9066</td>
<td>2.3629</td>
<td>35</td>
<td>3.2253</td>
<td>2.1589</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.7157</td>
<td>2.2563</td>
<td>19</td>
<td>3.9066</td>
<td>2.3629</td>
<td>36</td>
<td>2.7226</td>
<td>2.0276</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.6107</td>
<td>2.1589</td>
<td>20</td>
<td>3.9066</td>
<td>2.3629</td>
<td>37</td>
<td>2.7226</td>
<td>2.0276</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2.6107</td>
<td>2.1589</td>
<td>21</td>
<td>3.9066</td>
<td>2.3629</td>
<td>38</td>
<td>2.7226</td>
<td>2.0276</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.1284</td>
<td>2.1589</td>
<td>22</td>
<td>3.9066</td>
<td>2.3629</td>
<td>39</td>
<td>2.7226</td>
<td>2.0276</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3.1284</td>
<td>2.1589</td>
<td>23</td>
<td>3.9066</td>
<td>2.3629</td>
<td>40</td>
<td>2.7226</td>
<td>2.0276</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3.1584</td>
<td>2.1589</td>
<td>24</td>
<td>3.9066</td>
<td>2.3629</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Let \( y_i \) be the observed value of the response variable, and the covariates are \( x_i = (1, T_i)^T \). We consider modeling the data with Student-t right-censored regression model in (4.1) by setting \( \tau_i, \kappa_i = [c_i, \infty) \). First, we compare the results computed by the proposed methods with those provided by Massuia et al. (2015), where the degree of freedom is fixed at the value of \( \nu = 2 \). The results are listed in Table 9, where we can see that the estimates of parameters are divergent.

Furthermore, we do not fix the value of \( \nu \) in advance and estimate all parameters \( \{\beta, \sigma^2, \nu\} \), where \( \beta = (\beta_0, \beta_1)^T \). By adopting the pseudo OLS estimates \( \beta^{(0)} = (-4.9326, 3.7480)^T \),
\( \sigma^2(0) = 0.0235 \) and \( \nu(0) = 2 \) as the initial values of the parameters, the MLEs of the parameters are obtained via the EM algorithm by 63 iterations with precision \( 10^{-7} \) and are listed in Table 10. The standard deviations of the MLEs are approximated by the square roots of the diagonal elements of the inverse negative Hessian matrix. Both the Stds and the 95\% asymptotic Wald CIs are also provided in Table 10. However, we observed that the Wald CIs for \( \sigma^2 \) and \( \nu \) are invalid since the lower bounds are less than 0, which is due to that this is an asymptotic test and the sample size is not large enough in this case. Instead, we generate 1000 bootstrap samples to calculate the standard errors and two types of bootstrap CIs, i.e., SCI and PCI.

Table 9. Comparison of parameter estimation results for the insulation life data assuming fixed \( \nu = 2 \)

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our proposed method</td>
<td>-5.6376</td>
<td>4.1214</td>
<td>0.0143</td>
</tr>
<tr>
<td>Masssuia et al. (2015)</td>
<td>-5.7817</td>
<td>4.1964</td>
<td>0.0302</td>
</tr>
</tbody>
</table>

Table 10. Estimations of the parameters for the insulation life data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \sigma^2 )</th>
<th>( \nu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>-5.6369</td>
<td>4.1208</td>
<td>0.0138</td>
<td>1.8998</td>
</tr>
<tr>
<td>Std</td>
<td>0.5510</td>
<td>0.2581</td>
<td>0.0086</td>
<td>1.0206</td>
</tr>
<tr>
<td>Wald CI</td>
<td>[-6.7168, -4.5571]</td>
<td>[3.6149, 4.6267]</td>
<td>[-0.0030, 0.0307]</td>
<td>[-0.1006, 3.9002]</td>
</tr>
<tr>
<td>StdB</td>
<td>0.5370</td>
<td>0.2510</td>
<td>0.0077</td>
<td>0.9435</td>
</tr>
<tr>
<td>SCI</td>
<td>[-6.6870, -4.5819]</td>
<td>[3.6615, 4.6454]</td>
<td>[-0.0045, 0.0256]</td>
<td>[0.1913, 3.8900]</td>
</tr>
<tr>
<td>PCI</td>
<td>[-6.6781, -4.4958]</td>
<td>[3.6214, 4.6464]</td>
<td>[0.0010, 0.0288]</td>
<td>[0.7134, 4.3504]</td>
</tr>
<tr>
<td>log-L</td>
<td>-9.9662</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: StdB = the standard deviation of bootstrap samples.

7. Discussion

The effective truncation has been widely acknowledged in statistical modeling. For example, the univariate truncated normal distributions were extensively studied by Cohen (1949, 1950,
1959, 1961), Halperin (1952), and DePriest (1983). Tian et al. (2018) analyzed truncated data by a truncated normal distribution, in which the EM algorithm was developed via a SR, and explicit expressions were formulated. Properties for other popular truncated distributions have been summarized in Johnson et al. (1994, 1995) and Johnson et al. (2005).

In the paper, we carefully studied the statistical properties of the truncated $t$ distribution. By abandoning the wrong scale mixture through a truncated normal distribution, we recalculated the correct moments of a truncated $t$ distribution. It is worth mentioning that the moments of a doubly truncated $t$ distribution always exist, which is very different from the non-truncated situation. These results are important since when we adopt the EM algorithm to obtain the MLEs of parameters, they will be used. We proposed three valid SRs for a truncated $t$ variate, where the third SR motivated an EM algorithm and the correct truncated moments facilitated the EM algorithm’s implementation much more straightforward. The proposed method was also applied to a $t$ interval–censored regression model by utilizing the relationship between the truncated distribution and censored data.

The above results are merely effective for the univariate case. It is desirable to extend the method and derive all the valid properties for the multivariate truncated $t$ distribution as we have pointed out the same mistake that appeared in the multivariate case.

**Acknowledgments**

The authors are grateful to the editor and three anonymous referees’ valuable comments and suggestions for significant improvements of the paper. Chi ZHANG’s research was supported by National Natural Science Foundation of China (Grant no. 11801380). Guo-Liang TIAN’s research was fully supported by National Natural Science Foundation of China (Grant no. 11771199). Yu FEI’s research was fully supported by National Natural Science Foundation of China (Grant no. 11971421) and Yunling Scholar Research Fund of Yunnan Province.
Appendix A: Some Technical proofs

A.1 Derivation of moments

By using the Taylor series of function $e^{ty} = \sum_{n=0}^{\infty} \frac{(ty)^n}{n!}$, first we have

$$E(Y_0^k) = \frac{d^k M_Y(t)}{dt^k} \Big|_{t=0} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_{0}^{\infty} y^k \cdot e^{ty} \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} dy \Big|_{t=0}$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_{0}^{\infty} y^k \cdot (\sum_{n=0}^{\infty} \frac{(ty)^n}{n!}) \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} dy \Big|_{t=0}$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_{0}^{\infty} y^k \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} dy.$$

The first-order moment of $Y_0$ which is derived as

$$E(Y_0) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_{0}^{\infty} y \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} dy$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \cdot \left(-\frac{1}{\nu-1}\right) \left[(\nu + y^2)^{-\frac{\nu+1}{2}}\right]_{\alpha}^{\beta}$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - (\nu + \beta^2)^{-\frac{\nu+1}{2}}\right],$$

for $\nu > 0$ and $\nu \neq 1$.

The second-order moment of $Y_0$ is

$$E(Y_0^2) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_{0}^{\infty} y^2 \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} dy$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \cdot \left(-\frac{1}{\nu-1}\right) \int_{\alpha}^{\beta} y d(\nu + y^2)^{-\frac{\nu+1}{2}} dy$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu+1}{2}}\right]$$

$$+ \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_{\alpha}^{\beta} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}} dy$$

[let $y = z\sqrt{\frac{\nu}{\nu - 2}}$]

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu+1}{2}}\right]$$

$$+ \frac{\nu}{2c \cdot \nu^2 \cdot \sqrt{\nu - 2}} \int_{\alpha_1}^{\beta_1} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\nu - 2}} \left(1 + \frac{z^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}} dz$$
\[
\frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ \frac{\alpha}{(\nu + \alpha^2)^{\frac{\nu+1}{2}}} - \frac{\beta}{(\nu + \beta^2)^{\frac{\nu+1}{2}}} \right] + \frac{\nu [F_{\nu-2}(\beta_1) - F_{\nu-2}(\alpha_1)]}{(\nu - 2)c},
\]
for \( \nu > 2 \).

The third-order moment of \( Y_0 \) is derived as

\[
E(Y_0^3) = \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \int_\alpha^\beta y^3 \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} \, dy
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \cdot \left( -\frac{1}{\nu - 1} \right) \int_\alpha^\beta y^3 \, d(\nu + y^2)^{-\frac{\nu+1}{2}}
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ y^2(\nu + y^2)^{-\frac{\nu+1}{2}} \right]_\alpha^\beta + \int_\alpha^\beta 2y(\nu + y^2)^{-\frac{\nu+1}{2}} \, dy
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ \alpha^2(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - \beta^2(\nu + \beta^2)^{-\frac{\nu+1}{2}} \right]
\]

\[
+ \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \cdot \left( -\frac{2}{\nu - 3} \right) \left[ (\nu + y^2)^{-\frac{\nu+1}{2}} \right]_\alpha^\beta
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ \alpha^2(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - \beta^2(\nu + \beta^2)^{-\frac{\nu+1}{2}} \right]
\]

\[
+ \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ (\nu + \alpha^2)^{-\frac{\nu-3}{2}} - (\nu + \beta^2)^{-\frac{\nu-3}{2}} \right],
\]
for \( \nu > 0 \) and \( \nu \neq 1, \nu \neq 3 \).

The fourth-order moment of \( Y_0 \) is given by

\[
E(Y_0^4) = \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \int_\alpha^\beta y^4 \cdot (\nu + y^2)^{-\frac{\nu+1}{2}} \, dy
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \cdot \left( -\frac{1}{\nu - 1} \right) \int_\alpha^\beta y^3 \, d(\nu + y^2)^{-\frac{\nu+1}{2}}
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ y^3(\nu + y^2)^{-\frac{\nu+1}{2}} \right]_\alpha^\beta + \int_\alpha^\beta 3y^2(\nu + y^2)^{-\frac{\nu+1}{2}} \, dy
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ \alpha^3(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - \beta^3(\nu + \beta^2)^{-\frac{\nu+1}{2}} \right]
\]

\[
+ \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \cdot \left( -\frac{3}{\nu - 3} \right) \int_\alpha^\beta y^2 \, d(\nu + y^2)^{-\frac{\nu+1}{2}}
\]

\[
= \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ \alpha^3(\nu + \alpha^2)^{-\frac{\nu+1}{2}} - \beta^3(\nu + \beta^2)^{-\frac{\nu+1}{2}} \right]
\]

\[
+ \frac{\Gamma \left( \frac{\nu+1}{2} \right) \nu^\frac{3}{2}}{2c \cdot \Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right)} \left[ (\nu + \alpha^2)^{-\frac{\nu-3}{2}} - (\nu + \beta^2)^{-\frac{\nu-3}{2}} \right],
\]
for \( \nu > 0 \) and \( \nu \neq 1, \nu \neq 3 \).
\[
\begin{align*}
&\frac{\Gamma\left(\frac{\nu-1}{2}\right) \nu^\frac{\nu}{2}}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha^3(\nu + \alpha^2)^{-\frac{\nu-1}{2}} - \beta^3(\nu + \beta^2)^{-\frac{\nu-1}{2}}\right] \\
+ &\frac{3\Gamma\left(\frac{\nu-3}{2}\right) \nu^{\frac{\nu}{2}}}{4c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha(\nu + \alpha^2)^{-\frac{\nu-3}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu-3}{2}}\right] \\
+ &\frac{3\Gamma\left(\frac{\nu-3}{2}\right) \nu^{\frac{\nu}{2}}}{4c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \int_0^\beta \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu-3}{2}} \, dy \\
&\quad \text{[let } y = z \sqrt{\frac{\nu}{\nu - 4}}\text{]} \\
&\frac{\Gamma\left(\frac{\nu-1}{2}\right) \nu^\frac{\nu}{2}}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha^3(\nu + \alpha^2)^{-\frac{\nu-1}{2}} - \beta^3(\nu + \beta^2)^{-\frac{\nu-1}{2}}\right] \\
+ &\frac{3\Gamma\left(\frac{\nu-3}{2}\right) \nu^{\frac{\nu}{2}}}{4c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha(\nu + \alpha^2)^{-\frac{\nu-3}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu-3}{2}}\right] \\
+ &\frac{3\nu^2}{4c \cdot \nu^{-\frac{1}{2}} \cdot \nu^{-\frac{3}{2}}} \int_{\alpha_2}^{\beta_2} \frac{\Gamma\left(\frac{\nu-3}{2}\right) \Gamma\left(\frac{1}{2}\right)}{\sqrt{\nu - 4}} \left(1 + \frac{z^2}{\nu - 4}\right)^{-\frac{\nu-3}{2}} dz \\
&\left\{\alpha^3(\nu + \alpha^2)^{-\frac{\nu-1}{2}} - \beta^3(\nu + \beta^2)^{-\frac{\nu-1}{2}}\right] \\
+ &\frac{3\Gamma\left(\frac{\nu-3}{2}\right) \nu^{\frac{\nu}{2}}}{4c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha(\nu + \alpha^2)^{-\frac{\nu-3}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu-3}{2}}\right] \\
+ &\frac{3\nu^2}{4c \cdot \nu^{-\frac{1}{2}} \cdot \nu^{-\frac{3}{2}}} \int_{\alpha_2}^{\beta_2} \frac{\Gamma\left(\frac{\nu-3}{2}\right) \Gamma\left(\frac{1}{2}\right)}{\sqrt{\nu - 4}} \left(1 + \frac{z^2}{\nu - 4}\right)^{-\frac{\nu-3}{2}} dz \\
&\frac{\nu^2}{2c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha^3(\nu + \alpha^2)^{-\frac{\nu-1}{2}} - \beta^3(\nu + \beta^2)^{-\frac{\nu-1}{2}}\right] \\
+ &\frac{3\Gamma\left(\frac{\nu-3}{2}\right) \nu^{\frac{\nu}{2}}}{4c \cdot \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{1}{2}\right)} \left[\alpha(\nu + \alpha^2)^{-\frac{\nu-3}{2}} - \beta(\nu + \beta^2)^{-\frac{\nu-3}{2}}\right] \\
+ &\frac{3\nu^2}{4c \cdot \nu^{-\frac{1}{2}} \cdot \nu^{-\frac{3}{2}}} \int_{\alpha_2}^{\beta_2} \frac{\Gamma\left(\frac{\nu-3}{2}\right) \Gamma\left(\frac{1}{2}\right)}{\sqrt{\nu - 4}} \left(1 + \frac{z^2}{\nu - 4}\right)^{-\frac{\nu-3}{2}} dz \right]\frac{\nu^2}{(\nu - 2)(\nu - 4)c}, \\
\end{align*}
\]
for \(\nu > 4\).

### A.2 Limiting results

First note that \(\nu\) is non-negative, if \(\nu - \ell > 0\), we have

\[
(\nu + \alpha^2)^{\nu - \ell}/2 > (\alpha^2)^{\nu - \ell}/2 \quad \text{and} \quad (\nu + \beta^2)^{\nu - \ell}/2 > (\beta^2)^{\nu - \ell}/2.
\]

Given \(k, \ell\) which are non-negative integers, it follows that

\[
0 \leq \lim_{\alpha \to \infty} \left|\frac{\alpha^k}{(\nu + \alpha^2)^{\nu - \ell}/2}\right| = \lim_{\alpha \to \infty} \left|\frac{1}{\alpha^{\nu - \ell}}\right| = \lim_{\alpha \to \infty} \left|\frac{1}{\alpha^{\nu - k - \ell}}\right| = 0, \quad \text{if} \quad \nu > k + \ell,
\]

\[
0 \leq \lim_{\beta \to \infty} \left|\frac{\beta^k}{(\nu + \beta^2)^{\nu - \ell}/2}\right| = \lim_{\beta \to \infty} \left|\frac{1}{\beta^{\nu - \ell}}\right| = \lim_{\beta \to \infty} \left|\frac{1}{\beta^{\nu - k - \ell}}\right| = 0, \quad \text{if} \quad \nu > k + \ell.
\]

Combining with expressions in Lemmas 1–4, for \(\nu > k + \ell\), when \(\ell = 1, k = 0, 1, 2, 3\) and when \(\ell = 3, k = 0, 1\), we always have

\[
\lim_{\alpha \to \infty} \alpha^k(\nu + \alpha^2)^{-(\nu - \ell)/2} = 0 \quad \text{and} \quad \lim_{\beta \to \infty} \beta^k(\nu + \beta^2)^{-(\nu - \ell)/2} = 0.
\]

Finally, the first four moments for standard \(t(\nu)\) are easily obtained by using the above limiting results.
References


