UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 05-47
Statistics and Econometrics Series 09
July 2005

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

BAYESIAN INFERENCE FOR THE HALF-NORMAL AND HALF-T DISTRIBUTIONS

M.P. Wiper, F.J. Girón, A. Pewsey*

Abstract

In this article we consider approaches to Bayesian inference for the half-normal and half-t distributions. We show that a generalized version of the normal-gamma distribution is conjugate to the half-normal likelihood and give the moments of this new distribution. The bias and coverage of the Bayesian posterior mean estimator of the half-normal location parameter are compared with those of maximum likelihood based estimators. Inference for the half-t distribution is performed using Gibbs sampling and model comparison is carried out using Bayes factors. A real data example is presented which demonstrates the fitting of the half-normal and half-t models.

Keywords: Bias-correction; Gaussian-modulated gamma distribution; Gibbs sampling; likelihood based inference; model selection; right-truncated normal-gamma distribution.

* Wiper, Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: mwiper@est-econ.uc3m.es; Girón, Departamento de Estadística e Investigación Operativa, Universidad de Malaga; Pewsey, Departamento de Matemáticas, Universidad de Extremadura Departamento de Matemáticas, Universidad de Extremadura.

Bayesian inference for the half-normal and half-t distributions

M.P. Wiper*, F.J. Girón†, & A. Pewsey‡

July 21, 2005

Abstract

In this article we consider approaches to Bayesian inference for the half-normal and half-t distributions. We show that a generalized version of the normal-gamma distribution is conjugate to the half-normal likelihood and give the moments of this new distribution. The bias and coverage of the Bayesian posterior mean estimator of the half-normal location parameter are compared with those of maximum likelihood based estimators. Inference for the half-t distribution is performed using Gibbs sampling and model comparison is carried out using Bayes factors. A real data example is presented which demonstrates the fitting of the half-normal and half-t models.
KEY WORDS: Bias-correction; Gaussian-modulated gamma distribution; Gibbs sampling; likelihood based inference; model selection; right-truncated normal-gamma distribution.

1 Introduction

The half-normal distribution has been used as a model for (left) truncated data from application areas as diverse as fibre buckling (Haberle 1991), blowfly dispersion (Dobzhansky and Wright 1947), sports science physiology (Pewsey 2002, 2004) and, in particular, stochastic frontier modelling (Aigner *et al.*, 1977; Meeusen and van den Broeck, 1977). Likelihood based inference for the half-normal distribution has been considered by Pewsey (2002, 2004).

However, for heavy-tailed data, the half-normal distribution will not be an adequate model and then a half-t distribution might be considered as a more flexible alternative. For one of the few applications of this latter model, see Tancredi (2002).

The principal objective of this article is to illustrate that fully conjugate Bayesian inference can be carried out for the half-normal model and that Gibbs

*Departamento de Estadística, Universidad Carlos III de Madrid
†Departamento de Estadística e Investigación Operativa, Universidad de Malaga
‡Departamento de Matemáticas, Universidad de Extremadura

sampling techniques can be used to perform Bayesian inference for the parameters of the half-t model.

The article is structured as follows. In Section 2, we review the definition of the half-normal distribution and comment on likelihood based inference for its parameters. In Section 3, we illustrate how conjugate Bayesian inference for the half-normal distribution can be undertaken. Results for the posterior moments of the location and scale parameters are also given. In Section 4, we compare the properties of Bayesian point and interval estimators with those of their likelihood based counterparts. In Section 5, in which Bayesian inference for the half-t model is considered, we present a simple Gibbs sampling algorithm that can be used to sample the posterior parameter distributions. Section 6 focuses on model selection, and in Section 7 the half-normal and half-t two models are fitted to a real data set. Finally, in Section 8, we discuss our findings and contemplate some extensions to the work presented here.

2 The half-normal distribution and likelihood based inference

X has a half-normal distribution, with location and scale parameters ξ and η, that is $X|\xi,\eta \sim \mathcal{HN}(\xi,\eta)$, if

$$f(x|\xi,\eta) = \sqrt{\frac{2}{\pi}}\frac{1}{\eta}\exp\left\{-\frac{1}{2\eta^2}(x-\xi)^2\right\}, \tag{1}$$

where $x > \xi$, $-\infty < \xi < \infty$ and $\eta > 0$. Then $X = \xi + \eta|Z|$, where Z has a standard normal distribution. The moments of the half-normal distribution can be derived from general results for the truncated normal distribution, see e.g. Johnson *et al.* (1994). In particular,

$$E[X|\xi,\eta] = \xi + \eta\sqrt{\frac{2}{\pi}}. \tag{2}$$

Suppose that we observe a random sample of size n, $\mathbf{x} = (x_1,\ldots,x_n)$, from the half-normal distribution and wish to carry out inference for the unknown parameters of the model. Clearly, the maximum likelihood estimate (MLE) of the location parameter is given by $\hat{\xi} = x_{(1)} = \min\{x_i\}$. Pewsey (2002) gives the MLE of the scale parameter as $\hat{\eta} = \{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{(1)})^2\}^{1/2}$, while Pewsey (2004) demonstrates the superior sampling properties of the the bias-corrected estimate $\hat{\eta}_{BC} = \sqrt{\frac{n}{n-1}}\hat{\eta}$. In the latter paper it is also shown that point and interval estimation based on the bias-corrected estimate $\hat{\xi}_{BC} = \hat{\xi} - \hat{\eta}_{BC}\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)$, where $\Phi(\cdot)$ denotes the standard normal distribution function, outperforms that based on $\hat{\xi}$. The asymptotic, two tailed, $100(1-\alpha)\%$ confidence interval for ξ, incorporating bias-correction for η, proposed by Pewsey (2004) is

$$x_{(1)}+\log\left(\frac{\alpha}{2}\right)\hat{\eta}_{BC}\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)<\xi<x_{(1)}+\log\left(1-\frac{\alpha}{2}\right)\hat{\eta}_{BC}\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right). \tag{3}$$

This interval does not contain the MLE $\hat{\xi}=x_{(1)}$ as an endpoint and a one tailed interval is given by

$$x_{(1)}+\log\left(\alpha\right)\hat{\eta}_{BC}\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)<\xi<x_{(1)}. \tag{4}$$

An alternative to the use of classical inference is to use a Bayesian approach which has the advantage that if prior information is available, then this could also be incorporated. We explore this option in the following section.

3 Bayesian inference for the half-normal distribution

In order to undertake Bayesian inference for the half-normal model, it is first useful to reparameterize the distribution in terms of ξ and $\tau \stackrel{\text{def}}{=} \frac{1}{\eta^2}$. Then, we can write the half-normal model as

$$\begin{aligned} X|\xi,\tau &\sim \mathcal{HN}\left(\xi,\frac{1}{\tau}\right) \quad \text{if} \\ f(x|\xi,\tau) &= \sqrt{\frac{2\tau}{\pi}}\exp\left\{-\frac{\tau}{2}(x-\xi)^2\right\}, \quad x>\xi, -\infty<\xi<\infty, \tau>0. \end{aligned}$$

In Bayesian inference for the mean and precision parameters of an untruncated normal distribution, the natural conjugate prior distribution is a normal-gamma distribution; see e.g. Box and Tiao (1992). Here, we propose a generalization of this distribution which we shall call the *right-truncated normal-gamma* (RTNG) distribution which is defined below.

Definition 1 *We say that ξ,τ have a right-truncated normal-gamma distribution, $\xi,\tau \sim \mathcal{RTNG}(\xi_0,m,\alpha,a,b)$, where $-\infty<\xi_0<\infty$, $-\infty<m<\infty$, $\alpha>0$, $a>0$, and $b>0$, if*

$$f(\xi,\tau)=\frac{1}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(\alpha a)}}\right)}\frac{(b/2)^{a/2}}{\Gamma(a/2)}\sqrt{\frac{\alpha}{2\pi}}\tau^{\frac{a+1}{2}-1}\exp\left\{-\frac{\tau}{2}\left[b+\alpha(\xi-m)^2\right]\right\}, \tag{5}$$

for $\xi<\xi_0$, $\tau>0$, where $\Phi_d(\cdot)$ denotes the distribution function of the Student's t distribution with d degrees of freedom, that is

$$\begin{aligned} \Phi_d(z) &= \int_{-\infty}^{z}\phi_d(y)\,dy, \quad \textit{where} \\ \phi_d(y) &= \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\frac{1}{\sqrt{\pi d}}\left(1+\frac{y^2}{d}\right)^{-\frac{d+1}{2}} \end{aligned}$$

represents the density of the Student's t distribution at y, and obviously

$$\lim_{d\to\infty} \Phi_d(z) = \Phi(z)\,\forall\, z.$$

Clearly, when $\xi_0 \to \infty$, this distribution converges to the usual normal-gamma distribution.

It can easily be seen that this distribution results on assuming that $\kappa = \frac{\xi-m}{\sqrt{b/(a\alpha)}}$ has a Student's t distribution (with a degrees of freedom) truncated onto the region $\xi < \xi_0$, and that τ given ξ has a gamma distribution, i.e.

$$\begin{aligned}
f(\xi) &= \frac{1}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \sqrt{\frac{\alpha}{b\pi}} \left(1+\frac{1}{a}\left(\frac{\xi-m}{\sqrt{b/(a\alpha)}}\right)^2\right)^{-\frac{a+1}{2}} \\
&= \sqrt{\frac{a\alpha}{b}} \frac{\phi_a\left(\frac{\xi-m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)}, \quad \text{for } \xi < \xi_0, \quad \text{and} \\
\tau|\xi &\sim \mathcal{G}\left(\frac{a}{2}, \frac{b+\alpha(\xi-m)^2}{2}\right), \quad \text{that is} \\
f(\tau|\xi) &= \frac{\left(\frac{b+\alpha(\xi-m)^2}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \tau^{\frac{a}{2}-1} e^{-\frac{b+\alpha(\xi-m)^2}{2}\tau}, \quad \text{for } \tau > 0.
\end{aligned}$$

Noting that ξ given τ has a right-truncated normal distribution, we can derive the marginal density of τ. Thus,

$$f(\tau) = \frac{\Phi\left(\sqrt{\alpha\tau}(\xi_0-m)\right)}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \tau^{\frac{a}{2}-1} e^{-\frac{b}{2}\tau}.$$

This is a non-standard distribution which we shall refer to as a *Gaussian-modulated gamma* (GMG) distribution. We denote the fact as

$$\tau \sim \mathcal{GMG}\left(\sqrt{\alpha}(\xi_0-m), a, b\right).$$

The density for η can be derived from the density of τ via the usual change of variables formula. Thus, we have

$$f(\eta) = 2\frac{\Phi\left(\frac{\sqrt{\alpha}(\xi_0-m)}{\eta}\right)}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \eta^{-(a+1)} e^{-\frac{b}{2\eta^2}}. \tag{6}$$

Although the distributions of ξ, τ and η are non-standard, it is possible to derive their moments. The following theorem gives mean and variance formulae derived from the results in the appendices at the end of the paper.

Theorem 1 *Suppose that* $\xi, \tau \sim \mathcal{RTNG}(\xi_0, m, \alpha, a, b)$*, and let* $\eta = 1/\sqrt{\tau}$*. Then*

$$E[\xi] = m - \sqrt{\frac{b}{a\alpha}} \frac{a + \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)^2}{a-1} \frac{\phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \quad \text{for } a > 1,$$

$$V[\xi] = \frac{b}{a\alpha(a-2)} \left\{ a - \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right) \left(a + \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)^2\right) \frac{\phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \right\} + \frac{b}{a\alpha} \left(\frac{a + \left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)^2}{a-1} \frac{\phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(a\alpha)}}\right)} \right)^2 \quad \text{for } a > 2,$$

$$E[\tau] = \frac{\Phi_{a+2}\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha(a+2))}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{a}{b},$$

$$V[\tau] = \frac{\Phi_{a+4}\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha(a+4))}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{a(a+2)}{b^2} - (E[\tau])^2,$$

$$E[\eta] = \frac{\Phi_{a-1}\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha(a-1))}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{\Gamma\left(\frac{a-1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \sqrt{\left(\frac{b}{2}\right)} \quad \text{for } a > 1,$$

$$V[\eta] = \frac{\Phi_{a-2}\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha(a-2))}}\right)}{\Phi_a\left(\frac{\xi_0 - m}{\sqrt{b/(\alpha a)}}\right)} \frac{b}{a-2} - (E[\eta])^2 \quad \text{for } a > 2.$$

Proof

The moments of ξ follow from Theorem 4 in Appendix 1, noting that $\xi = m + \sqrt{\frac{b}{a\alpha}}\kappa$, where κ has a right-truncated t distribution with parameters $\frac{\xi_0 - m}{\sqrt{\frac{b}{a\alpha}}}$ and a. The moments of τ follow from Theorem 5 in Appendix 2, writing $\gamma = \sqrt{\alpha}(\xi_0 - m)$. Similarly, the moments of η can be derived directly from Theorem 5 by noting that $\eta = \tau^{-\frac{1}{2}}$.

$\diamond$

Recalling that the normal-gamma density is conjugate to the normal likelihood, it can be seen immediately that the RTNG distribution is conjugate to the half-normal likelihood. Thus we have the following theorem.

Theorem 2 *If* $X|\xi,\tau \sim \mathcal{HN}\left(\xi, \frac{1}{\tau}\right)$ *and* $\xi,\tau \sim \mathcal{RTNG}(\xi_0, m, \alpha, a, b)$ *then:*

1. *The marginal density of* X *is*

$$f(x) = 2\sqrt{\frac{a\alpha}{b(\alpha+1)}} \frac{\Phi_{a+1}\left(\frac{\min\{\xi_0,x\}-\frac{\alpha m+x}{\alpha+1}}{\sqrt{\frac{b+\frac{\alpha}{\alpha+1}(x-m)^2}{(\alpha+1)(a+1)}}}\right)}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(\alpha a)}}\right)} \phi_a\left(\frac{x-m}{\sqrt{\frac{b(\alpha+1)}{\alpha a}}}\right).$$

The mean of this density is

$$E[X] = m - \sqrt{\frac{b}{a\alpha}} \frac{a + \left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)^2}{a-1} \frac{\phi_a\left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{b/(a\alpha)}}\right)} + \sqrt{\frac{2}{\pi}} \frac{\Phi_{a-1}\left(\frac{\xi_0-m}{\sqrt{\frac{b}{(a-1)\alpha}}}\right)}{\Phi_a\left(\frac{\xi_0-m}{\sqrt{\frac{b}{a\alpha}}}\right)} \frac{\Gamma\left(\frac{a-1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \left(\frac{b}{2}\right)^{\frac{1}{2}}$$

for $a > 1$.

2. *Given a random sample of data,* $\mathbf{x} = (x_1, \ldots, x_n)$*, then the posterior density of* $\xi, \tau|\mathbf{x}$ *is also RTNG,*

$$\xi, \tau|\mathbf{x} \sim \mathcal{RTNG}(\xi^\star, m^\star, \alpha^\star, a^\star, b^\star),$$

where

$$\begin{aligned}
\xi^\star &= \min\{x_{(1)}, \xi_0\}, \\
m^\star &= \frac{\alpha m + n\bar{x}}{\alpha + n}, \\
\alpha^\star &= \alpha + n, \\
a^\star &= a + n, \\
b^\star &= b + (n-1)s^2 + \frac{\alpha n}{\alpha + n}(m - \bar{x})^2.
\end{aligned}$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$.

Proof

1. The formula for the predictive density follows by writing

$$f(x) = \int\int f(x|\xi,\tau) f(\xi,\tau)\, d\xi\, d\tau,$$

and noting that the integrand is proportional to a RTNG density, that is

$$\mathcal{RTNG}\left(\min\{\xi_0, x\}, \frac{\alpha m + 1}{\alpha + 1}, a+1, b + \frac{\alpha}{\alpha+1}(m-x)^2\right),$$

and then equating terms. The formula for the predictive mean follows by noting that

$$E[X] = E[E[X|\xi,\tau]] = E\left[\xi + \sqrt{\frac{2}{\pi\tau}}\right],$$

and then using the results of Theorems 1 and 5 (in Appendix 2).

2. This follows immediately from the usual normal-gamma updating formulae; see e.g. Box and Tiao (1992).

$\diamond$

Often, prior information may not be available and in such cases a non-informative prior should be used. As ξ represents a location parameter and τ is a scale parameter, then the natural non-informative prior is given by $f(\xi, \tau) \propto 1/\tau$. The posterior distribution given this prior is identified in Theorem 3.

Theorem 3 *The joint posterior distribution of ξ, τ is*

$$\xi, \tau|\mathbf{x} \sim \mathcal{RTNG}(x_{(1)}, \bar{x}, n, n-1, (n-1)s^2).$$

Proof

The form of the distribution follows immediately on multiplying the likelihood and prior, and the parameters follow from Theorem 1.

$\diamond$

Henceforth, we shall consider Bayesian inference for the half-normal distribution using the improper prior defined above.

For the untruncated normal distribution, the maximum likelihood estimators and confidence intervals for the mean and variance coincide with the Bayesian estimators and highest posterior density intervals when the non-informative prior distribution described above is used, although this is no longer true when Bayesian credible and frequentist confidence regions for both parameters are considered. However, for the half-normal distribution, the classical and Bayesian estimators and intervals are clearly different. In the following section, the properties of Bayesian point and interval estimates are compared with those of their bias-corrected, likelihood based counterparts described in Section 2.

4 Simulation based comparison of estimators

In this section we present the results from a simulation experiment designed to compare the sampling properties of bias-corrected likelihood based and Bayesian point and interval estimators of ξ and η. Specifically, we investigate the bias and root mean square error (RMSE) of the point estimators, and the coverage and length of the interval estimators.

Thus, we undertook a Monte Carlo study in which, for various values of n, 100,000 samples of size n were simulated from a half-normal distribution with parameters $\xi = 0$ and $\eta = 1$.

Table 1 presents the estimated biases and root mean squared errors of the uncorrected maximum likelihood estimator $X_{(1)}$, the bias corrected estimator $\hat{\xi}_{BC}$ and the Bayesian posterior mean, $E[\xi|\mathbf{X}]$.

From Table 1, it can be seen that the bias of the posterior mean estimator is consistently lower than that of the maximum likelihood estimator $X_{(1)}$ but

Table 1: Estimated biases and RMSE's of the uncorrected and bias corrected maximum likelihood estimators and the Bayesian posterior mean estimator.

	$X_{(1)}$		$\hat{\xi}_{BC}$		$E[\xi\|\mathbf{X}]$	
n	bias	RMSE	bias	RMSE	bias	RMSE
5	0.2156	0.2882	-0.0031	0.2229	-0.0628	0.2440
10	0.1153	0.1575	-0.0015	0.1157	-0.0111	0.1192
20	0.0591	0.0823	-0.0011	0.0591	-0.0029	0.0597
50	0.0246	0.0345	-0.0001	0.0245	-0.0005	0.0245
100	0.0125	0.0176	3×10^{-5}	0.0125	-0.0002	0.0125
1000	0.0012	0.0018	9×10^{-7}	0.0012	2×10^{-6}	0.0012

slightly higher than that of the bias corrected estimator. The bias corrected estimator has lower RMSE for small sample sizes ($n \leq 20$) and the Bayesian and bias corrected estimators have lower RMSE's for larger sample sizes.

Table 2 gives the estimated coverage probabilities and lengths of the nominally 95% two sided confidence interval as given in Equation 3, the alternative, one sided interval given in Equation 4 and the Bayesian 95% highest posterior density interval for ξ.

Table 2: Estimated coverage probabilities and lengths of likelihood based and Bayesian nominally 95% intervals for ξ.

	Frequentist intervals				Bayesian interval	
	2 sided		1 sided			
n	coverage	length	coverage	length	coverage	length
5	0.911	0.801	0.902	0.655	0.951	0.910
10	0.935	0.428	0.931	0.349	0.950	0.398
20	0.944	0.222	0.943	0.181	0.951	0.191
50	0.948	0.091	0.947	0.074	0.950	0.076
100	0.948	0.046	0.947	0.037	0.950	0.038
1000	0.951	0.005	0.951	0.004	0.949	0.004

It can be seen from Table 2 that the coverage of the highest posterior density interval for ξ is identical to, or very close to, the nominal coverage of 0.95, even for small samples. The two sided interval has slightly worse coverage in small samples and is, in general somewhat wider. The coverage of the one sided interval is poorer although this interval is slightly narrower than the Bayesian interval.

We also considered the bias and coverage properties of estimators of η. Table 3 gives the biases of the two likelihood based estimators of η and the Bayesian posterior mean. One can see that, in general, the posterior mean slightly overestimates the value of η whereas the likelihood based estimators give slight underestimates.

Table 3: Estimated biases and RMSE's of the maximum likelihood, bias-corrected and Bayesian posterior mean estimators of η.

	$\hat{\eta}$		$\hat{\eta}_{BC}$		$E[\eta\|\mathbf{X}]$	
n	bias	RMSE	bias	RMSE	bias	RMSE
5	-0.2278	0.3738	-0.1366	0.3584	0.2019	0.5035
10	-0.1184	0.2483	-0.0707	0.2407	0.0694	0.2737
20	-0.0602	0.1678	-0.0359	0.1646	0.0298	0.1742
50	-0.0240	0.1028	-0.0140	0.1020	0.0114	0.1043
100	-0.0121	0.0718	-0.0071	0.0715	0.0055	0.0722
1000	-0.0012	0.0224	-0.0007	0.0224	0.0005	0.0224

Finally, Table 4 gives the estimated coverages and lengths of the classical, nominally 95%, confidence interval for η, a Bayesian credible interval with a nominal 2.5% in each tail and the highest posterior density interval.

We can see from Table 4 that the coverages of the Bayesian and classical intervals are very close to the nominal values, although generally the Bayesian intervals are slightly wider than the classical interval.

Thus, we can conclude that, Bayesian point and interval estimates of the location and scale parameter of the half-normal distribution have relatively good classical properties comparable with those of maximum likelihood based estimators.

In the following section, we now consider a more general model than the half-normal distribution, that is the half-t distribution.

Table 4: Estimated coverage probabilities and lengths of likelihood based and Bayesian nominally 95% intervals for η.

	Frequentist interval		Credible interval		HPD interval	
n	coverage	length	coverage	length	coverage	length
5	0.943	1.937	0.949	2.321	0.938	2.308
10	0.944	1.057	0.949	1.156	0.940	1.148
20	0.947	0.675	0.949	0.707	0.939	0.701
50	0.948	0.405	0.949	0.417	0.941	0.413
100	0.949	0.282	0.949	0.284	0.947	0.281
1000	0.950	0.088	0.950	0.088	0.950	0.088

5 The half-t distribution

The random variable X has a half-t distribution, $X|\xi,\tau,d \sim \mathcal{HT}_d(\xi, 1/\tau)$, if

$$\begin{aligned} f(x|\xi,\tau,d) &= 2\frac{\Gamma\left(\frac{d+1}{2}\right)\sqrt{\tau}}{\Gamma\left(\frac{d}{2}\right)\sqrt{d\pi}}\left[1+\frac{1}{d}\left(\sqrt{\tau}(x-\xi)\right)^2\right]^{-\frac{d+1}{2}} \\ &= 2\sqrt{\tau}\phi_d\left(\sqrt{\tau}(x-\xi)\right) \quad \text{for } x>\xi, -\infty<\xi<\infty, \tau>0, d>0. \end{aligned}$$

As $d \to 0$, the right-hand tail of the distribution becomes increasingly heavier relative to that of the limiting half-normal distribution, obtained as $d \to \infty$.

Little work on inference for the half-t distribution appears to have been published, but see Tancredi (2002). In addition, there is no simple way of conducting Bayesian inference directly for the half-t distribution. However, latent variables can be introduced so as to define a Gibbs sampling algorithm similar to the one of Geweke (1993) designed to perform Bayesian inference for the t distribution.

Suppose that $X|\xi,\tau,d \sim \mathcal{HT}_d(\xi, 1/\tau)$ and let $\theta_i|d$ be independently gamma distributed, $\mathcal{G}\left(\frac{d}{2},\frac{d}{2}\right)$, for $i = 1,\dots,n$. Then

$$X_i|\xi,\tau,\theta_i \sim \mathcal{HN}\left(\xi, \frac{1}{\tau\theta_i}\right).$$

Suppose now that $\mathbf{x}$ is a random sample of size n from the half-t distribution. Given the non-informative improper prior[1] $f(\xi,\tau) \propto \frac{1}{\tau}$ as earlier, and an independent prior $f(d)$, we have the following conditional posterior distributions

$$\begin{aligned} f(\xi,\tau,\boldsymbol{\theta},d|\mathbf{x}) &\propto \tau^{\frac{n}{2}-1}\exp\left\{-\frac{\tau}{2}\sum_{i=1}^n\theta_i(x_i-\xi)^2\right\}\frac{(d/2)^{\frac{nd}{2}}}{\Gamma(d/2)^n} \\ &\quad \prod_{i=1}^n\theta_i^{\frac{d+1}{2}-1}\exp\left\{-\frac{d}{2}\sum_{i=1}^n\theta_i\right\}f(d), \\ \tau|\mathbf{x},\xi,\boldsymbol{\theta},d &\sim \mathcal{G}\left(\frac{n}{2},\frac{\sum_{i=1}^n\theta_i(x_i-\xi)^2}{2}\right), \\ \xi|\mathbf{x},\tau,\boldsymbol{\theta} &\sim \mathcal{TN}\left(\frac{\sum_{i=1}^n\theta_i x_i}{\sum_{i=1}^n\theta_i},\frac{1}{\tau\sum_{i=1}^n\theta_i}\right), \quad \text{a truncated normal with } \xi<x_{(1)}, \\ \theta_i|\mathbf{x},\xi,\tau,d &\sim \mathcal{G}\left(\frac{d+1}{2},\frac{d+\tau(x_i-\xi)^2}{2}\right), \quad \text{independently for } i=1,\dots,n, \\ f(d|\mathbf{x},\xi,\tau,\boldsymbol{\theta}) &\propto \frac{(d/2)^{\frac{nd}{2}}}{\Gamma(d/2)^n}\prod_{i=1}^n\theta_i^{\frac{d+1}{2}}\exp\left\{-\frac{d}{2}\sum_{i=1}^n\theta_i\right\}f(d), \quad \text{for } d>0. \end{aligned}$$

Note that, as observed by Geweke (1993), it is important to use a proper prior distribution for d because an improper prior essentially implies the half-normal model, due to the fact that the prior mass as $d \to \infty$ always infinitely

[1] It is straightforward to extend the inference to the case where a RTNG prior is used.

exceeds the mass for any finite d. Suppose then that, as suggested by Geweke (1993), we use an exponential prior with parameter β. In this case,

$$f(d|\mathbf{x},\xi,\tau,\boldsymbol{\theta}) \propto \frac{(d/2)^{\frac{nd}{2}}}{\Gamma(d/2)^n} \prod_{i=1}^n \theta_i^{\frac{d+1}{2}} \exp\left\{-\frac{d}{2}\left(\beta + \sum_{i=1}^n \theta_i\right)\right\}.$$

We can now use a Gibbs sampling algorithm to simulate a sample from the joint posterior parameter distribution as follows.

1. $t = 0$. `Select initial values` $\xi^{(0)}$, $\tau^{(0)}$, $d^{(0)} = 100$.
2. `Sample` $\theta_i^{(t+1)} \sim \theta_i|\mathbf{x},\xi^{(t)},\tau^{(t)},d^{(t)}$ `for` $i = 1,\ldots,n$.
3. `Sample` $\tau^{(t+1)} \sim \tau|\mathbf{x},\xi^{(t)},\boldsymbol{\theta}^{(t+1)}$.
4. `Sample` $\xi^{(t+1)} \sim \xi|\mathbf{x},\tau^{(t+1)},\boldsymbol{\theta}^{(t+1)}$.
5. `Sample` $d^{(t+1)} \sim d|\mathbf{x},\boldsymbol{\theta}^{(t+1)}$.
6. $t = t + 1$. `Go to 2.`

Some aspects of this algorithm deserve commentary. Firstly, the starting values for the algorithm could be chosen, for example, by setting some large initial value for d and then using the posterior modal or mean estimates for ξ and τ calculated assuming the half-normal model. Secondly, in order to sample the truncated normal distribution for ξ, a rejection algorithm developed by Geweke (1991) which uses the fact that the tail of a normal distribution behaves similarly to an exponential distribution can be used. Robert (1995) also provides a simple algorithm. Thus, the only part of this procedure that is problematic is the sampling of d. One simple method is to use a Metropolis Hastings step with a candidate distribution having mean the current value of d, for example sampling $\tilde{d} \sim \mathcal{G}\left(k, \frac{k}{d}\right)$ for some suitably chosen value of k. Then the candidate is accepted with probability

$$\min\left\{\frac{f(\tilde{d}|\mathbf{x},\xi,\tau,\theta)}{f(d|\mathbf{x},\xi,\tau,\theta)} \frac{g(d,\tilde{d})}{g(\tilde{d},d)}\right\},$$

where $g(\tilde{d},d)$ is the generating density.

6 Model Selection

From a Bayesian perspective, an informal way of comparing the half-normal and half-t models would be to examine the posterior distribution of the degrees of freedom parameter d given the half-t model. If most of the mass of the posterior distribution of d is centred on large values of d, this would provide evidence in favour of the half-normal model.

A more formal approach is to use Bayes factors, see e.g. Jeffreys (1961) or Kass and Raftery (1995). Given the data, $\mathbf{x}$, the Bayes factor for comparing two models $\mathcal{M}_1$ (here the half-normal model) and $\mathcal{M}_2$ (half-t) with prior probabilities $P(\mathcal{M}_1)$ and $P(\mathcal{M}_2)$ is defined as

$$\begin{aligned} B_{12} &= \frac{P(\mathcal{M}_1|\mathbf{x})}{P(\mathcal{M}_2|\mathbf{x})}\frac{P(\mathcal{M}_2)}{P(\mathcal{M}_1)} \\ &= \frac{f(\mathbf{x}|\mathcal{M}_1)}{f(\mathbf{x}|\mathcal{M}_2)}. \end{aligned}$$

That is, the Bayes factor is the ratio of the predictive densities of the data under the two models.

Usually, when improper prior distributions are used under the different models, the Bayes factor does not exist. However, in our case, the parameter space and prior distribution for ξ, τ ($f(\xi,\tau) \propto 1/\tau$) is the same under both models which implies that the Bayes factor is well defined as long as the prior distribution for d under the half-t model is proper.

In order to calculate the Bayes factor, note first that under the half-normal model, the marginal density of the data (up to the integrating constant of the prior) can be calculated analytically. Thus, for the data, $\mathbf{x}$, we have

$$\begin{aligned} f(\mathbf{x}|\mathcal{M}_1) &\propto \int_{-\infty}^{x_{(1)}}\int_0^\infty \frac{1}{\tau} f(\mathbf{x}|\xi,\tau,\mathcal{M}_1)\, d\tau\, d\xi \\ &= \frac{2\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{n}}\left(\frac{2}{\sqrt{\pi(n-1)}s}\right)^{n-1}\Phi_{n-1}\left(\frac{\xi_0-\bar{x}}{s/\sqrt{n}}\right). \end{aligned}$$

For the half-t model, we can use an algorithm of Chib (1995) and Chib and Jeliazkov (2001) to calculate an approximate marginal likelihood from the output of the Gibbs sampler. For given values of the parameters d, ξ, τ, for example the posterior modes, we have, from Bayes Theorem,

$$\log f(\mathbf{x}|\mathcal{M}_2) = \log f(\mathbf{x}|d,\xi,\tau,\mathcal{M}_2) - \log\tau + \log f(d|\mathcal{M}_2) - \log f(d,\xi,\tau|\mathbf{x},\mathcal{M}_2),$$

where the first term on the right hand side is the log-likelihood, the next two terms represent the log-prior and the last term is the log-posterior. The log-likelihood and log-prior can be directly evaluated for given values of d, ξ, τ, and the log-posterior can be expressed as

$$\log f(d,\xi,\tau|\mathbf{x},\mathcal{M}_2) = \log f(\xi|\mathbf{x},\mathcal{M}_2) + \log f(\tau|\mathbf{x},\xi,\mathcal{M}_2) + \log f(d|\mathbf{x},\xi,\tau,\mathcal{M}_2).$$

Now, $\log f(\xi|\mathbf{x},\mathcal{M}_2)$ can be directly estimated from the Gibbs sampler output and $\log f(\tau|\mathbf{x},\xi,\mathcal{M}_2)$ can be estimated by fixing ξ, running the Gibbs sampler for a further set of iterations and applying the algorithm of Chib (1995). $\log f(d|\mathbf{x},\xi,\tau,\mathcal{M}_2)$ can be estimated either directly by simple, unidimensional numerical integration, or by running the sampler through a further set of iterations with ξ and τ fixed and using the method of Chib and Jeliazkov (2001).

7 Example

In this section, we re-analyze the data considered in Pewsey (2002, 2004). These data consist of the body fat measurements of 102 elite Australian athletes.

First of all, we fitted a half-normal distribution model to the data using both bias-corrected likelihood based and Bayesian techniques. In this case, $\hat{\xi} = 5.63$ and the Bayesian posterior mean is 5.57. The latter is identical to $\hat{\xi}_{BC}$. The bias-corrected and Bayesian 95% intervals for ξ are $[5.41, 5.63)$ and $[5.44, 5.63)$, respectively.

When joint credible and confidence regions for ξ and η are calculated, we can see further differences between the classical and Bayesian results. Figure 1 illustrates a classical 95% confidence region as given in Pewsey (2004) and Bayesian 50%, 95% and 99% contours along with the posterior mean and mode of (ξ, η). We can see that the area of the Bayesian 95% region is somewhat smaller than that of the classical region.

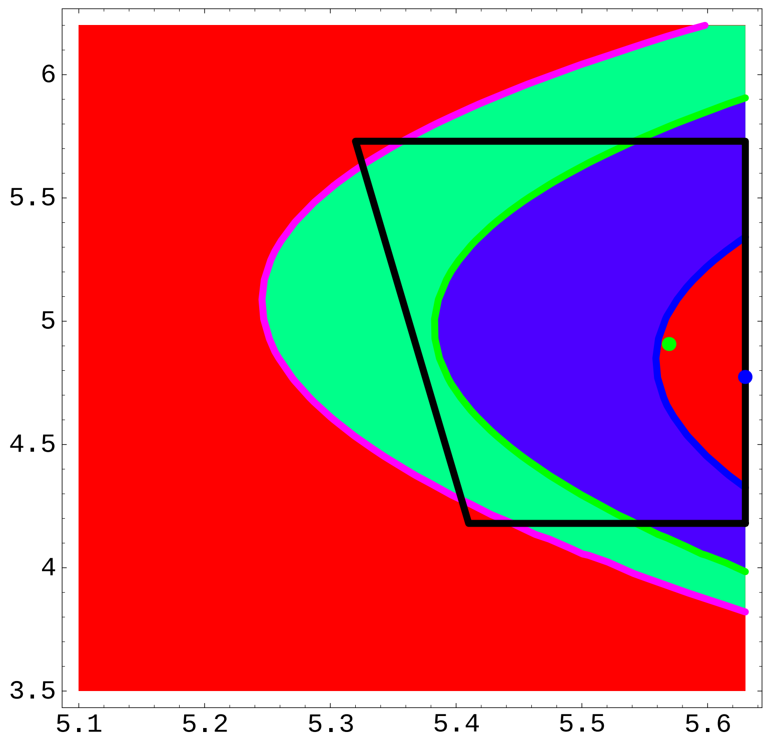

Figure 1: A classical 95% confidence region (rectilinear region) and Bayesian 50%, 95% and 99% credible regions (parabolic regions) for (ξ, η) together with the posterior mean (grey dot) and mode (black dot) for (ξ, η).

The $\mathcal{HN}(\hat{\xi}_{BC}, \hat{\eta}_{BC})$ and Bayesian predictive distribution functions are compared in Figure 2. As might be expected given a sample size as large as 102, the fitted distribution functions are almost identical. However, this plot also suggests that the half-normal model does not provide a particularly good fit to the data.

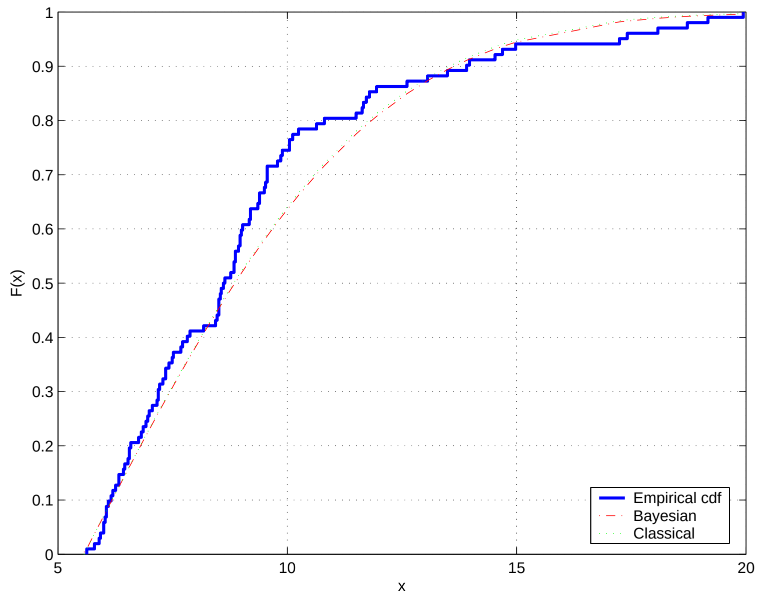

Figure 2: Empirical and fitted cumulative distribution functions.

In an attempt to more adequately describe the distribution of the data, we considered the fit of the half-t model. The prior for d was assumed to be exponential with mean $E[d] = 20$, and we ran a Gibbs sampler for 100,000 iterations (plus a burn in of 10,000 iterations) with starting values for ξ, τ close to posterior means under the half-normal model and a large initial value for d.

In Figure 3, we present a relative frequency histogram of the values of d generated by the Gibbs sampler together with a superimposed kernel density estimate of the posterior distribution. Clearly, the distribution of d is very heavy-tailed, although the posterior probability of d exceeding 50 is somewhat less than 0.05. Moreover, the posterior mean of d is around 10.

In order to compare the two models, the Bayes factor was calculated. The Bayes factor in favour of the half-normal model was estimated to be $B_{12} \approx 1/4$. According to the scale of evidence of Kass and Raftery (1995), this corresponds to positive evidence in favour of the half-t model, and suggests that the half-t model should be preferred to the half-normal.

Finally, we explored the effects of fitting the half-t model using different values of the prior mean for d. In Figure 4, the predictive cdf's under a variety of different prior means are given. We can see that the predictive cdf under the half-normal model ($E[d] = \infty$) is markedly different from the predictive

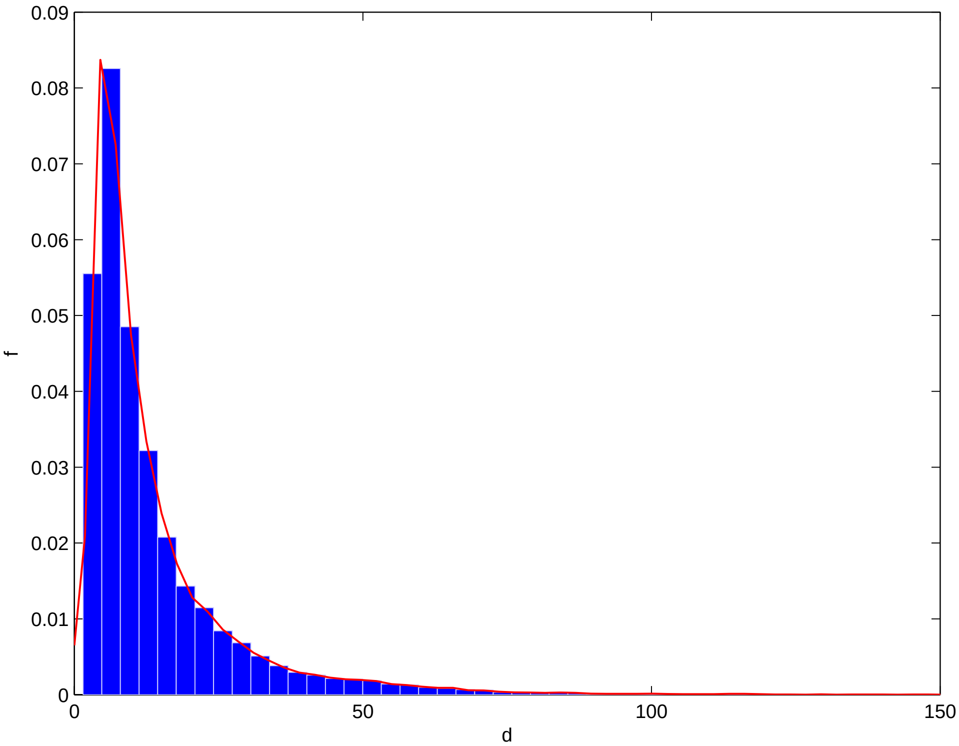

Figure 3: Relative frequency histogram of the posterior density of d together with a superimposed kernel density estimate.

cdf's under the half-t model, these latter cdf's being very similar for values of $E[d]$ ranging between 20 and 50. Furthermore, 95% highest posterior density intervals for ξ were calculated for each of the priors. These intervals varied between $[5.47, 5.63)$ (for $E[d] = 20$) and $[5.46, 5.63)$ (for $E[d] = 50$), in each case being marginally narrower than the limiting interval $[5.44, 5.63)$ under the half-normal model. Thus, at least in terms of the estimation of ξ, there appears to be relatively little sensitivity to the choice of prior mean for d.

8 Discussion

In this paper we have illustrated how Bayesian inference can be carried out for the half-normal and half-t distributions. In particular, we have shown how conjugate inference can be implemented for the half-normal model and, further, that Bayesian methods assuming the usual non-informative or objective prior perform well in terms of frequentist properties such as bias and coverage.

A number of extensions to the work presented here are of interest. For instance, two generalizations would be to consider Bayesian inference for the

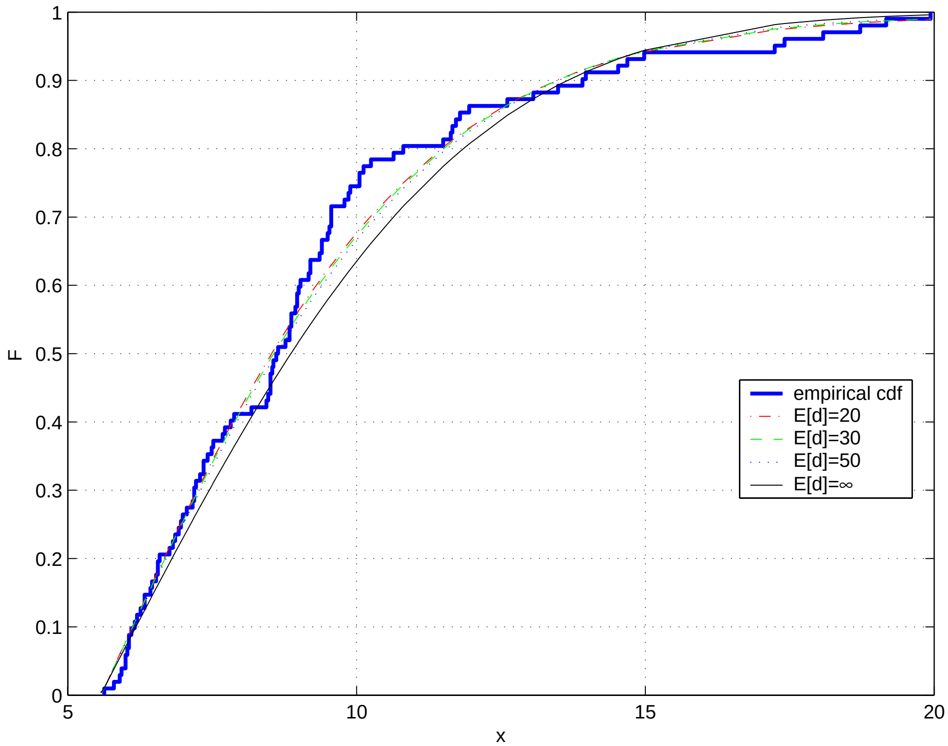

Figure 4: Empirical distribution function of the body fat data together with predictive half-t distribution functions for various prior mean values of d.

truncated normal and folded normal distributions. Interest in the latter distribution dates back to the work of Elandt (1961). In addition, the half-normal and half-t distributions are limiting cases of the skew-normal (Azzalini, 1985) and skew-t (Mukhopadhyay and Vidakovic, 1995; Jones and Faddy, 2003; Azzalini and Capitanio, 2003) distributions, respectively. Bayesian inference for the skew-normal distribution, as well as for the more general context of skew-elliptical distributions, is considered by Liseo (2004). Given the focus of this paper, it is of interest to note that an extension of the skew-normal distribution was originally proposed by O'Hagan and Leonard (1976) as a potential skew prior when there is uncertainty about an inequality constraint in the Bayesian estimation of the mean of the normal distribution.

Finally, whilst we have informally explored sensitivity to the choice of prior for the degrees of freedom of the half-t distribution, it would also be possible to undertake a more formal sensitivity analysis, see e.g. Berger (1994) or Ríos Insua and Ruggeri (2000) for a review of this area.

Acknowledgements

Michael Wiper was partially supported in this research by the project SEJ2004-03303 and Javier Girón by project SEJ2004-02447.

References

1. Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production models. *Journal of Econometrics*, **6**, 21–37.

2. Azzalini, A., 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

3. Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of Royal Statistical Society, Series B*, **65**, 367–389.

4. Berger, J., 1994. An overview of robust Bayesian analysis. *Test*, **3**, 5–124.

5. Box, G., Tiao, G., 1992. *Bayesian Inference in Statistical Analysis*. New York: Wiley.

6. Chib, S., 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90**, 1313–1321.

7. Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis Hastings output. *Journal of the American Statistical Association*, **96**, 270–281.

8. Dobzhansky, T., Wright, S., 1947. Genetics of natural populations X. Dispersal rates in *drosophila pseudoobscura*. *Genetics*, **28**, 304–340.

9. Elandt, R.C., 1961. The folded normal distribution: two methods for estimating parameters from moments. *Technometrics*, **3**, 551–562.

10. Geweke, J., 1991. Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints. In E. M. Keramidas (ed.), *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, 571–578. Fairfax: Interface Foundation of North America, Inc.

11. Geweke, J., 1993. Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, **8 Supplement**, 19–40.

12. Haberle, J.G., 1991. Strength and failure mechanisms of unidirectional carbon fibre-reinforced plastics under axial compression. Unpublished Ph.D. thesis, Imperial College, London, U.K.

13. Jeffreys, H., 1961). *Theory of Probability* (3rd ed.). Oxford: University Press.

14. Jones, M.C, Faddy, M.J., 2003. A skew extension of the t distribution, with applications. *Journal of Royal Statistical Society, Series B*, **65**, 159–174.

15. Johnson, N., Kotz, S., Balakrishnan, N., 1994. *Continuous Univariate Distributions*, Vol. 1 (2nd ed.). New York: Wiley.

16. Kass, R., Raftery, A., 1995. Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

17. Liseo, B., 2004. Skew-elliptical Distributions in Bayesian Inference. In M. G. Genton (ed.), *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*, 153–171. Boca Raton, Florida: Chapman & Hall/CRC.

18. Meeusen, W.J., van den Broeck, J., 1977. Efficiency estimation from Cobb Douglas production functions with composed error. *International Economic Review*, **8**, 435–444.

19. Mukhopadhyay, S., Vidakovic, B., 1995. Efficiency of linear Bayes rules for a normal mean: skewed prior class. *The Statistician*, **44**, 389–397.

20. O'Hagan, A., Leonard, T., 1976. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **63**, 201–203.

21. Pewsey, A., 2002. Large-sample inference for the general half-normal distribution. *Communications in Statistics – Theory and Methods*, **31**, 1045–1054.

22. Pewsey, A., 2004. Improved likelihood based inference for the general half-normal distribution. *Communications in Statistics – Theory and Methods*, **33**, 197–204.

23. Ríos Insua, D., Ruggeri, F., 2000. *Robust Bayesian Analysis.* New York: Springer Verlag.

24. Robert, C.P., 1995. Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121–125.

25. Tancredi, A., 2002. Accounting for heavy tails in stochastic frontier models. Working Paper n. 2002.16, Dipartimento di Scienze Statistiche, Università di Padova, Italy.

Appendix 1: The right-truncated Student's t distribution

In this appendix, we outline the properties of the marginal distribution of ξ. As noted in Section 3, if $\xi, \tau \sim \mathcal{RTNG}(\xi_0, m, \alpha, a, b)$, the marginal distribution of $\kappa = \frac{\xi - m}{\sqrt{b/(a\alpha)}}$ is a standard t distribution with a degrees of freedom, truncated for $\xi < \xi_0$. In this subsection, we define this distribution and give its properties.

Definition 2 *We say that a random variable T has a right-truncated t distribution with $a > 0$ degrees of freedom and truncation parameter c where $-\infty < c < \infty$, i.e. $T|a, c \sim t_a^-(c)$, if*

$$f(t|a,c) = \frac{1}{\Phi_a(c)} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \frac{1}{\sqrt{a\pi}} \left(1 + \frac{t^2}{a}\right)^{-\frac{a+1}{2}} = \frac{\phi_a(t)}{\Phi_a(c)}, \quad \textit{for } t < c.$$

The following theorem gives the odd and even moments of the right-truncated t distribution.

Theorem 4 *If $T|a, c \sim t_a^-(c)$ then for $s = 0, 1, 2, \ldots$ the odd moments of T are given by*

$$E[T^{2s+1}|a,c] = -(a + c^2) \frac{\phi_a(c)}{\Phi_a(c)} \sum_{i=0}^{s} c^{2(s-i)} a^i \frac{1}{2(s+1)} \prod_{j=0}^{i} \frac{2(s+1-j)}{(a - 2(s-j) - 1)},$$

for $s < \frac{a-1}{2}$.

Also, for $s = 1, 2, \ldots,$ the even moments of T are given by

$$E[T^{2s}|a,c] = a^s \prod_{i=1}^{s} \frac{2i-1}{a-2i} - (a+c^2) \frac{\phi_a(c)}{\Phi_a(c)} \sum_{i=0}^{s-1} c^{2(s-i)-1} a^i \frac{1}{2s+1} \prod_{j=0}^{i} \frac{2(s-j)+1}{a - 2(s-j)},$$

for $s < \frac{a}{2}$.

Proof

In the following, we drop the dependence on a and c throughout. Firstly, we calculate the formulae for $E[T]$ and $E[T^2]$ directly.

$$\begin{aligned}
E[T] &= \int_{-\infty}^{c} t \frac{\phi_a(t)}{\Phi_a(c)} \, dt \\
&= \frac{1}{\Phi_a(c)} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \frac{1}{\sqrt{a\pi}} \int_{-\infty}^{c} t \left(1 + \frac{t^2}{a}\right)^{-\frac{a+1}{2}} \, dt \\
&= \frac{1}{\Phi_a(c)} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \frac{1}{\sqrt{a\pi}} \left[-\frac{a}{a-1} \left(1 + \frac{t^2}{a}\right)^{-\frac{a-1}{2}} \right]_{-\infty}^{c} \quad \text{if } a > 1
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{a}{a-1}\frac{1}{\Phi_a(c)}\left[\left(1+\frac{t^2}{a}\right)\phi_a(t)\right]_{-\infty}^{c} \\
&= -\frac{a+c^2}{a-1}\frac{\phi_a(c)}{\Phi_a(c)}, \quad \text{which is the odd moment formula for } s=0. \\
E[T^2] &= \int_{-\infty}^{c} t\times t\frac{\phi_a(t)}{\Phi_a(c)}\,dt \quad \text{and integrating by parts, we find, for } a>2, \\
&= -\frac{a}{a-1}\frac{1}{\Phi_a(c)}\left[t\left(1+\frac{t^2}{a}\right)\phi_a(t)\right]_{-\infty}^{c} + \\
&\quad \frac{a}{a-1}\frac{1}{\Phi_a(c)}\int_{-\infty}^{c}\left(1+\frac{t^2}{a}\right)\phi_a(t)\,dt \\
\Rightarrow E[T^2] &= \frac{a}{a-1} - \frac{a}{a-1}\frac{\phi_a(c)}{\Phi_a(c)}c\left(1+\frac{c^2}{a}\right) + \frac{1}{a-1}E[T^2] \\
\Rightarrow E[T^2] &= \frac{a}{a-2} - \frac{c(a+c^2)}{a-2}\frac{\phi_a(c)}{\Phi_a(c)},
\end{aligned}
$$

which is the even moment formula corresponding to $s = 1$. More generally, for any $r < a$, we have:

$$
\begin{aligned}
E[T^r] &= \int_{-\infty}^{c} t^{r-1}\times t\frac{\phi_a(t)}{\Phi_a(c)}\,dt \\
&= -\frac{a}{a-1}\frac{1}{\Phi_a(c)}\left[t^{r-1}\left(1+\frac{t^2}{a}\right)\phi_a(t)\right]_{-\infty}^{c} + \\
&\quad \frac{a}{a-1}\frac{1}{\Phi_a(c)}\int_{-\infty}^{c}(r-1)t^{r-2}\left(1+\frac{t^2}{a}\right)\phi_a(t)\,dt \\
&= -\frac{c^{r-1}(a+c^2)}{a-1}\frac{\phi_a(c)}{\Phi_a(c)} + \frac{a(r-1)}{a-1}E[T^{r-2}] + \frac{r-1}{a-1}E[T^r] \\
\Rightarrow E[T^r] &= -\frac{c^{r-1}(a+c^2)}{a-r}\frac{\phi_a(c)}{\Phi_a(c)} + \frac{a(r-1)}{a-r}E[T^{r-2}].
\end{aligned}
$$

Now we can use induction to obtain the result. In the case of the odd moments, assuming the formula given in the theorem is correct for $k = 0, \ldots, s-1$, we have

$$
\begin{aligned}
E[T^{2s+1}] &= -\frac{c^{2s+1-1}(a+c^2)}{a-2s-1}\frac{\phi_a(c)}{\Phi_a(c)} + \frac{a(2s)}{a-2s-1}E[T^{2s-1}] \\
&= -\frac{c^{2s}(a+c^2)}{a-2s-1}\frac{\phi_a(c)}{\Phi_a(c)} - \frac{a(2s)}{a-2s-1}(a+c^2)\frac{\phi_a(c)}{\Phi_a(c)}\times \\
&\quad \sum_{i=0}^{s-1} c^{2(s-1-i)}a^i\frac{1}{2(s-1+1)}\prod_{j=0}^{i}\frac{2(s-1+1-j)}{a-2(s-1-j)-1} \\
&= -(a+c^2)\frac{\phi_a(c)}{\Phi_a(c)}\left\{\frac{c^{2s}}{a-2s-1}+\right.
\end{aligned}
$$

$$
\left. \frac{a}{a-2s+1}\sum_{i=0}^{s-1} c^{2(s-i)}(a)^i \prod_{j=0}^{i} \frac{2(s-j)}{(a-2(s-j)-1)} \right\}
$$

$$
= \quad -(a+c^2)\frac{\phi_a(c)}{\Phi_a(c)}\sum_{i=0}^{s} c^{2(s-i)} a^i \frac{1}{2(s+1)} \prod_{j=0}^{i} \frac{2(s+1-j)}{(a-2(s-j)-1)},
$$

which proves the result.

Now, for the even moments, assuming the general formula is valid for $k = 1, \ldots, s-1$, we have

$$
\begin{aligned}
E[T^{2s}] &= -\frac{c^{2s-1}(a+c^2)}{a-2s-1}\frac{\phi_a(c)}{\Phi_a(c)} + \frac{a(2s-1)}{a-2s}E[T^{2s-2)}] \\
&= -\frac{c^{2s-1}(a+c^2)}{a-2s-1}\frac{\phi_a(c)}{\Phi_a(c)} + \frac{a(2s-1)}{a-2s} \times \left\{ a^{s-1}\prod_{i=1}^{s-1}\frac{2i-1}{a-2i} - \right. \\
&\quad \left. (a+c^2)\frac{\phi_a(c)}{\Phi_a(c)}\sum_{i=0}^{s-1} c^{2(s-1-i)-1} a^i \frac{1}{2(s-1)+1}\prod_{j=0}^{i}\frac{2(s-1-j)+1}{a-2(s-1-j)} \right\} \\
&= a^s \prod_{i=1}^{s}\frac{2i-1}{a-2i} - (a+c^2)\frac{\phi_a(c)}{\Phi_a(c)}\left\{ \frac{c^{2s-1}}{a-2s-1} \right. \\
&\quad \left. + \frac{a}{a-2s}\sum_{i=0}^{s-1} c^{2(s-1-i)-1} a^i \prod_{j=0}^{i}\frac{2(s-1-j)+1}{a-2(s-1-j)} \right\} \\
&= a^s \prod_{i=1}^{s}\frac{2i-1}{a-2i} - (a+c^2)\frac{\phi_a(c)}{\Phi_a(c)}\sum_{i=0}^{s-1} c^{2(s-i)-1} a^i \frac{1}{2s+1}\prod_{j=0}^{i}\frac{2(s-j)+1}{a-2(s-j)}.
\end{aligned}
$$

◇

Appendix 2: The Gaussian-modulated gamma distribution

Here we examine the properties of the distribution of τ.

Definition 3 *We say that a random variable τ has a Gaussian-modulated gamma (GMG) distribution with parameters $-\infty < \gamma < \infty$ and $a, b > 0$ if*

$$
f(\tau) = \frac{\Phi\left(\gamma\sqrt{\tau}\right)}{\Phi_a\left(\gamma\sqrt{\frac{a}{b}}\right)} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \tau^{\frac{a}{2}-1} e^{-\frac{b}{2}\tau}.
$$

In this case, we write $\tau \sim \mathcal{GMG}(\gamma, a, b)$.

The following theorem gives the non-central moments of this distribution.

Theorem 5 *If* $\tau \sim \mathcal{GMG}(\gamma, a, b)$*, then*

$$E\left[\tau^{\frac{c}{2}}\right] = \frac{\Phi_{a+c}\left(\gamma\sqrt{\frac{a+c}{b}}\right)}{\Phi_a\left(\gamma\sqrt{\frac{a}{b}}\right)} \frac{\Gamma\left(\frac{a+c}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \left(\frac{2}{b}\right)^{\frac{c}{2}}, \quad \textit{for } c > -a.$$

Proof

Consider the expression

$$\tau^{\frac{c}{2}} f(\tau) = \frac{1}{\Phi_a\left(\gamma\sqrt{\frac{a}{b}}\right)} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}}}{\Gamma\left(\frac{a}{2}\right)} \Phi\left(\gamma\sqrt{\tau}\right) \tau^{\frac{a+c}{2}-1} e^{-\frac{b}{2}\tau}.$$

Ignoring constant terms, this is proportional to a GMG density with parameters $\gamma, a+c, b$. Equating terms gives

$$E\left[\tau^{\frac{c}{2}}\right] = \frac{\Phi_{a+c}\left(\gamma\sqrt{\frac{a+c}{b}}\right)}{\Phi_a\left(\gamma\sqrt{\frac{a}{b}}\right)} \frac{\left(\frac{b}{2}\right)^{\frac{a}{2}} \Gamma\left(\frac{a+c}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\left(\frac{b}{2}\right)^{\frac{a+c}{2}}},$$

and the result follows immediately.

Note that as the tail behaviour of the density close to zero is the same as that of the gamma density, the integration is finite if and only if $a + c > 0$, that is if $c > -a$.

$\diamond$