Posterior inference on the degrees of freedom parameter in multivariate-t regression models
Chib, S.; Osiewalski, J.; Steel, M.F.J.

Publication date:
1992

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models

by

Siddhartha Chib, Jacek Osiewalski and Mark F.J. Steel


Reprint Series no. 89
CENTER FOR ECONOMIC RESEARCH

Research Staff

Helmut Bester
Eric van Damme

Board

Helmut Bester
Eric van Damme, director
Arie Kapteyn

Scientific Council

Eduard Bomhoff
Willem Buiter
Jacques Drèze
Theo van de Klundert
Simon Kuijpers
Jean-Jacques Laffont
Merton Miller
Stephen Nickell
Pieter Ruys
Jacques Sijben

Erasmus University Rotterdam
Yale University
Université Catholique de Louvain
Tilburg University
Groningen University
Université des Sciences Sociales de Toulouse
University of Chicago
University of Oxford
Tilburg University
Tilburg University

Residential Fellows

Svend Albaek
Pramila Krishnan
Jan Magnus
Eduardo Siandra
Hideo Suehiro

European University Institute
San Francisco State University
Tilburg University
UCLA
Kobe University

Doctoral Students

Roel Beetsma
Hans Bloemen
Sjaak Hurkens
Frank de Jong
Pieter Kop Jansen

Address: Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
Phone: +31 13 663050
Telex: 52426 kub n1
Telefax: +31 13 663066
E-mail: "center@htikub5.bitnet"

ISSN 0924-7874

1992
Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models

by

Siddhartha Chib,
Jacek Osiewalski and
Mark F.J. Steel

Reprinted from Economics Letters,
Vol. 37, No. 4, 1991

Reprint Series
no. 89
Posterior inference on the degrees of freedom parameter in multivariate-t regression models *

Siddhartha Chib
University of Missouri, Columbia, MO 65211, USA

Jacek Osiewalski
Academy of Economics, Kraków, Poland

Mark F.J. Steel
Tilburg University, 5000 LE Tilburg, The Netherlands

Received 15 April 1991
Accepted 3 October 1991

This paper considers the nonlinear regression model with errors that follow the multivariate Student-t distribution with \( \nu \) degrees of freedom. We specify general conditions on the overall prior structure under which the prior of \( \nu \) is not updated by the sample information, and provide an example in which learning about \( \nu \) is not precluded.

1. Introduction

An interesting topic in regression analysis is the effect of non-Normal error distributions. In particular, distributions with heavier tails, like the Student \( t \), provide attractive alternatives. The seminal work of Zellner (1976) examines the consequences of adopting a multivariate Student \( t \) (WVt) distribution instead of the usual multivariate Normal (MVN). Zellner (1976) shows that under some restrictions on the prior, the marginal posterior of the regression parameter, \( \beta \), is unaffected by the MVt assumption. This result was generalized further by Chib et al. (1988), Osiewalski (1991), and Osiewalski and Steel (1990).

Most of the papers that have adopted the MVt framework have made one important assumption, i.e., the degrees of freedom, \( \nu \), of the MVt error distribution is known. If this assumption is relaxed, the maximum likelihood method cannot be used to estimate \( \nu \) [cf. Zellner (1976, footnote 6)]. However, a method of moments estimator does exist, and has been provided by Singh (1988).

* This paper was written while the second author was a visiting fellow at CORE, Louvain-la-Neuve, and the third author acknowledges a senior research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).
Especially since the MVN is a limiting case of the MVt (for \( v \to \infty \)), it seems particularly interesting to let the data speak for themselves and indicate how far their distribution deviates from Normality. In this paper we pursue the unknown \( v \) from an entirely Bayesian point of view and focus on the prior to posterior mapping for \( v \). However, due to the fact that in the multivariate framework the entire sample is but one realization from the data density, the observations are not very informative on the properties of the underlying sampling density. We find that prior beliefs on \( v \) are sometimes not updated at all. This outcome is shown to hold under certain proper priors and under the 'usual' Jeffreys' type improper prior on the scalar precision parameter \( \tau \). The main results are derived under fairly weak restrictions for a general nonlinear model with an unknown covariance matrix.

Finally, some comments about notation. The symbol \( p \) is used generically to denote density functions, whether proper or not. Specific types of density functions used in the sequel are \( f_S(w \mid m, C) \) for a MVN density on \( w \in \mathbb{R}^n \) with mean \( m \) and covariance matrix \( C \), \( f_V(w \mid f, m, C) \) denotes a MVt density with degrees of freedom \( f \), location \( m \) and dispersion \( C \). On \( v \in \mathbb{R}_+ \) a gamma density with parameters \( a, b > 0 \) is denoted by \( f_G(v \mid a, b) \), whereas a beta distribution on \( v \in (0, c) \) with \( a, b > 0 \) has the density function \( f_B(v \mid a, b, c) \). Conditional independence of \( g \) and \( \ell \) given \( q \) is denoted by \( g \perp \ell \mid q \).

### 2. The model

Consider the nonlinear regression model in which a \( n \)-vector of observations, \( y \), satisfies

\[
y = h(X, \beta) + \epsilon, \quad p(\epsilon \mid X, \beta, \eta, \tau, \nu) = f_V(\epsilon \mid \nu, 0, \tau^{-1}V(X, \eta)),
\]

where \( X: n \times r \) is a set of regressors, \( \beta \) is the regression coefficient vector, \( h(X, \beta) \) is a vector function of \((X, \beta)\), and \( \epsilon \) is a MVt distributed error vector. More specifically, we make the following assumptions about the quantities in (1):

1. Define \( \omega = (\beta, \nu, \tau, \eta) \) and \( (\beta, \eta) = \vartheta \), where \( \beta \in B \subseteq \mathbb{R}^4 \), and \( \eta \in H \subseteq \mathbb{R}^4 \), \( \vartheta \in \Theta \subseteq B \times H \), \( \tau \in \mathbb{R}_+ \) and \( \nu \in \mathbb{R}_+ \).

2. \( h(X, \beta): n \times 1 \) is a known function of \( X \) and \( \beta \).

3. \( V(X, \eta): n \times n \) is a positive definite matrix, and a known function of \( X \) and \( \eta \).

4. \( X \) is a random matrix such that the joint density \( p(X, \omega) \) factorizes as \( p(X)p(\omega) \), where \( p(\omega) \) is the prior density of the parameters. We also assume \( p(X) \) is a proper (predictive) density.

Note that the degrees of freedom parameter, \( \nu \), need not be integer. The only requirement is that it be strictly positive. The sampling model in (1) is a nonlinear, elliptical version of the model considered by Zellner (1976). We obtain Zellner's model, which we will refer to as the linear spherical model, simply by letting

1. \( r = k \) and \( h(X, \beta) = X\beta \).
and

A3. \( V(X, \eta) = I_n \), so that \( \vartheta = \beta \).

The sampling density resulting from (1) is the following \( t \) density

\[
p(y \mid X, \omega) = \frac{\gamma^{\nu/2}}{\sqrt{\pi \nu}} \left( 1 + \frac{z^2}{\nu} \right)^{-\nu/2},
\]

which can be represented as a scale mixture of Normal densities

\[
p(y \mid X, \omega) = \int_0^\infty p(y \mid X, z, \omega) p(z \mid X, \omega) \, dz,
\]

where

\[
p(y \mid X, z, \omega) = f_\nu^n \left( y \mid h(X, \beta), (z\tau)^{-1}V(X, \eta) \right),
\]

and

\[
p(z \mid X, \omega) = p(z \mid \nu) = f_\nu \left( z \left| \frac{\nu}{2}, \frac{\nu}{2} \right. \right).
\]

Note that the proper mixing density, \( p(z \mid X, \omega) \), is independent of \( X \) and the parameters other than \( \nu \). In order to facilitate the posterior analysis in the following sections, we will make use of the representation (3)–(5), and consider the following joint Bayesian model (given \( X \))

\[
p(y, z, \omega \mid X) = p(y \mid X, z, \omega) p(z \mid \nu) p(\omega).
\]

Although we focus only on MVt models here, it can be shown that the results carry over to general scale mixtures of Normals as used in Jammalamadaka et al. (1987), Chib et al. (1988) and Osiewalski (1991), whereas Osiewalski and Steel (1990) derive the equivalent of Theorem 2 for any member in the entire class of elliptical densities.

3. Prior to posterior analysis: proper priors

The focus will be on the prior to posterior analysis for \( \nu \in \mathbb{R}_+ \) in (1). By Bayes theorem, the posterior of \( \nu \) is

\[
p(\nu \mid y, X) = \int p(\omega \mid y, X) \, d\tau d\vartheta,
\]

where \( p(\omega \mid y, X) = \int p(y, z, \omega \mid X) \, dz / \int p(y, z, \omega \mid X) \, d\omega \, dz \) is the posterior of \( \omega = (\vartheta, \tau, \nu) \) and \( p(y, z, \omega \mid X) \) is the joint Bayesian model given in (6). In this section, we shall provide sufficient conditions under which the posterior of \( \nu \), \( p(\nu \mid y, X) \), is the same as the prior, \( p(\nu) \). We restrict attention to prior of \( \omega \) that are proper. Consider the transformation

\[
(\tau, z) \rightarrow (\phi, z), \quad \phi = \tau z,
\]
with the Jacobian equal to $1/z$. Note that (4) and (7) allow us to conclude that $y$ and $(z, \nu)$ are conditionally independent given $X$, $\theta$ and $\phi$:

$$y \perp (z, \nu) \mid X, \theta, \phi. \tag{8}$$

From (7) and the fact that $X \perp (z, \omega)$ [which results from A.4 and eq. (5)], it ensues that

$$X \perp (z, \theta, \phi, \nu). \tag{9}$$

Using the properties of conditional independence [see e.g. Mouchart and Rolin (1984)] we infer from (8) and (9) that $y \perp \nu \mid X, \theta, \phi$ and $X \perp \nu \mid \theta, \phi$, i.e.

$$\nu \perp (y, X) \mid \theta, \phi. \tag{10}$$

The latter clearly indicates the fundamental result that if $\nu$ and $(\theta, \phi)$ are independent, then $(X, y)$ and $\nu$ are unconditionally independent, and the marginal posterior of $\nu$ is identical to its marginal prior. In other words, the marginalized likelihood

$$p(y \mid X, \nu) = \int p(y, \theta, \phi \mid X, \nu) \, d\phi \, d\theta$$

$$= \int \int \nu(y \mid h(X, \beta), \phi^{-1} \nu(X, \eta)) p(\theta, \phi) \, d\phi \, d\theta, \tag{11}$$

is constant in $\nu$. Thus, in this case the data cannot modify our prior opinions about $\nu$. We summarize this fact in the following theorem which provides a sufficient condition for the impossibility of updating the prior of $\nu$ under proper priors.

**Theorem 1.** Consider (1) under assumptions A.1–A.4 and with proper $p(\omega)$. If

$$\nu \perp (\theta, \phi), \tag{12}$$

where $\phi$ is defined in (7), then

$$\nu \perp (y, X). \tag{13}$$

**Proof.** As we have argued, (10) follows from the hypotheses. However, (10) and (12) are (jointly) equivalent to $\nu \perp (y, X, \theta, \phi)$. This directly leads to the conclusion stated in (13). \qed

In order to illustrate a situation where Theorem 1 holds, i.e., where the prior of $\nu$ is not updated due to independence between $\nu$ and $(\theta, \phi)$, we consider the following example.

**Example 1.** Suppose that we have the proper prior structure

(i) $p(\theta, \tau, \nu) = p(\theta) p(\nu) p(\tau \mid \theta, \nu),$

(ii) $p(\nu) = 0$ if $\nu \leq a,$

(iii) $p(\tau \mid \theta, \nu) f_\theta(\tau \mid \frac{a}{2}, \frac{\nu - a}{2}, \frac{\nu}{d(\theta)}),$
where $a > 0$ and $d(\theta)$ is some known positive function of $\theta$ or any positive constant. Using the transformation $\phi = \tau z$, we obtain

$$p(z | \theta, \phi, \nu) = f_G \left( z - \frac{\phi}{\nu} d(\theta) \frac{\nu - a}{2}, \frac{\nu}{2} \right),$$

which is only nonzero in the support of $\tau$ and implicitly imposes the other prior constraint that $\nu > a$. Furthermore, we get

$$p(\theta, \phi, \nu) = p(\theta) f_G \left( \frac{\phi}{2}, \frac{d(\theta)}{2} \right) p(\nu),$$

so that (12) holds and the prior of $\nu$ will remain unaffected by the data. Indeed, the marginalized likelihood (11) becomes

$$p(y | X, \nu) = p(y | X) = \int f_\nu \left( y | a, h(X, \beta), \frac{d(\theta)}{a} \right) V(X, \eta) p(\theta) \, d\theta.$$

Provided this integral is finite, the results still go through under improper $p(\theta)$. The proper beta density of $\tau$ is what crucially triggers the lack of updating, which is quite particular to (iii). All other proper priors for $\tau$ we have tried fail to produce this result (see e.g. Example 2).

4. Prior to posterior analysis: Improper priors

We now turn to the specially interesting case of the usual improper prior of $\tau$, which is not covered by Theorem 1, since the arguments made there are not necessarily valid for distributions that are not proper.

Theorem 2. Consider (1) under A.1–A.4. Suppose that $p(\omega)$ is a product measure, $p(\omega) = p(\tau)p(\theta)p(\nu)$, and

$$p(\tau) \propto \tau^{-1}, \quad \tau \in \mathbb{R}_+,$$

(14)

where the prior of $\nu$ is proper and functionally independent of $(\tau, \theta)$. If the posterior of $\nu$ exists then $p(\nu | y, X) = p(\nu)$.

Proof. From (5), (7) and (14),

$$p(\phi, z, \theta, \nu) \propto \phi^{-1} p(z | \nu) p(\theta) p(\nu)$$

and the marginal posterior of $\nu$ is

$$p(\nu | y, X) \propto p(\nu) \int_{\theta_0}^{\theta_1} \int_0^\infty f_N(y | h(X, \beta), \phi^{-1} V(X, \eta)) \phi^{-1} d\phi p(z | \nu) \, dz \, p(\theta) \, d\theta,$$

$$\propto p(\nu) \int_{\theta_0}^{\theta_1} p(\theta) |V(X, \eta)|^{-1/2}$$

$$\times \left[ ((y - h(X, \beta))' V(X, \eta)^{-1}(y - h(X, \beta))) \right]^{-n/2} \, d\theta,$$

(15)

so the result is immediate. \( \Box \)
Theorem 2 says that, if the integral in (15) is finite, the improper prior (14) of the precision parameter together with the (functional) prior independence between \( \theta \) and \( \nu \) are sufficient to prevent updating the prior of \( \nu \). The marginalized likelihood \( p(y \mid X, \nu) \), which is then proportional to the integral in (15), is constant in \( \nu \). The result in Theorem 2 holds under a Jeffreys' type prior on \( \tau \). If instead, we are informative on \( \tau \) without using (iii), we can extract some information on \( \nu \) directly from the data, even if we are not informative on the coefficients.

**Example 2.** In the linear spherical model, obtained by A.2' and A.3', we adopt the following improper prior:

\[
\begin{align*}
(iv) \quad p(\theta, \tau, \nu) &= p(\beta) p(\tau) p(\nu), \\
(v) \quad p(\beta) &= c, \quad \beta \in \mathbb{R}^k, \\
(vi) \quad p(\tau) &= \mathcal{G} \left( \frac{a}{2}, \frac{b}{2}, \frac{c}{2} \right),
\end{align*}
\]

where \( a, b, c \) are positive constants. In this example, \( X \) is of full column rank \( k \), and the complete Bayesian model (6) is given by (in terms of \( \phi = \tau z \))

\[
p(y, \beta, \phi, z, \nu \mid X) = c f_{\mathcal{G}}(y \mid X\beta, \phi^{-1} I_n) f_{\mathcal{G}} \left( \phi \mid \frac{a}{2}, \frac{b}{2}, \frac{c}{2} \right) f_{\mathcal{G}} \left( z \mid \frac{\nu}{2}, \tau \right) p(\nu).
\]

Letting \( \hat{\beta} = (X'X)^{-1}X'y \), \( s^2 = (y - X\hat{\beta})' (y - X\hat{\beta}) \), and performing analytical integration w.r.t. \( \beta \) and \( \phi \) we obtain the following formula for the posterior pdf of \( \nu \):

\[
p(\nu \mid y, X) \propto p(\nu) \int_0^{\infty} (b + zs^2)^{-1/2 \chi_2(n-k)} z^{1/2 \chi_2(n-k)} f_{\mathcal{G}} \left( z \mid \frac{\nu}{2}, \frac{\nu}{2} \right) \, dz.
\] (16)

Clearly, there is updating for \( \nu \) now. In the limit case as \( b \to 0 \), the integral in (16) exists if \( \nu > a \), so if we add (ii) we obtain in the limit

\[
p(\nu \mid y, X) \propto p(\nu) \nu^{n/2} \left( \Gamma \left( \frac{\nu - a}{2} \right) / \left( \Gamma \left( \frac{\nu}{2} \right) \right) \right).
\] (17)

If we also let \( a \to 0 \), the prior in (vi) will tend to the Jeffreys' type prior in (14) and updating will be lost, which is also immediate from (17).

5. Conclusion

This paper focuses on a general class of nonlinear MVt error models and discusses Bayesian inference on the degrees of freedom parameter \( \nu \). We have provided sufficient conditions under which the prior of \( \nu \) is not updated by the sample and examined a special case with updating.

References


Osiewalski, J. and M.F.J. Steel, 1990, Robust Bayesian inference in elliptical regression models, Center discussion paper 9032 (Tilburg University, Tilburg).

Singh, R.S., 1988, Estimation of error variance in linear regression models with errors having multivariate Student-\(t\) distributions with unknown degrees of freedom, Economics Letters 27, 47–53.

Reprint Series, CentER, Tilburg University, The Netherlands:


No. 5  Th. ten Raa and F. van der Ploeg, A statistical approach to the problem of negatives in input-output analysis, *Economic Modelling*, vol. 6, no. 1, 1989, pp. 2 - 19.


No. 8  Th. van de Klundert and F. van der Ploeg, Wage rigidity and capital mobility in an optimising model of a small open economy, *De Economist*, vol. 137, nr. 1, 1989, pp. 47 - 75.


<table>
<thead>
<tr>
<th>No. 54</th>
<th>No. 55</th>
<th>No. 56</th>
<th>No. 57</th>
<th>No. 58</th>
<th>No. 59</th>
<th>No. 60</th>
<th>No. 61</th>
<th>No. 62</th>
<th>No. 63</th>
<th>No. 64</th>
<th>No. 65</th>
<th>No. 66</th>
</tr>
</thead>
</table>


