Contents

List of Tables

List of Figures

Bayesian inference for time series regression models with multivariate t autoregressions on errors

Student: Yi-Chuan Yeh Advisors: Dr. Jack C. Lee
Dr. Tsung I. Lin

Institute of Statistics
National Chiao Tung University
Hsinchu, Taiwan

Abstract

This thesis considers a Bayesian approach to the regression model with autoregressive multivariate t errors, whose conditional variance satisfies a kind of generalized autoregressive conditional heteroscedastic model. We present the approximate Bayesian posterior and predictive inferences under a non-informative prior. Markov chain Monte Carlo computational schemes are developed for precisely accounting for the posterior uncertainties. To enhance the computational efficiency, we provide a fast method to compute the inverse autocorrelation matrix of an AR(p) process. A real example of the U.S. interest rates is conducted to demonstrate our methodologies.

Keywords: Approximate inference; MCMC; Predictive distribution; Reparameterization

1. Introduction

There exist plenty of theses in the literature concerning Bayesian techniques on the regression model with autoregressive time series errors. Chib (1993) and McCulloch and Tsay (1994) employ Gibbs sampler methods for regression models with autoregressive Gaussian error process by conditioning on initial observations. Chen et al. (2004) provide an efficient Bayesian estimation procedure with the exact likelihood.

Regression models with the multivariate t distributed error terms have received considerable attentions since Zellner (1976) who investigated the theoretical framework from maximum likelihood and Bayesian viewpoints. The generalization of the Bayesian treatment for regression models with uncorrelated elliptical errors has been discussed by various authors, e.g., Chib et al. (1988), Osiewalski (1991), Arellano-Valle et al. (2000) and Kim and Mallick (2003). Recently, Tarami and Pourahmadi (2003) consider a time series model whose error terms arise from a multivariate t process with AR(p) autocorrelations. They provide exact likelihood equations for computing the maximum likelihood estimates and show that the multivariate t AR model is a generalized autoregressive conditional heteroscedastic model similar to those in Engle (1982) and Bollerslev (1986).

In this thesis, we extend Tarami and Pourahmadi (2003) to consider a regression model with multivariate t autoregressions on errors. Under improper prior distributions, we show a general Bayesian analysis of the model and computational techniques of Markov chain Monte Carlo (MCMC) methods. In order to reduce computing burden for the inverse of the AR(p) autocorrelation matrix, a fast computing program is also given.

In the next section, we present the approximate Bayesian (AB) posterior and predictive inferences for the model. In Section 3, we show how to implement MCMC methods to generate posterior samples and use them to predict future values and volatilities. In Section 4, a real example of the U.S. interest rates is conducted to demonstrate our methodologies. In Section 5 we briefly summarize and discuss future issues and in the Appendix we give the technical derivations of the proposed approach.

2. Approximate Bayesian inference

2.1. The model

An $n \times 1$ random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is said to follow a multivariate t distribution with the location vector $\boldsymbol{\mu}$, scaling covariance matrix $\boldsymbol{\Sigma}$ and degrees-of-freedom ν if $\boldsymbol{Y}$ has the following density function

$$g(\boldsymbol{Y}) = \frac{\Gamma\big((\nu+n)/2\big)}{\Gamma(\nu/2)(\pi\nu)^{n/2}} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{(\boldsymbol{Y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})}{\nu}\right)^{-(n+\nu)/2}. \quad (1)$$

We shall say $\boldsymbol{Y}$ has a $\mathrm{Mt}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ with $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ for $\nu > 1$, and $\mathrm{Cov}(\boldsymbol{Y}) = \nu/(\nu-2)\boldsymbol{\Sigma}$ for $\nu > 2$. It is easy to establish the following two propositions:

Proposition 1. *Let* $u = (\boldsymbol{Y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})$*. If* $\boldsymbol{Y} \sim \mathrm{Mt}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$*, then*

$$\int_0^\infty u^{n/2-1} g(u)\,du = \Gamma(n/2)\pi^{-n/2},$$

where $g(\cdot)$ *is the density (1).*

Proof: The sketch of the proof is given in Appendix A.

Proposition 2. *If* $\boldsymbol{Y} \sim \mathrm{Mt}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *and suppose that* $\boldsymbol{Y}$*,* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *are partitioned as*

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

where $\boldsymbol{Y}^{(1)}$ *and* $\boldsymbol{\mu}^{(1)}$ *are* $m \times 1$ *vectors,* $\boldsymbol{\Sigma}_{11}$ *is an* $m \times m$ *matrix,* $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^{\mathrm{T}}$ *is an* $m \times (n-m)$ *matrix, and* $\boldsymbol{\Sigma}_{22}$ *is an* $(n-m) \times (n-m)$ *matrix. Then*

(a) $\boldsymbol{Y}^{(1)} \sim \mathrm{Mt}_m(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}, \nu)$.

(b) $\boldsymbol{Y}^{(2)} \big| \boldsymbol{Y}^{(1)} = \boldsymbol{y}^{(1)} \sim \mathrm{Mt}_{n-m}(\boldsymbol{\mu}_{2\cdot 1},\ w\boldsymbol{\Sigma}_{22\cdot 1},\ \nu + m)$.

where $\boldsymbol{\mu}_{2\cdot 1} = \boldsymbol{\mu}^{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}^{(1)} - \boldsymbol{\mu}^{(1)})$, $w = \big(\nu + (\boldsymbol{y}^{(1)} - \boldsymbol{\mu}^{(1)})^{\mathrm{T}}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}^{(1)} - \boldsymbol{\mu}^{(1)})\big)/(\nu + m)$ *and* $\boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Proof: The proof is referred to Anderson (2003).

The AR regression model with multivariate t innovations can be written in matrix form as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{Mt}_n(\boldsymbol{0},\ \sigma^2\boldsymbol{C}_n(\boldsymbol{\phi}),\ \nu), \tag{2}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$, $\boldsymbol{X} = [\boldsymbol{1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k]$ is an $n \times (k+1)$ matrix, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^{\mathrm{T}}$ is a $(k+1)$-dimensional vector. $\boldsymbol{C}_n(\boldsymbol{\phi})$ is an $n \times n$ matrix which is an autocorrelation matrix. In the model, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ is the vector of error terms such that

$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \cdots + \phi_p\varepsilon_{t-p} + a_t, \tag{3}$$

where $\{a_t\}$ is a sequence of white noise process with mean zero and a constant variance. In this thesis, we assume that the innovations of (3) have an uncorrelated multivariate t distribution. The autocovariance matrix for $\boldsymbol{\varepsilon}$ is $\mathrm{Cov}(\boldsymbol{\varepsilon}) =$

$\sigma_\nu^2\left[\rho_{|g-h|}\right]$ $(g,\ h = 1, 2, \ldots, n)$, where $\sigma_\nu^2 = \nu\sigma^2/(\nu-2)$ is a scaling variance of ε_t. Notice that ρ_j's are implicit functions of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ which satisfy the Yule-Walker equations (Box et al., 1994)

$$\rho_i = \phi_1\rho_{i-1} + \ldots + \phi_p\rho_{i-p} \quad (i = 1, \ldots, n-1),\ \rho_0 = 1,$$

and the roots of $1-\phi_1 B - \ldots - \phi_p B^p = 0$ need to lie outside the unit circle for assuring the stationarity condition. For model (2), the likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \nu)$ is

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \nu \mid \boldsymbol{Y}) \propto (\sigma^2)^{-n/2}|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}\left(\nu + \frac{S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\phi})}{\sigma^2}\right)^{-(n+\nu)/2}, \qquad (4)$$

where $S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\phi}) = (\boldsymbol{Y} - \boldsymbol{X\beta})^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})(\boldsymbol{Y} - \boldsymbol{X\beta})$.

Typically, any existing software takes enormous time to directly execute matrix inversion for extremely large n. In Appendix B, we give a theorem with explicit formula for computing $\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})$ to remedy this potential problem. In order to facilitate the estimating procedure and achieve the admissibility of $\boldsymbol{\phi}$, we perform the reparameterization scheme of Barndorff-Nielsen and Schou (1973) as follows:

$$\phi_i^{(i)} = \gamma_i, \quad \phi_j^{(i)} = \phi_j^{(i-1)} - \gamma_i\phi_{i-j}^{(i-1)}, \quad j = 1, 2, \ldots, i-1, \qquad (5)$$

where $\phi_j^{(p)} = \phi_j = \phi_j^{(j)} - \phi_{j+1}^{(j+1)}\phi_1^{(j)} - \phi_{j+2}^{(j+2)}\phi_2^{(j+1)} - \cdots - \phi_p^{(p)}\phi_{p-j}^{(p-1)}$, for $j = 1, \ldots, p-1$.

Note that (5) is a one-to-one and onto transformation which reparameterizes $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p) \in \mathbb{C}^p$ in terms of the partial autocorrelations $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p) \in [-1, 1]^p$. In the following, we provide on R-langauge program to compute (5).

```
pacf.to.phi=function(pacf)
{
  p=length(pacf)
  if(p==1) phi=pacf
  if(p>1)
  {
    Phi=matrix(diag(pacf), p, p)
    for (i in 2:p)
        for (j in 1:(i-1))
             Phi[i, j]=Phi[i-1, j]-Phi[i,i]*Phi[i-1, i-j]
    phi=Phi[p,]
  }
  return(phi)
}
```

2.2. Posterior inference

Treating $\boldsymbol{\gamma}$ as a model parameter, we reparameterize (4) as a function of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}, \nu)$. With no prior knowledge about the parameters, we assume the prior distribution is *a priori* independent such that

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}, \nu) = \sigma^{-2}\pi(\boldsymbol{\beta})\prod_{i=1}^{p}\pi(\gamma_i)\pi(\nu) \propto \sigma^{-2}\pi(\nu). \tag{6}$$

A flat prior for $\boldsymbol{\beta}$ is adopted, thus $\pi(\boldsymbol{\beta}) \propto$ constant. Since $-1 < \gamma_i < 1$ for $i = 1, \ldots, p$, it is straightforward to employ the uniform $[-1, 1]$ distribution for γ_i's.

Under the above assumption, the joint posterior density is

$$\begin{aligned}
& p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}, \nu | \boldsymbol{Y}) \\
\propto\ & \pi(\nu)\Big(S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma})\Big)^{-(n+2)/2}\Big|\boldsymbol{C}_n(\boldsymbol{\gamma})\Big|^{-1/2}\Big(\frac{n\sigma^2}{S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma})}\Big)^{\nu/2-1} \\
& \times\Big(1 + \frac{\nu\sigma^2}{S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma})}\Big)^{-(n+\nu)/2}. \qquad (7)
\end{aligned}$$

Reparameterize σ^2 by $\eta = n\sigma^2/S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, $0 < \eta < \infty$, (7) can be rewritten as

$$p(\boldsymbol{\beta}, \eta, \boldsymbol{\gamma}, \nu | \boldsymbol{Y}) \propto \pi(\nu)\Big|\boldsymbol{C}_n(\boldsymbol{\gamma})\Big|^{-1/2}\Big(S(\boldsymbol{\beta}, \boldsymbol{\gamma})\Big)^{-n/2}\eta^{\nu/2-1}\Big(1 + \frac{\nu}{n}\eta\Big)^{-(n+\nu)/2}. \qquad (8)$$

Obviously, $\eta \mid \boldsymbol{\beta}, \boldsymbol{\gamma} \sim \mathrm{F}_{\nu,\, n}$. Integrating (8) w.r.t. η, we obtain the following posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$:

$$\begin{aligned}
& p(\boldsymbol{\beta}, \boldsymbol{\gamma} | \boldsymbol{Y}) \\
\propto\ & |\boldsymbol{C}_n(\boldsymbol{\gamma})|^{-1/2}\Big(\hat{\xi}(\boldsymbol{\gamma}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\boldsymbol{\gamma})\boldsymbol{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\Big)^{-(m+k+1)/2}, \qquad (9)
\end{aligned}$$

where $m = n-k-1$, $\hat{\xi}(\boldsymbol{\gamma}) = S(\boldsymbol{Y}, \hat{\boldsymbol{\beta}}, \boldsymbol{\gamma})$ and $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\boldsymbol{\gamma})\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\boldsymbol{\gamma})\boldsymbol{Y}$. It is noted that the posterior distributions of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and ν are independent. Under the improper prior (6), the result is identical to Chib et al. (1991). Integrating (9) w.r.t. $\boldsymbol{\beta}$, we obtain

$$p(\boldsymbol{\gamma}|\boldsymbol{Y}) \propto |\boldsymbol{C}_n(\boldsymbol{\gamma})|^{-1/2}|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\boldsymbol{\gamma})\boldsymbol{X}|^{-1/2}\hat{\xi}^{-m/2}(\boldsymbol{\gamma}). \qquad (10)$$

Following Ljung and Box (1980), the approximate posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} \mid \boldsymbol{Y} \mathrel{\dot{\sim}} \mathrm{Mt}_{k+1}\Big(\hat{\boldsymbol{\beta}}^{\star},\ \hat{\xi}^{\star}\big(m\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\hat{\boldsymbol{\gamma}})\boldsymbol{X}\big)^{-1},\ m\Big), \qquad (11)$$

where $\hat{\boldsymbol{\beta}}^{\star} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}})$ and $\hat{\xi}^{\star} = \hat{\xi}(\hat{\boldsymbol{\gamma}})$ with $\hat{\boldsymbol{\gamma}}$ optimizing (10). This is easily done by "optim" routine of the statistical package R with the bounded constraint of $(-1,\ 1)$

on γ_i's. Let $\mathrm{F}_{1-\alpha}(\nu_1, \nu_2)$ be the upper $(1-\alpha)$ quantile of the F distribution with degrees-of-freedoms (ν_1, ν_2). We apply the result that if $\boldsymbol{Y} \sim \mathrm{Mt}_n(\boldsymbol{\mu},\ \boldsymbol{\Sigma},\ \nu)$, then $(\boldsymbol{Y}-\boldsymbol{\mu})^{\mathrm{T}}(\nu\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}) \sim \frac{n}{\nu}\mathrm{F}(n,\ \nu)$. Hence, an approximate $(1-\alpha)$ posterior region for $\boldsymbol{\beta}$ can be constructed from

$$\hat{\xi}^{\star^{-1}}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^{\star})^{\mathrm{T}}\Big(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\hat{\gamma})\boldsymbol{X}\Big)(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^{\star}) \leq \frac{k+1}{m}\mathrm{F}_{1-\alpha}(k+1,\ m). \tag{12}$$

2.3. Predictive inference

Let $\boldsymbol{y}_f$ be a $q \times 1$ vector of future responses of $\boldsymbol{Y}$ and $\boldsymbol{x}_f$ be the regressors corresponding to $\boldsymbol{y}_f$. We thus have $\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{y}_f \end{bmatrix} \sim \mathrm{Mt}_{n+q}\Big(\boldsymbol{X}^*\boldsymbol{\beta},\ \sigma^2\boldsymbol{\Omega}^*(\boldsymbol{\gamma}),\ \nu\Big)$, where $\boldsymbol{X}^* = (\boldsymbol{X}^{\mathrm{T}},\ \boldsymbol{x}_f^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\Omega}^*(\boldsymbol{\gamma})$ is an $(n+q) \times (n+q)$ scaling correlation matrix partitioned conformably with $\boldsymbol{Y}^* = (\boldsymbol{Y}^{\mathrm{T}},\ \boldsymbol{y}_f^{\mathrm{T}})^{\mathrm{T}}$, i.e., $\boldsymbol{\Omega}^*(\boldsymbol{\gamma}) = \begin{bmatrix} \boldsymbol{\Omega}_{11}(\boldsymbol{\gamma}) & \boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}) \\ \boldsymbol{\Omega}_{21}(\boldsymbol{\gamma}) & \boldsymbol{\Omega}_{22}(\boldsymbol{\gamma}) \end{bmatrix}$. Notice that $\boldsymbol{\Omega}_{11}(\boldsymbol{\gamma}) = \boldsymbol{C}_n(\boldsymbol{\gamma})$, $\boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}) = \boldsymbol{\Omega}_{21}^{\mathrm{T}}(\boldsymbol{\gamma})$ is an $n \times q$ scaling autocorrelation matrix between $\boldsymbol{Y}$ and $\boldsymbol{y}_f$ and $\boldsymbol{\Omega}_{22}(\boldsymbol{\gamma})$ is a $q \times q$ scaling autocorrelation matrix corresponding to $\boldsymbol{y}_f$.

The joint posterior distribution of $\boldsymbol{y}_f$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{\gamma}, \nu)$ is

$$\begin{aligned} & p(\boldsymbol{y}_f, \boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{\gamma}, \nu|\boldsymbol{Y}) \\ \propto\ & \pi(\nu)(\sigma^{-2})^{(n+q)/2+1}|\boldsymbol{\Omega}^*(\boldsymbol{\gamma})|^{-1/2}\left(1+\frac{\sigma^{-2}S(\boldsymbol{Y}^*, \boldsymbol{\beta}, \boldsymbol{\gamma})}{\nu}\right)^{-(\nu+n+q)/2}, \end{aligned}$$

where

$$S(\boldsymbol{Y}^*, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}^{*^{-1}}(\boldsymbol{\gamma})(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}). \tag{13}$$

Following Osiewalski and Stell (1993), we transform $(\boldsymbol{y}_f, \boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{\gamma}, \nu)$ to $(\boldsymbol{y}_f, \boldsymbol{\beta}, u, \boldsymbol{\gamma}, \nu)$, where $u=\sigma^{-2}S(\boldsymbol{Y}^*,\boldsymbol{\beta},\boldsymbol{\gamma})$. By Proposition 1, it will yield

$$\begin{aligned} p(\boldsymbol{y}_f,\boldsymbol{\beta},\boldsymbol{\gamma}|\boldsymbol{Y}) &\propto |\boldsymbol{\Omega}^*(\boldsymbol{\gamma})|^{-1/2}S(\boldsymbol{Y}^*,\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma}),\boldsymbol{\gamma})^{-(n+q)/2} \\ &\quad\times\left(1+\frac{\big(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma})\big)^{\mathrm{T}}\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{\Omega}^{*-1}(\boldsymbol{\gamma})\boldsymbol{X}^*\big(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma})\big)}{S(\boldsymbol{Y}^*,\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma}),\boldsymbol{\gamma})}\right)^{-(n+q)/2}, \end{aligned} \tag{14}$$

where $S(\boldsymbol{Y}^*,\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma}),\boldsymbol{\gamma})$ is $S(\boldsymbol{Y}^*,\boldsymbol{\beta},\boldsymbol{\phi})$ in (13) with $\boldsymbol{\beta}$ replaced by

$$\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma})=\big(\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{\Omega}^{*^{-1}}(\boldsymbol{\gamma})\boldsymbol{X}^*\big)^{-1}\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{\Omega}^{*^{-1}}(\boldsymbol{\gamma})\boldsymbol{Y}^*.$$

Integrating (14) w.r.t. $\boldsymbol{\beta}$, we obtain

$$p(\boldsymbol{y}_f,\boldsymbol{\gamma}|\boldsymbol{Y})\propto|\boldsymbol{\Omega}^*(\boldsymbol{\gamma})|^{-1/2}S(\boldsymbol{Y}^*,\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma}),\boldsymbol{\gamma})^{-(m+q)/2}|\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{\Omega}^{*^{-1}}(\boldsymbol{\gamma})\boldsymbol{X}^*|^{-1/2}. \tag{15}$$

Since $S(\boldsymbol{Y}^*,\hat{\boldsymbol{\beta}}^*(\boldsymbol{\gamma}),\boldsymbol{\gamma})=S(\boldsymbol{Y},\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})\Big(1+\dfrac{(\boldsymbol{y}_f-\boldsymbol{\mu}_f)^{\mathrm{T}}\boldsymbol{\Omega}_f^{-1}(\boldsymbol{\gamma})(\boldsymbol{y}_f-\boldsymbol{\mu}_f)}{S(\boldsymbol{Y},\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})}\Big)$, where

$$\begin{aligned} \boldsymbol{\mu}_f &= \boldsymbol{x}_f\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})+\boldsymbol{\Omega}_{21}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})\big(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})\big), \\ \boldsymbol{\Omega}_f &= \boldsymbol{\Omega}_{22}(\boldsymbol{\gamma})-\boldsymbol{\Omega}_{21}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{12}(\boldsymbol{\gamma})+\big(\boldsymbol{x}_f-\boldsymbol{\Omega}_{21}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})\boldsymbol{X}\big)^{\mathrm{T}} \\ &\quad\times\big(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})\boldsymbol{X}\big)^{-1}\big(\boldsymbol{x}_f-\boldsymbol{\Omega}_{21}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})\boldsymbol{X}\big), \end{aligned} \tag{16}$$

upon integrating (15) w.r.t. $\boldsymbol{y}_f$, we have

$$p^*(\boldsymbol{\gamma}|\boldsymbol{Y})\propto|\boldsymbol{\Omega}^*(\boldsymbol{\gamma})|^{-1/2}|\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{\Omega}^{*^{-1}}(\boldsymbol{\gamma})\boldsymbol{X}^*|^{-1/2}S(\boldsymbol{Y},\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})^{-m/2}|\boldsymbol{\Omega}_f|^{1/2}. \tag{17}$$

It is noted that the functional form (17) is different from (10).

With arguments similar to (11), we have the following approximate predictive distribution of $\boldsymbol{y}_f$:

$$\boldsymbol{y}_f\mid\boldsymbol{Y}\ \dot{\sim}\ \mathrm{Mt}_q\Big(\hat{\boldsymbol{\mu}}_f,\ \frac{\hat{\boldsymbol{\Omega}}_fS(\boldsymbol{Y},\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\hat{\boldsymbol{\gamma}})}{m},\ m\Big), \tag{18}$$

where $\hat{\boldsymbol{\mu}}_f$ and $\hat{\boldsymbol{\Omega}}_f$ are $\boldsymbol{\mu}_f$ and $\boldsymbol{\Omega}_f$ in (16) with $\hat{\boldsymbol{\gamma}}$ optimizing (17). An approximate $(1-\alpha)$ posterior region for $\boldsymbol{y}_f$ can be obtained from

$$(\boldsymbol{y}_f-\hat{\boldsymbol{\mu}}_f)^{\mathrm{T}}\frac{\hat{\boldsymbol{\Omega}}_f^{-1}}{S(\boldsymbol{Y},\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\hat{\boldsymbol{\gamma}})}(\boldsymbol{y}_f-\hat{\boldsymbol{\mu}}_f)\leq\frac{q}{m}\mathrm{F}_{1-\alpha}(q,\ m).$$

3. Markov chain Monte Carlo inference

3.1. Implementation

Given the state $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\beta}^{(t)}, {\sigma^2}^{(t)}, \boldsymbol{\gamma}^{(t)})$ and an appropriate degrees-of-freedom ν, which can be obtained by the method of moments proposed by Singh (1988, Eq. 2.5) from OLS residuals. The proof is given in Appendix D. The following algorithm is one sweep of MCMC sampler for simulating $\boldsymbol{\theta}^{(t+1)}$ from the posterior distributions (7).

Step 1: Sample $\boldsymbol{\beta}^{(t+1)}$ from $\mathrm{Mt}_{k+1}\left(\hat{\boldsymbol{\beta}}^{(t)}, \frac{\nu {\sigma^2}^{(t)} + \hat{\xi}^{(t)}}{\nu + m}\left(\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\gamma}^{(t)})\boldsymbol{X}\right)^{-1}, \nu{+}m\right)$, where $\hat{\boldsymbol{\beta}}^{(t)} = \left(\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\gamma}^{(t)})\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\gamma}^{(t)})\boldsymbol{Y}$ and $\hat{\xi}^{(t)} = S(\boldsymbol{Y}, \hat{\boldsymbol{\beta}}^{(t)}, \boldsymbol{\gamma}^{(t)})$.

Step 2: Sample $\eta^{(t+1)}$ from $\mathrm{F}_{\nu,n}$, then transform it back to

$$ {\sigma^2}^{(t+1)} = \frac{1}{n}\eta^{(t+1)} S(\boldsymbol{Y}, \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\gamma}^{(t)}). $$

Step 3: Sample $\boldsymbol{\gamma}^{(t+1)}$ via the MH algorithm from

$$ f(\boldsymbol{\gamma}) \propto |\boldsymbol{C}_n(\boldsymbol{\gamma})|^{-1/2}\left(\nu + \frac{S(\boldsymbol{Y}, \boldsymbol{\beta}^{(t+1)}, \boldsymbol{\gamma})}{{\sigma^2}^{(t+1)}}\right)^{-(n+\nu)/2}. $$

To implement the MH algorithm for generating $\boldsymbol{\gamma}$ at the $(t+1)$st iteration, we can transform $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}^* = (\gamma_1^*, \ldots, \gamma_p^*) \in \mathbb{R}^p$, $\mathbb{R} = (-\infty, \infty)$, where, $\gamma_i^* = \log\big((1 + \gamma_i)/(1 - \gamma_i)\big)$ $(i = 1, \ldots, p)$. We then apply the M-H algorithm to $g(\boldsymbol{\gamma}^*) = f(\boldsymbol{\gamma})J_{\boldsymbol{\gamma}^*}$, where $J_{\boldsymbol{\gamma}^*} = \prod_{i=1}^p \left(2e^{\gamma_i^*}/(1 + e^{\gamma_i^*})^2\right)$ is the Jacobain of transforming $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}^*$.

A p-dimensional multivariate normal distribution with mean $\boldsymbol{\gamma}^{*(t)}$ and covariance matrix $c^2\boldsymbol{\Sigma}_{\boldsymbol{\gamma}^*}^{(t)}$ is chosen as the proposal distribution, where the scale $c \approx 2.4/\sqrt{p}$, as suggested in Gelman et al. (2004). The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}^*}^{(t)}$ can be estimated by inverting the sample information matrix given $\boldsymbol{\gamma}^{*^{(t)}}$. Having obtained $\boldsymbol{\gamma}^*$ from the M-H algorithm, we transform it back to $\boldsymbol{\gamma}$ by $\gamma_i = (e^{\gamma_i^*} - 1)/(e^{\gamma_i^*} + 1)$ $(i = 1, \ldots, p)$, then transform $\boldsymbol{\gamma}$ back to $\boldsymbol{\phi}$ by inverting (5).

3.2. Forecasting future values and volatilities

The predictive distribution of $\boldsymbol{y}_f$ is $p(\boldsymbol{y}_f|\boldsymbol{Y}) = \int f(\boldsymbol{y}_f \mid \boldsymbol{Y}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{Y})d\boldsymbol{\theta}$. By Proposition 2, we have

$$f(\boldsymbol{y}_f|\boldsymbol{Y}, \boldsymbol{\theta}) \stackrel{\text{d}}{=} \text{Mt}_q\left(\boldsymbol{y}_f \mid \boldsymbol{\mu}_{2\cdot1},\ w\boldsymbol{\Omega}_{22\cdot1},\ \nu + n\right), \tag{19}$$

where $\boldsymbol{\mu}_{2\cdot1} = \boldsymbol{x}_f\boldsymbol{\beta} + \boldsymbol{\Omega}_{21}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$, $\omega = \dfrac{\nu + (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\text{T}}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})}{\nu + n}$, and $\boldsymbol{\Omega}_{22\cdot1} = \boldsymbol{\Omega}_{22}(\boldsymbol{\gamma}) - \boldsymbol{\Omega}_{21}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{12}(\boldsymbol{\gamma})$.

Let $\boldsymbol{\theta}^{(g)}$ be the generated sample at the gth iteration of the MCMC sampler after the convergence is achieved. The predictive distribution of $\boldsymbol{y}_f$ can be approximated by Monte Carlo integration from the MCMC samples

$$p(\boldsymbol{y}_f|\boldsymbol{Y}) \approx \frac{1}{G}\sum_{g=1}^{G} \text{Mt}_q\left(\boldsymbol{y}_f \mid \boldsymbol{\mu}_{2\cdot1}^{(g)},\ w^{(g)}\boldsymbol{\Sigma}_{22\cdot1}^{(g)},\ \nu + n\right).$$

For the prediction of future values, it is straight forward to generate $\boldsymbol{y}_f^{(g)}$ from (19) given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(g)}$. The forecast of $\boldsymbol{y}_f$ can be computed by

$$\hat{\boldsymbol{y}}_f = \hat{E}(\boldsymbol{y}_f|\boldsymbol{Y}) = \frac{1}{G}\sum_{g=1}^{G} \boldsymbol{y}_f^{(g)}.$$

Let h_{n+1}^2 be the volatility of Y_{n+1} based on $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ for $n \geq p$. Tarami and Pourahmadi (2003) have shown that h_{n+1}^2 is more general than the GARCH(n,1) model. In Appendix E, we provide a different proof based on our considered model.

In case of $q = 1$, we can predict h^2_{n+1} using the sample variance of the MCMC samples $y_f^{(g)}$ $(g = 1, \ldots, G)$, i.e.,

$$\hat{h}^2_{n+1} = \widehat{\mathrm{Var}}(y_f|\boldsymbol{Y}) = \frac{1}{G-1}\sum_{g=1}^{G}(y_f^{(g)} - \hat{y}_f)^2. \tag{20}$$

4. An application: the U.S. interest rates

We consider the 1-year, 3-year and 10-year Treasury constant maturity rates over the period January 1975 to December 2004 for 360 observations. The series are obtained from the Federal Reserve Bank of St Louis on its website

`http://www.stls.frb.org/fred`.

Figure 1 shows the time plots of the three series with solid line denoting the 10-year rate, dashed line the 3-year rate and dotted line the 1-year rate. The three series moved in tandem in the sampling period, indicating that they are highly correlated. To avoid the unit-root behavior of the series, we take the first difference of them to obtain the changed series of interest rates. Let x_{1t}, x_{2t} and y_t denote the changes in the 1-year, 3-year and 10-year interest rates, respectively. We consider an ordinary linear regression model $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t$. The fitted model is given by

$$y_t = -.0.3723x_{1t} + 1.1413x_{2t} + \epsilon_t, \quad \hat{\sigma}_\epsilon = 0.0989,$$

with $R^2 = 91.77\%$. The standard errors of the above regression coefficients are 0.0292, 0.0343, respectively. Both coefficients are highly significant.

Figure 2 displays the time series plot of the residuals and its sample PACF. Obviously, the residual series has some serial correlations at lags 2, 3, and 4. We modify (4.) with an AR(4) model for the error process.

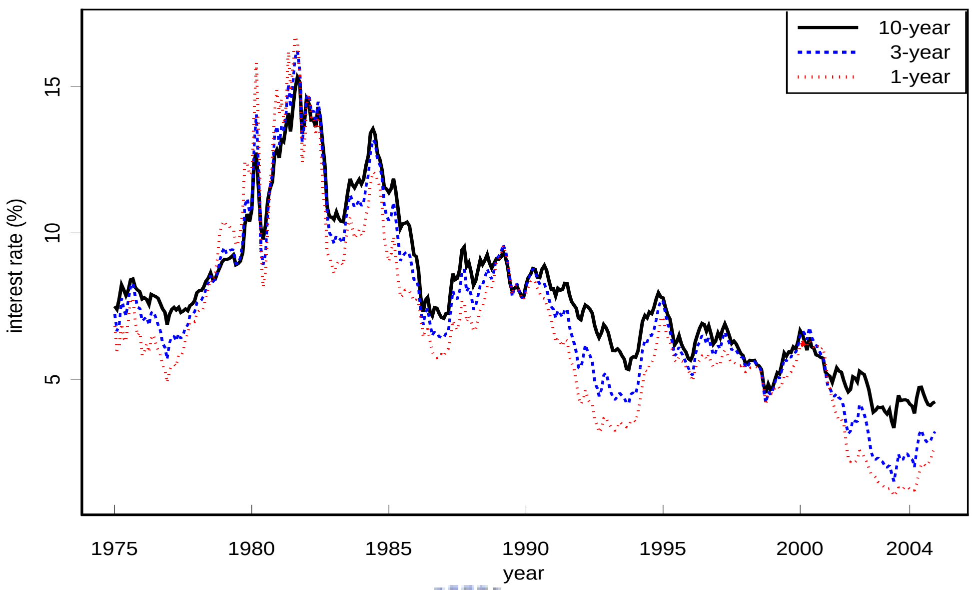

Figure 1: The time series plots of the U.S. interest rates from Jan. 1975 to Dec. 2004.

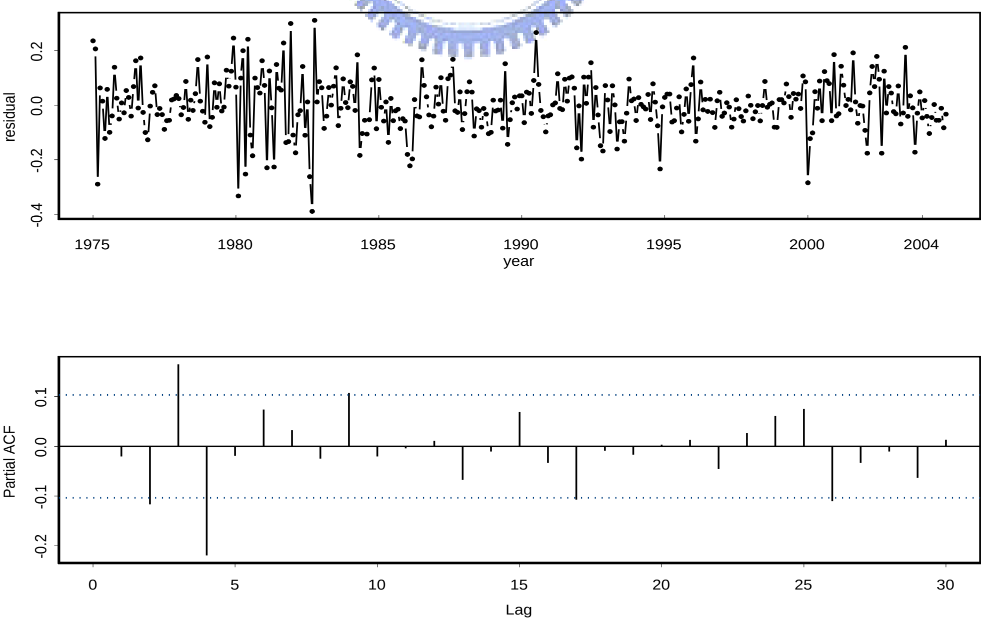

Figure 2: The sample PACF and time series plot of OLS residuals.

Using the traditional approach, we obtain the modified model

$$y_t = -.0.3838x_{1t} + 1.1640x_{2t} + \epsilon_t, \quad \epsilon_t = a_t - 0.1269a_{t-2} + 0.1536a_{t-3} - 0.2189a_{t-4},$$

with $\hat{\sigma}_a = 0.0921$ and $R^2 = 92.8\%$. The standard errors of fitted coefficients are 0.0278, 0.0330, 0.0505, 0.0499, 0.050, respectively. All estimates are significant at the 5% level .

Figure 3 shows the sample ACF, sample PACF and the Q-Q plot of fitted normal residuals. The model no longer has serial correlation within the shocks, however, its tail is heavier than the normal distribution. In addition, we show the time plot of the standardized shocks $\{\tilde{a}^2\}$, and the associated sample ACF and PACF of their squared series in Figure 4. The sample ACF and sample PACF clearly exhibit the existence of conditional heteroscedasticity.

To demonstrate our proposed methodology, we fit an AR(4) model with multivariate t errors, the estimated degrees-of-freedom is 7.82 using the MME of Singh (1988) from the OLS residuals. The MVT-AR(4) model for U.S. interest rates is given as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{Mt}_n(\mathbf{0},\ \sigma^2\boldsymbol{C}_n(\boldsymbol{\phi}),\ \nu), \tag{21}$$

where $\boldsymbol{Y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]$ with $\boldsymbol{X}_j = (x_{j1}, \ldots, , x_{jn})$, $j = 1, 2$, $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3, \phi_4)$, $\nu = 7.82$ and $n = 359$.

To implement our proposed approximate Bayesian and MCMC procedures, we

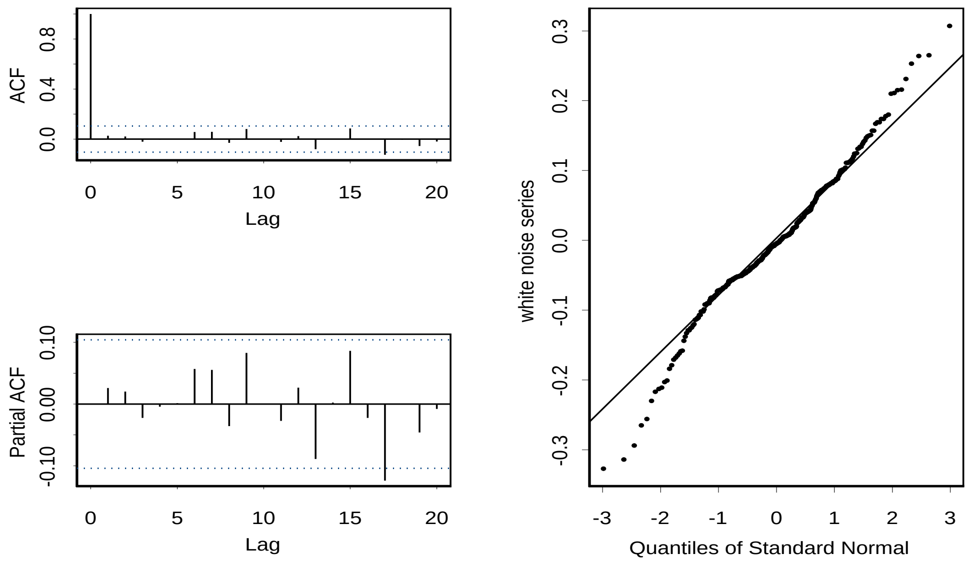

Figure 3: Sample ACF, PACF and the Q-Q plot of fitted normal residuals.

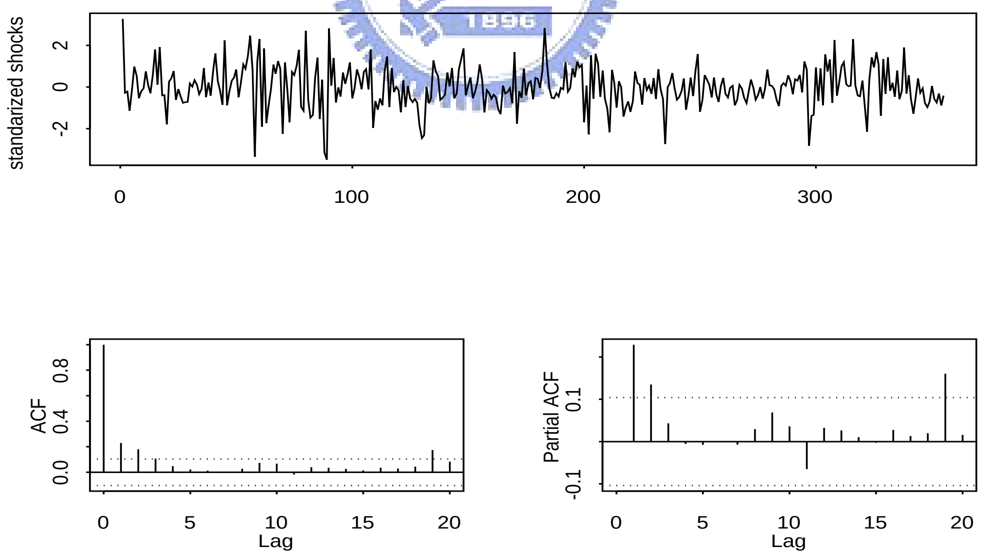

Figure 4: The time series plot of standardized normal residuals, and the associated sample ACF and PACF of the squared sequences.

make use of (5) to reparameterize $\boldsymbol{\phi}$. It leads to

$$
\begin{aligned}
\gamma_1 &= \frac{-\phi_4^2\phi_3 - \phi_1\phi_4 + \phi_2\phi_1\phi_4 + \phi_4\phi_3 + \phi_1 + \phi_2\phi_3}{1 + \phi_4^3 + \phi_2\phi_4^2 - \phi_4^2 - \phi_3\phi_1\phi_4 - \phi_1^2\phi_4 - \phi_4 - \phi_1\phi_3 - \phi_2 - \phi_3^2}, \\
\gamma_2 &= \frac{\phi_1\phi_3 + \phi_1^2\phi_4 + \phi_4\phi_3^2 + \phi_4^2\phi_3\phi_1 + \phi_2\phi_4 - \phi_2\phi_4^3 + \phi_2 - \phi_2\phi_4^2}{1 - \phi_3^2 - 2\,\phi_3\phi_1\phi_4 - \phi_4^2\phi_1^2 - 2\,\phi_4^2 + \phi_4^4}, \\
\gamma_3 &= \frac{\phi_3 + \phi_1\phi_4}{1 - \phi_4^2}, \\
\gamma_4 &= \phi_4.
\end{aligned}
$$

We carry out MCMC by running seven independent parallel chains with different initial values for each chain "over-dispersed" around ± 3 standard deviations of the *maximum a posteriori* (MAP) estimates. For each chain, we implemented 26,000 iterations. We monitored the convergence by examining the *multivariate potential scale reduction factor* (MPSRF) proposed by Brooks and Gelman (1998) of the 7 chains. The convergence occurred after 1,000 iterations. Discarding the first 1,000 iterations as a "burn-in" for each chain, we then stored one imputed parameter value for every 5 iterations to reduce the autocorrelation. Hence, we have 35,000 realizations from the target posterior distribution. Figure 5 displays the convergence diagrams and histograms of the posterior samples of each parameter for one of the seven chains.

The estimates of parameters and their standard errors based on ML and AB approaches, together with the summary statistics of converged MCMC samples, including the mean, standard deviation, median and 95% HPD interval, are listed in Table 1. It can be seen that both ML and AB estimates are similar, but the estimates using MCMC are somewhat different and have a bit larger standard errors.

To compare the confidence and posteior inferences for $\boldsymbol{\beta} = (\beta_1, \beta_2)$ among ML,

Table 1: Summary statistics based on ML estimation, approximate Bayesian, and MCMC sampling methods.

Parameter	ML		App. Bayesian		MCMC				
	Est	Sd	Est	Sd	Mean	Sd	Median	2.5%	97.5%
β_1	–0.3804	0.0283	–0.3803	0.0283	–0.3813	0.0321	–0.3807	–0.4452	–0.3196
β_2	1.1560	0.0336	1.1558	0.0335	1.1584	0.0550	1.1568	1.0547	1.2709
σ^2	0.0088	0.0045	0.0088	0.0045	0.0092	0.0047	0.0084	0.0024	0.0205
ϕ_1	0.0267	0.0515	0.0282	0.0515	0.0326	0.0549	0.0316	–0.0729	0.1401
ϕ_2	–0.1412	0.0509	–0.1414	0.0509	–0.1400	0.0543	–0.1400	–0.2473	–0.0327
ϕ_3	0.1708	0.0442	0.1704	0.0442	0.1721	0.0532	0.1735	0.0668	0.2766
ϕ_4	–0.2319	0.0478	–0.2317	0.0478	–0.2320	0.0550	–0.2328	–0.3415	–0.1197

AB and MCMC approaches, the 95% confidence region constructed by (25), the 95% posterior confidence region by (12), and the scatter plot of randomly selected 5000 MCMC samples for $\boldsymbol{\beta} = (\beta_1, \beta_2)$ are superimposed in Figure 6. In the figure, we found that the two regions constructed by ML and AB are nearly the same, however, it only contains about 75% of the MCMC sample. This reveals that posterior inferences based on MCMC samples can be quite different from the ML and AB approaches for our considered model. With this phenomenon, we have conducted a simulation study to compare the converge probabilities among the three approaches. We found the MCMC approach has better performance than AB and ML approaches as ν is small, while they have similar performances as ν is large. We skip the details to save space.

As pointed out by Tarami and Pourahmadi (2003), the multivariate t process has the advantage of describing the evolution of volatility and has a general GARCH-type, which is fully robust as opposed to Gaussian or i.i.d. univariate t distribution. Using the Bayesian prediction procedure described in Section 3.2, we reuse the converged MCMC samples to generate volatility estimates according to (20). The estimated volatility process for model (21) is shown in Figure 7. This plot appears to be reasonable since the estimated volatilities exhibit conditional heteroscedasticity.

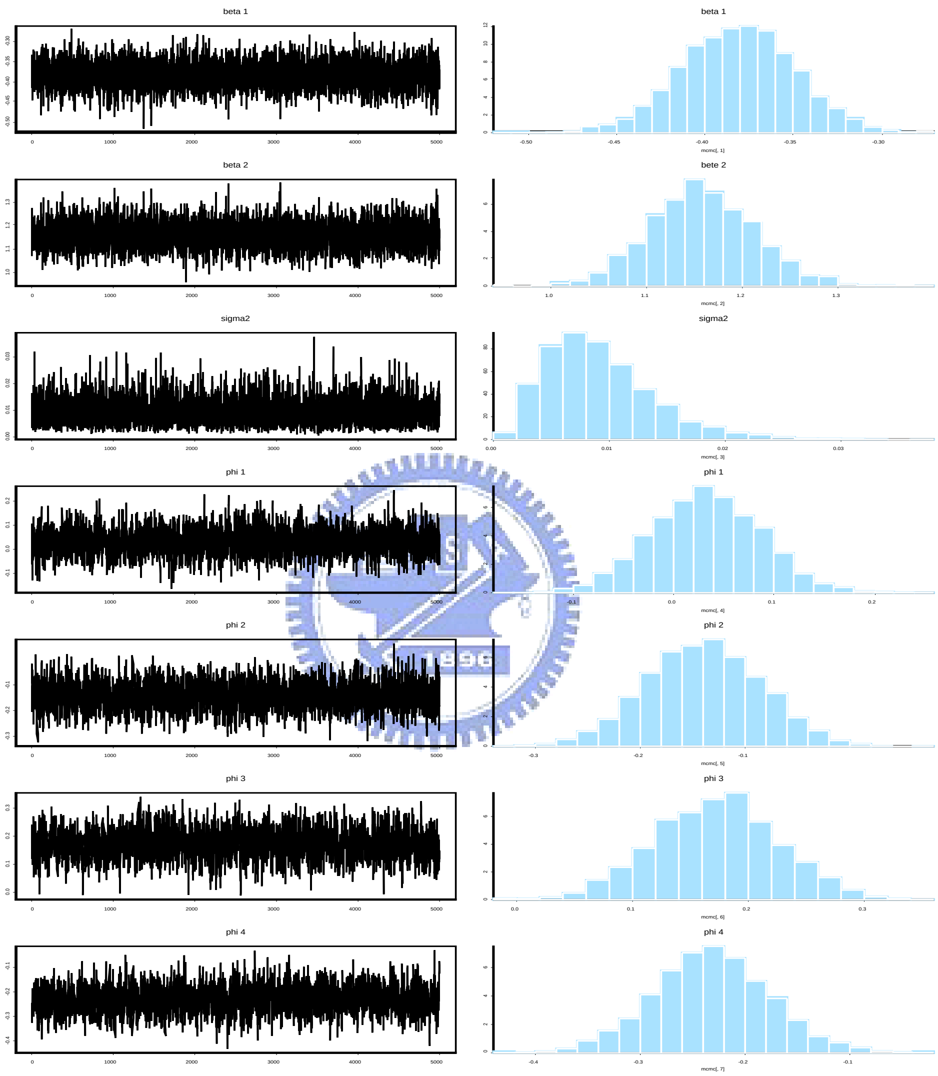

Figure 5: Convergence diagrams and histograms of the distribution of each parameter.

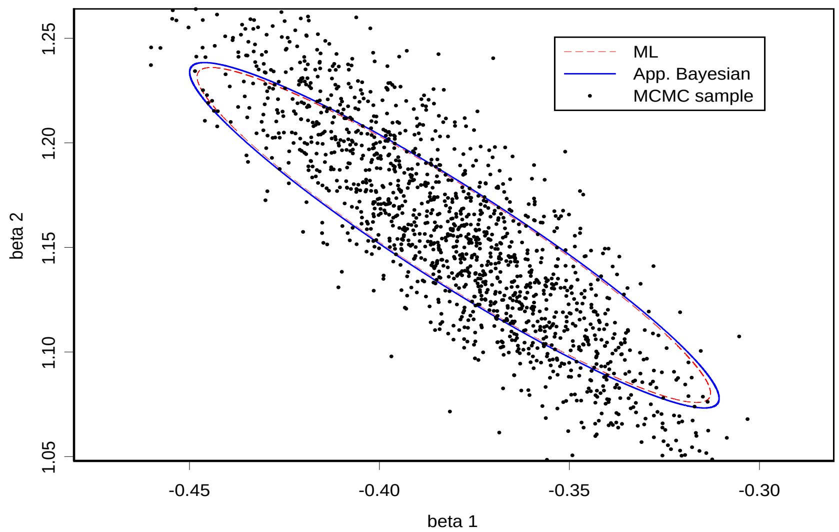

Figure 6: The ML, approximate Bayesian, and MCMC confidence ellipsoid for $\boldsymbol{\beta}$.

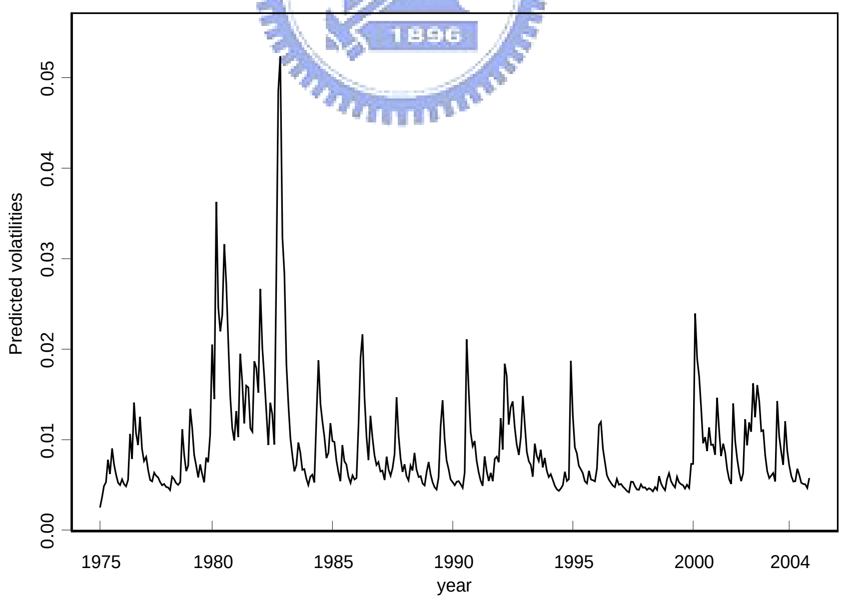

Figure 7: Estimated volatilities of the 10-year U.S. interest rate.

5. Discussion

With the chosen non-informative prior, we present the practical approximate Bayesian approach and a straightforward MCMC sampling scheme for regression models when the error vector is distributed as a multivariate t process with AR(p) serial correlations. Based on this work, these results can be readily extended to ARMA models with the same reparameterization on moving average parameters (Monahan, 1984). This model is suitable for the fitting of financial time series data since it explains autocorrelation and conditional heteroscedasticity simultaneously. Also, we have shown how to estimate volatilities from the generated MCMC samples.

For the multivariate t process with non-informative prior information, the posterior distribution of degrees-of-freedom ν is the same as its prior distribution and hence is estimated by MME. Future work will seek other Bayesian treatments to update posterior samples of ν based on other types of prior and compare relative merits among them.

Appendix A: Proof of Proposition 1

If $u^* \sim \boldsymbol{F}_{n,\nu}$, then

$$\int_0^\infty \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{n}{\nu}\right)^{n/2} {u^*}^{n/2-1} \left(1+\frac{n}{\nu}u^*\right)^{-(n+\nu)/2} du^* = 1.$$

It leads to

$$\int_0^\infty {u^*}^{n/2-1} \left(1+\frac{n}{\nu}u^*\right)^{-(n+\nu)/2} du^* = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{n+\nu}{2}\right)} \left(\frac{\nu}{n}\right)^{n/2}. \tag{22}$$

Let $\boldsymbol{Y} \sim \mathrm{Mt}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $\boldsymbol{u} = (\boldsymbol{Y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})$. The density of $\boldsymbol{u}$ is

$$g_{n,\nu}(\boldsymbol{u}) = \frac{\Gamma((n+\nu)/2)}{\Gamma(\nu/2)} (\nu\pi)^{-n/2} \left(1+\frac{\boldsymbol{u}}{\nu}\right)^{-(n+u)/2}.$$

Therefore,

$$\begin{aligned} & \int_0^\infty u^{n/2-1} g_{n,\nu}(\boldsymbol{u}) d\boldsymbol{u} \\ = & \int_0^\infty \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} (\nu\pi)^{-n/2} \boldsymbol{u}^{n/2-1} \left(1+\frac{\boldsymbol{u}}{\nu}\right)^{-\frac{n+\nu}{2}} d\boldsymbol{u}. \end{aligned}$$

Define $\boldsymbol{u}^* = \boldsymbol{u}/n$ and from the result of (22), it can be seen that

$$\begin{aligned} & \int_0^\infty \boldsymbol{u}^{n/2-1} g_{n,\nu}(\boldsymbol{u}) d\boldsymbol{u} \\ = & \frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} (\nu\pi)^{-n/2} n^{n/2} \int_0^\infty {\boldsymbol{u}^*}^{n/2-1} \left(1+\frac{n}{\nu}\boldsymbol{u}^*\right)^{-\frac{n+\nu}{2}} d\boldsymbol{u}^* \\ = & \Gamma(n/2)\pi^{-n/2}. \end{aligned}$$

This completes the proof.

Appendix B: An inverse of the AR(p) correlation matrix

Theorem 1. *Let* $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ *be the observations from a stationary autoregressive process of order p, i.e.,* $Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} = a_t$*, where* $\{a_t\}$ *is a independent zero mean white noise process with a constant variance. Denoting* $\boldsymbol{C}_n(\boldsymbol{\phi})$ *as the autocorrelation matrix of* $\boldsymbol{Z}$ *and*

$$\boldsymbol{M} = \left[\begin{array}{c|c} \boldsymbol{I}_p & \mathbf{0}_{p\times(n-p)} \\ \hline \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{array}\right],$$

where $[\boldsymbol{M}_{21} \;\; \boldsymbol{M}_{22}] = [m_{ij}]$ *is an* $(n-p)\times n$ *submatrix of* $\boldsymbol{M}$ *with*

$$m_{ij} = \begin{cases} -\phi_k, & j = i+p-k,\; k = 1,2,\ldots,p, \\ 1, & j = i+p, \\ 0, & \text{otherwise}. \end{cases}$$

Let

$$\boldsymbol{M}_{21}^{(p)} = \begin{bmatrix} -\phi_p & -\phi_{p-1} & \cdots & -\phi_1 \\ & -\phi_p & & -\phi_2 \\ & & \ddots & \vdots \\ 0 & & & -\phi_p \end{bmatrix} \text{ and } \boldsymbol{M}_{22}^{(p)^{\mathrm{T}}} = \begin{bmatrix} 1 & -\phi_1 & \cdots & -\phi_{p-1} \\ & 1 & & -\phi_{p-2} \\ & & \ddots & \vdots \\ & 0 & & 1 \end{bmatrix}$$

be the $p\times p$ *leading principle minor of* $\boldsymbol{M}_{21}$ *and* $\boldsymbol{M}_{22}$*, respectively. Then the inverse autocorrelation matrix of* $\boldsymbol{C}_n$ *is given by*

$$\boldsymbol{C}_n^{-1} = \begin{bmatrix} \boldsymbol{M}_{21}^{\mathrm{T}}\boldsymbol{M}_{21} + \boldsymbol{\Psi}_p & \boldsymbol{M}_{21}^{\mathrm{T}}\boldsymbol{M}_{22} \\ \boldsymbol{M}_{22}^{\mathrm{T}}\boldsymbol{M}_{21} & \boldsymbol{M}_{22}^{\mathrm{T}}\boldsymbol{M}_{22} \end{bmatrix}, \quad n \geq p, \tag{23}$$

where $\boldsymbol{\Psi}_p = \boldsymbol{M}_{22}^{(p)}\boldsymbol{M}_{22}^{(p)^{\mathrm{T}}} - \boldsymbol{M}_{21}^{(p)^{\mathrm{T}}}\boldsymbol{M}_{21}^{(p)}$.

For the proof of above theorem, see Lee et al. (2004). In general, the inversion of an $n \times n$ matrix needs $O(n^2)$ operations, while using Theorem 1 it takes only $O(n)$ operations. An R program for computing (23) is given below.

```
arp.cov.inv = function(phi, n)
 {
   p = length(phi)
   M21 = M22 = matrix(0, p, p)
   temp.phi = c(1, -phi)
   for( i in 1:p)
   {
     M21[i, i:p] = - rev(phi)[1: (p-i+1)]
     M22[i:p, i] = temp.phi[1: (p-i+1)]
   }
   Omega.inv = M22 %*% t(M22) - t(M21) %*% (M21)
   if(N != p){
     M2 = matrix(0, n-p, n)
     for(i in 1:(n-p))
     {
       M2[i, i:(i+p)] = rev(temp.phi)
     }
     Lambda = t(M2) %*% M2
     Lambda[1:p,1:p] = Lambda[1:p,1:p] + Omega.inv
     Cov=Lambda
   }
   else Cov=Omega.inv
 }
```

Appendix C: Maximum likelihood inferences

Using the profile likelihood approach, the ML estimates of the parameters $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\phi}$ denoted as $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$, $\hat{\sigma}^2_{\mathrm{ML}}$ and $\hat{\boldsymbol{\phi}}_{\mathrm{ML}}$ for model (2) are given as:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{ML}} &= \Big(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\hat{\boldsymbol{\phi}}_{\mathrm{ML}})\boldsymbol{X}\Big)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\hat{\boldsymbol{\phi}}_{\mathrm{ML}})\boldsymbol{Y}, \\
\hat{\sigma}^2_{\mathrm{ML}} &= \frac{1}{n}\Big(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{ML}}\Big)^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\hat{\boldsymbol{\phi}}_{\mathrm{ML}})\Big(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{ML}}\Big), \\
\hat{\boldsymbol{\phi}}_{\mathrm{ML}}(\hat{\boldsymbol{\gamma}}) &= \underset{\gamma_{\phi_1},\ldots,\gamma_{\phi_p}}{\operatorname{argmax}}\ -\frac{n}{2}\log\Big(\hat{\sigma}^2_{\mathrm{ML}}\Big(\boldsymbol{\phi}(\boldsymbol{\gamma})\Big)\Big)+\frac{1}{2}\log\Big|\boldsymbol{C}_n^{-1}\Big(\boldsymbol{\phi}(\boldsymbol{\gamma})\Big)\Big|.
\end{aligned}$$

The Fisher information for $\boldsymbol{\theta}=(\boldsymbol{\beta},\sigma^2,\boldsymbol{\phi})$ is

$$\boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}=\begin{bmatrix}\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{I}_{\boldsymbol{\beta}\sigma^2} & \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\phi}}\\ \boldsymbol{I}_{\boldsymbol{\beta}\sigma^2} & \boldsymbol{I}_{\sigma^2\sigma^2} & \boldsymbol{I}_{\sigma^2\boldsymbol{\phi}}\\ \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\phi}} & \boldsymbol{I}_{\sigma^2\boldsymbol{\phi}} & \boldsymbol{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}\end{bmatrix}. \tag{24}$$

The elements of (24) are

$$\begin{aligned}
\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{\nu+n}{\sigma^2\,(\nu+n+2)}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}, \qquad \boldsymbol{I}_{\boldsymbol{\beta}\sigma^2}=\mathbf{0}, \qquad \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\phi}}=\mathbf{0}, \\
\boldsymbol{I}_{\sigma^2\sigma^2} &= \frac{n\nu}{2\sigma^4(\nu+n+2)}, \\
\boldsymbol{I}_{\sigma^2\phi_i} &= \frac{\nu}{2\sigma^2(\nu+n+2)}\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\right), \quad i=1,\ldots,p, \\
\boldsymbol{I}_{\phi_i\phi_j} &= \frac{1}{2(\nu+n+2)}\left((\nu+n)\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\right)\right. \\
&\qquad \left.-\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\right)\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\right)\right), \quad i,\ j=1,\ldots,p.
\end{aligned}$$

Under some regularity conditions, the standard errors for $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}=(\hat{\boldsymbol{\beta}}_{\mathrm{ML}},\hat{\sigma}^2_{\mathrm{ML}},\hat{\boldsymbol{\phi}}_{\mathrm{ML}})$ can be estimated by taking the square root of the corresponding diagonal elements of $\mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})$. Approximately, a $(1-\alpha)$ confidence region for $\boldsymbol{\beta}$ can be constructed from

$$(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\mathrm{ML}})^{\mathrm{T}}\left(\frac{\nu+n}{\hat{\sigma}^2_{\mathrm{ML}}(\nu+n+2)}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}_n^{-1}(\hat{\boldsymbol{\phi}}_{\mathrm{ML}})\boldsymbol{X}\right)(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\mathrm{ML}})\le\chi^2_{m_1}(\alpha), \tag{25}$$

where $\chi^2_{m_1}(\alpha)$ denotes the 100(1–α) quantile of a chi-square distribution with m_1 degrees-of-freedom.

Proof of (24):

The log-likelihood function for $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})$ is

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) \propto -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}log|\boldsymbol{C}_n(\boldsymbol{\phi})| - \frac{n+\nu}{2}\log(1 + \frac{S(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\phi})}{\nu\sigma^2}).$$

For notational simplicity, let $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$ and $\Delta = \boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}$. The score vectors of $(\boldsymbol{\beta}, \sigma^2, \phi_i)$ are

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\beta}} &= (n+\nu)\frac{\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2 + \Delta}, \\
\mathbf{s}_{\sigma^2} &= -\frac{n}{2\sigma^2} + \frac{n+\nu}{2\sigma^2}(\frac{\Delta}{\nu\sigma^2+\Delta}), \\
\mathbf{s}_{\phi_i} &= -\frac{1}{2}\left[\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right) - (n+\nu)\frac{\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2+\Delta}\right].
\end{aligned}$$

Let $\psi_i = \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})$ and $\psi_j = \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})$. The elements of the Hessian matrix are as follows:

$$\begin{aligned}
\mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= (n+\nu)\frac{\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})(-\boldsymbol{X})(\nu\sigma^2+\Delta) - \boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}(-2\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e})^\top}{(\nu\sigma^2+\Delta)^2} \\
&= (n+\nu)\left(\frac{-\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{\nu\sigma^2+\Delta} + \frac{2\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{(\nu\sigma^2+\Delta)^2}\right), \\
\mathbf{H}_{\boldsymbol{\beta}\sigma^2} &= (n+\nu)\frac{-\nu(\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e})}{(\nu\sigma^2+\Delta)^2} = -\nu(n+\nu)\frac{\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}, \\
\mathbf{H}_{\boldsymbol{\beta}\phi_i} &= -(n+\nu)\left(\frac{\boldsymbol{X}^\top \psi_i \boldsymbol{e}}{\nu\sigma^2+\Delta} - \frac{1}{(\nu\sigma^2+\Delta)^2}\left(\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}\boldsymbol{e}^\top \psi_i \boldsymbol{e}\right)\right), \\
\mathbf{H}_{\sigma^2\phi_i} &= -\frac{1}{2}\left(-(n+\nu)\frac{(-\nu)\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) = -\frac{\nu(n+\nu)}{2}\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}, \\
\mathbf{H}_{\sigma^2\sigma^2} &= \frac{n}{2\sigma^4} - \left(\frac{(n+\nu)}{2\sigma^4}\left(\frac{\Delta}{\nu\sigma^2+\Delta}\right) - \left(\frac{n+\nu}{2\sigma^2}\right)\frac{-\nu\Delta}{(\nu\sigma^2+\Delta)^2}\right) \\
&= \frac{n}{2\sigma^4} - \frac{(n+\nu)}{2\sigma^4}\left(\frac{\Delta}{\nu\sigma^2+\Delta} + \frac{\sigma^2\nu\Delta}{(\nu\sigma^2+\Delta)^2}\right),
\end{aligned}$$

$$
\begin{aligned}
\mathbf{H}_{\phi_i\phi_j} &= \frac{1}{2}\Bigg(\mathrm{tr}\left(\psi\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\right) - \mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i\partial\phi_j}\right) \\
&\quad -(n+\nu)\frac{\boldsymbol{e}^\top\left(2\psi_i\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i\partial\phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)\boldsymbol{e}}{\nu\sigma^2+\Delta} \\
&\quad +(n+\nu)\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}\boldsymbol{e}^\top\psi_j\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\Bigg).
\end{aligned}
$$

The Fisher information matrix is obtained by the negative expectation of the Hessian matrix. We list several important formulae which are useful to obtain the Fisher information matrix.

(a)

$$
\begin{aligned}
E\left(\frac{1}{\nu\sigma^2+\Delta}\right) &= E\left[\frac{1}{\nu\sigma^2+(\boldsymbol{Y}-\boldsymbol{X\beta})^\top\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})(\boldsymbol{Y}-\boldsymbol{X\beta})}\right] \\
&= \frac{1}{\nu\sigma^2}E\left[\left(1+\frac{\boldsymbol{e}^\top\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}\right)^{-1}\right] \\
&= \frac{1}{\nu\sigma^2}\int\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\left(1+\frac{\boldsymbol{e}^\top\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}\right)^{-\frac{n+\nu+2}{2}}d\boldsymbol{Y} \\
&= \frac{1}{\nu\sigma^2}\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\int\left(1+\frac{\boldsymbol{e}^\top\left(\frac{\nu}{\nu+2}\boldsymbol{C}_n(\boldsymbol{\phi})\right)^{-1}\boldsymbol{e}}{(\nu+2)\sigma^2}\right)^{-\frac{n+\nu+2}{2}}d\boldsymbol{Y} \\
&= \frac{1}{\nu\sigma^2}\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\left(\frac{\Gamma(\frac{\nu+2}{2})\left(\pi(\nu+2)\sigma^2\right)^{n/2}}{\Gamma(\frac{n+\nu+2}{2})|\frac{\nu}{\nu+2}\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}\right) \\
&= \frac{1}{\nu\sigma^2}\frac{\frac{\nu}{2}(\nu+2)^{n/2}}{\frac{n+\nu}{2}\nu^{n/2}}\left(\frac{\nu}{\nu+2}\right)^{n/2} = \frac{1}{\sigma^2(n+\nu)}.
\end{aligned}
$$

(b)

$$
\begin{aligned}
E\left(\frac{\Delta}{\nu\sigma^2+\Delta}\right) &= E\left(\frac{\Delta+\nu\sigma^2-\nu\sigma^2}{\nu\sigma^2+\Delta}\right)\\
&= E\left(1-\frac{\nu\sigma^2}{\nu\sigma^2+\Delta}\right)\\
&= 1-\nu\sigma^2 E\left(\frac{1}{\nu\sigma^2+\Delta}\right)\\
&= 1-\nu\sigma^2\frac{1}{\sigma^2(n+\nu)}\\
&= \frac{n}{n+\nu}.
\end{aligned}
$$

(c)

$$
\begin{aligned}
& E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top}{\nu\sigma^2+\Delta}\right)\\
=\; & E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top}{\nu\sigma^2+\boldsymbol{e}^\top\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}\right)\\
=\; & \frac{1}{\nu\sigma^2}E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top}{1+\frac{\boldsymbol{e}^\top\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}}\right)\\
=\; & \frac{1}{\nu\sigma^2}\int \boldsymbol{e}\boldsymbol{e}^\top\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\left(1+\frac{\boldsymbol{e}^\top\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}\right)^{-\frac{n+\nu+2}{2}}d\boldsymbol{Y}\\
=\; & \frac{1}{\nu\sigma^2}\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\int \boldsymbol{e}\boldsymbol{e}^\top\left(1+\frac{\boldsymbol{e}^\top\left(\frac{\nu}{\nu+2}\boldsymbol{C}_n(\boldsymbol{\phi})\right)^{-1}\boldsymbol{e}}{(\nu+2)\sigma^2}\right)^{-\frac{n+\nu+2}{2}}d\boldsymbol{Y}\\
=\; & \frac{1}{\nu\sigma^2}\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\left(\frac{\Gamma(\frac{\nu+2}{2})\left(\pi(\nu+2)\sigma^2\right)^{n/2}}{\Gamma(\frac{n+\nu+2}{2})|\frac{\nu}{\nu+2}\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}\right)E(\boldsymbol{e}\boldsymbol{e}^\top)\\
=\; & \frac{1}{\nu\sigma^2}\frac{\frac{\nu}{2}(\nu+2)^{n/2}}{\frac{n+\nu}{2}\nu^{n/2}}\left(\frac{\nu}{\nu+2}\right)^{n/2}\sigma^2\boldsymbol{C}_n(\boldsymbol{\phi})\\
=\; & \frac{\boldsymbol{C}_n(\boldsymbol{\phi})}{n+\nu}.
\end{aligned}
$$

(d)

$$
\begin{aligned}
& E\left(\frac{1}{(\nu\sigma^2+\Delta)^2}\right) \\
= \; & \frac{1}{\nu^2\sigma^4} E\left(\frac{1}{[1+\frac{\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}]^2}\right) \\
= \; & \frac{1}{\nu^2\sigma^4} \int \frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}} \left(1+\frac{\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}\right)^{-\frac{n+\nu+4}{2}} d\boldsymbol{Y} \\
= \; & \frac{1}{\nu^2\sigma^4} \frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}} \left(\frac{\Gamma(\frac{\nu+4}{2})\left(\pi(\nu+4)\sigma^2\right)^{n/2}}{\Gamma(\frac{n+\nu+4}{2})|\frac{\nu}{\nu+4}\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}\right) \\
= \; & \frac{1}{\nu^2\sigma^4} \frac{(\frac{\nu}{2})(\frac{\nu+2}{2})}{(\frac{n+\nu}{2})(\frac{n+\nu+2}{2})} \\
= \; & \frac{\nu+2}{\sigma^4\nu(n+\nu)(n+\nu+2)}.
\end{aligned}
$$

(e)

$$
\begin{aligned}
& E\left(\frac{\Delta}{(\nu\sigma^2+\Delta)^2}\right) \\
= \; & E\left(\frac{\Delta+\nu\sigma^2-\nu\sigma^2}{(\nu\sigma^2+\Delta)^2}\right) \\
= \; & E\left(\frac{1}{\nu\sigma^2+\Delta} - \frac{\nu\sigma^2}{(\nu\sigma^2+\Delta)^2}\right) \\
= \; & E\left(\frac{1}{\nu\sigma^2+\Delta}\right) - \nu\sigma^2 E\left(\frac{1}{(\nu\sigma^2+\Delta)^2}\right) \\
= \; & \frac{1}{\sigma^2(n+\nu)} - \nu\sigma^2\frac{\nu+2}{\sigma^4\nu(n+\nu)(n+\nu+2)} \\
= \; & \frac{n}{\sigma^2(n+\nu)(n+\nu+2)}.
\end{aligned}
$$

(f)

$$
\begin{aligned}
& E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top}{(\nu\sigma^2+\Delta)^2}\right) \\
= \ & \frac{1}{\nu^2\sigma^4} E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top}{\left(1+\frac{\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}\right)^2}\right) \\
= \ & \frac{1}{\nu^2\sigma^4}\int \boldsymbol{e}\boldsymbol{e}^\top \frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\left(1+\frac{\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{\nu\sigma^2}\right)^{-\frac{n+\nu+4}{2}} d\boldsymbol{Y} \\
= \ & \frac{1}{\nu^2\sigma^4}\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\int \boldsymbol{e}\boldsymbol{e}^\top \left(1+\frac{\boldsymbol{e}^\top\left(\frac{\nu}{\nu+4}\boldsymbol{C}_n(\boldsymbol{\phi})\right)^{-1}\boldsymbol{e}}{(\nu+4)\sigma^2}\right)^{-\frac{n+\nu+4}{2}} d\boldsymbol{Y} \\
= \ & \frac{1}{\nu^2\sigma^4}\frac{\Gamma(\frac{n+\nu}{2})|\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}{\Gamma(\frac{\nu}{2})(\pi\nu\sigma^2)^{n/2}}\left(\frac{\Gamma(\frac{\nu+4}{2})\left(\pi(\nu+4)\sigma^2\right)^{n/2}}{\Gamma(\frac{n+\nu+4}{2})|\frac{\nu}{\nu+4}\boldsymbol{C}_n(\boldsymbol{\phi})|^{-1/2}}\right)E(\boldsymbol{e}\boldsymbol{e}^\top) \\
= \ & \frac{1}{\nu^2\sigma^4}\frac{(\frac{\nu}{2})(\frac{\nu+2}{2})}{(\frac{n+\nu}{2})(\frac{n+\nu+2}{2})}\frac{\nu}{\nu+2}\sigma^2\boldsymbol{C}_n(\boldsymbol{\phi}) \\
= \ & \frac{\boldsymbol{C}_n(\boldsymbol{\phi})}{\sigma^2(n+\nu)(n+\nu+2)}.
\end{aligned}
$$

(g)

$$
\begin{aligned}
& E\left(\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}}{\nu\sigma^2+\Delta}\right) \\
= \ & E\mathrm{tr}\left(\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}}{\nu\sigma^2+\Delta}\right) \\
= \ & E\mathrm{tr}\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top\psi_i}{\nu\sigma^2+\Delta}\right) \\
= \ & \mathrm{tr}E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top\psi_i}{\nu\sigma^2+\Delta}\right) \\
= \ & \frac{1}{n+\nu}\mathrm{tr}\left(\boldsymbol{C}_n(\boldsymbol{\phi})\psi_i\right) \\
= \ & \frac{1}{n+\nu}\mathrm{tr}\left(\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right) \\
= \ & \frac{1}{n+\nu}\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\right).
\end{aligned}
$$

(h)

$$\begin{aligned}
E\left(\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) &= E\text{tr}\left(\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right)\\
&= E\text{tr}\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top\psi_i}{(\nu\sigma^2+\Delta)^2}\right)\\
&= \text{tr}E\left(\frac{\boldsymbol{e}\boldsymbol{e}^\top\psi_i}{(\nu\sigma^2+\Delta)^2}\right)\\
&= \frac{1}{\sigma^2(n+\nu)(n+\nu+2)}\text{tr}\left(\boldsymbol{C}_n(\boldsymbol{\phi})\psi_i\right)\\
&= \frac{1}{\sigma^2(n+\nu)(n+\nu+2)}\text{tr}\left(\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)\\
&= \frac{1}{\sigma^2(n+\nu)(n+\nu+2)}\text{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\right).
\end{aligned}$$

Proof of (h):

Using the fact that

$$E\left(\frac{\partial^2\log f(\boldsymbol{Y})}{\partial\phi_i\partial\phi_j}\right) = -E\left(\frac{\partial\log f(\boldsymbol{Y})}{\partial\phi_i}\frac{\partial\log f(\boldsymbol{Y})}{\partial\phi_j}\right),$$

and apply the following equation $E\left(\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}\boldsymbol{e}^\top\psi_j\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right)$, we first calculate that

$$\begin{aligned}
&\frac{\partial^2\log f(\boldsymbol{Y})}{\partial\phi_i\partial\phi_j}\\
=\;& \frac{1}{2}\text{tr}\left(\psi_i\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\right) - \frac{1}{2}\text{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i\partial\phi_j}\right) + \frac{n+\nu}{2}\left(\frac{\boldsymbol{e}^\top\psi_i\boldsymbol{e}\boldsymbol{e}^\top\psi_j\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right)\\
&-\frac{(n+\nu)}{2}\left(\frac{\boldsymbol{e}^\top\left(2\psi_i\frac{\partial\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2\boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i\partial\phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)\boldsymbol{e}}{(\nu\sigma^2+\Delta)}\right),
\end{aligned}$$

and

$$
\begin{aligned}
& - \quad \frac{n+\nu}{2} E\left(\frac{\boldsymbol{e}^{\top}\left(2\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)\boldsymbol{e}}{\nu\sigma^2+\Delta}\right) \\
= \quad & -\frac{n+\nu}{2} E\mathrm{tr}\left(\frac{\boldsymbol{e}^{\top}\left(2\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)\boldsymbol{e}}{\nu\sigma^2+\Delta}\right) \\
= \quad & -\frac{n+\nu}{2} E\mathrm{tr}\left(\frac{\boldsymbol{e}\boldsymbol{e}^{\top}\left(2\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)}{\nu\sigma^2+\Delta}\right) \\
= \quad & -\frac{n+\nu}{2} \mathrm{tr}E\left(\frac{\boldsymbol{e}\boldsymbol{e}^{\top}\left(2\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)}{\nu\sigma^2+\Delta}\right) \\
= \quad & -\frac{n+\nu}{2}\frac{1}{n+\nu}\mathrm{tr}\left(2\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n(\boldsymbol{\phi})\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right) \\
= \quad & -\mathrm{tr}\left(\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right) \\
= \quad & -\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& E\left(\frac{\partial^2 \log f(\boldsymbol{Y})}{\partial \phi_i \partial \phi_j}\right) \\
= \quad & \frac{1}{2}\mathrm{tr}\left(\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\right) + \frac{n+\nu}{2}E\left(\frac{\boldsymbol{e}^{\top}\psi_i \boldsymbol{e}\boldsymbol{e}^{\top}\psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) \\
& -\frac{(n+\nu)}{2} E\left(\frac{\boldsymbol{e}^{\top}\left(2\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi}) - \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial^2 \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i \partial \phi_j}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\right)\boldsymbol{e}}{\nu\sigma^2+\Delta}\right) \\
= \quad & \frac{1}{2}\mathrm{tr}\left(\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) - \mathrm{tr}\left(\psi_i \frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) + \frac{n+\nu}{2}E\left(\frac{\boldsymbol{e}^{\top}\psi_i \boldsymbol{e}\boldsymbol{e}^{\top}\psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right).
\end{aligned}
$$

Also, we need to calculate

$$
\begin{aligned}
& -\left(\frac{\partial \log f(\boldsymbol{Y})}{\partial \phi_i}\right)\left(\frac{\partial \log f(\boldsymbol{Y})}{\partial \phi_j}\right) \\
= & -\frac{1}{4}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) \\
& +\frac{n+\nu}{4}\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{\nu\sigma^2+\Delta}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) \\
& +\frac{n+\nu}{4}\left(\frac{\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{\nu\sigma^2+\Delta}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right) - \frac{(n+\nu)^2}{4}\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{n+\nu}{4}E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{\nu\sigma^2+\Delta}\right) \\
= & \frac{n+\nu}{4}\frac{1}{n+\nu}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right) = \frac{1}{4}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right).
\end{aligned}
$$

Hence, we can get

$$
\begin{aligned}
& -E\left(\frac{\partial \log f(\boldsymbol{Y})}{\partial \phi_i}\right)\left(\frac{\partial \log f(\boldsymbol{Y})}{\partial \phi_j}\right) \\
= & -\frac{1}{4}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) - \frac{(n+\nu)^2}{4}E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) \\
& +\frac{(n+\nu)}{4}E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{\nu\sigma^2+\Delta}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) \\
& +\frac{(n+\nu)}{4}E\left(\frac{\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{\nu\sigma^2+\Delta}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right) \\
= & -\frac{1}{4}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) - \frac{(n+\nu)^2}{4}E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) \\
& +\frac{(n+\nu)}{4}\frac{1}{n+\nu}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) \\
& +\frac{(n+\nu)}{4}\frac{1}{n+\nu}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right) \\
= & \frac{1}{4}\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right)\operatorname{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) - \frac{(n+\nu)^2}{4}E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}\boldsymbol{e}^\top \psi_j \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) \\
= & \frac{1}{(n+\nu)(n+\nu+2)}\Bigg[\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right)\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right) \\
& +2\mathrm{tr}\left(\psi_i\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_j}\right)\Bigg].
\end{aligned}
$$

Therefore, the elements of Fisher information matrix can be obtained by

$$
\begin{aligned}
\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -(n+\nu)E\left(\frac{-\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{\nu\sigma^2+\Delta}+\frac{2\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}\boldsymbol{e}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{(\nu\sigma^2+\Delta)^2}\right) \\
&= -(n+\nu)\left(\frac{-\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{\sigma^2(n+\nu)}+\frac{2\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{C}_n(\boldsymbol{\phi})\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{\sigma^2(n+\nu)(n+\nu+2)}\right) \\
&= \frac{(\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X})(n+\nu+2)-2\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}}{\sigma^2(n+\nu+2)} \\
&= \frac{\nu+n}{\sigma^2\,(\nu+n+2)}\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{X}, \\
\boldsymbol{I}_{\boldsymbol{\beta}\sigma^2} &= \nu(n+\nu)E\left(\frac{\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right)=\boldsymbol{0}, \\
\boldsymbol{I}_{\boldsymbol{\beta}\phi_i} &= (n+\nu)E\left(\frac{\boldsymbol{X}^\top \psi_i \boldsymbol{e}}{\nu\sigma^2+\Delta}-\frac{\boldsymbol{X}^\top \boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\boldsymbol{e}\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right)=\boldsymbol{0}, \\
\boldsymbol{I}_{\sigma^2\phi_i} &= \frac{\nu(n+\nu)}{2}E\left(\frac{\boldsymbol{e}^\top \psi_i \boldsymbol{e}}{(\nu\sigma^2+\Delta)^2}\right) \\
&= \frac{\nu(n+\nu)}{2}\frac{1}{\sigma^2(n+\nu)(n+\nu+2)}\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right) \\
&= \frac{\nu}{2\sigma^2(n+\nu+2)}\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial \phi_i}\right),
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{I}_{\sigma^2\sigma^2} &= -\frac{n}{2\sigma^4} + \frac{n+\nu}{2\sigma^4} E\left(\frac{\Delta}{\nu\sigma^2+\Delta} + \frac{\sigma^2\nu\Delta}{(\nu\sigma^2+\Delta)^2}\right) \\
&= -\frac{n}{2\sigma^4} + \frac{n+\nu}{2\sigma^4}\left(\frac{n}{n+\nu} + \frac{n\sigma^2\nu}{\sigma^2(n+\nu)(n+\nu+2)}\right) \\
&= \frac{n\nu}{2\sigma^4(\nu+n+2)}, \\
\boldsymbol{I}_{\phi_i\phi_j} &= \frac{1}{2(\nu+n+2)}\Bigg((\nu+n)\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\right) \\
&\quad -\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_i}\right)\mathrm{tr}\left(\boldsymbol{C}_n^{-1}(\boldsymbol{\phi})\frac{\partial \boldsymbol{C}_n(\boldsymbol{\phi})}{\partial\phi_j}\right)\Bigg).
\end{aligned}
$$

Appendix D: Derivation of the method of moment estimator of ν

For regression model with the error term $\boldsymbol{\varepsilon} \sim \mathrm{Mt}_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n, \nu)$, the p.d.f. of $\boldsymbol{\epsilon}$ is

$$
\begin{aligned}
f(\boldsymbol{\varepsilon}) &= \frac{\Gamma\big((\nu+n)/2\big)(\sigma^2)^{-n/2}}{\Gamma(\nu/2)(\pi\nu)^{n/2}}\left(1+\frac{\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}}{\nu\sigma^2}\right)^{-(\nu+n)/2} \\
&= \frac{\nu^{\nu/2}\Gamma\big((n+\nu)/2\big)}{\Gamma(\nu/2)\pi^{n/2}}\frac{1}{(\sigma^2)^{n/2}}\left(\nu+\frac{\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}}{\sigma^2}\right)^{-(\nu+n)/2}.
\end{aligned}
$$

The second and the fourth moments of ε_i are

$$
E(\varepsilon_i^2) = \frac{\nu\sigma^2}{\nu-2},
$$

and

$$
E(\varepsilon_i^4) = \frac{3\nu^2\sigma^4}{(\nu-2)(\nu-4)}, \quad i = 1, \ldots, n.
$$

Define

$$
m_1 = E(n^{-1}\sum_{i=1}^{n}\varepsilon_i^2) = E\Big(n^{-1}\sum_{i=1}^{n}(y_i - x_i^{\mathrm{T}}\boldsymbol{\beta})^2\Big) = \frac{\nu\sigma^2}{\nu-2},
$$

and

$$
m_2 = E\Big(n^{-1}\sum_{i=1}^{n}\varepsilon_i^4\Big) = E\Big(n^{-1}\sum_{i=1}^{n}(y_i - x_i^{\top}\boldsymbol{\beta})^4\Big) = \frac{3\nu^2\sigma^4}{(\nu-2)(\nu-4)}.
$$

Hence with $a = m_2/m_1^2$, we get $a = 3(\nu-2)/(\nu-4)$. Since the usual least square estimator of $\boldsymbol{\beta}$, given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y},$$

is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$, it is reasonable to estimate a by

$$\hat{a} = \frac{E(n^{-1}\sum_{i=1}^{n}(y_i - x_i^{\top}\hat{\boldsymbol{\beta}})^4)}{\left(E(n^{-1}\sum_{i=1}^{n}(y_i - x_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})^2)\right)^2} = \frac{n^{-1}\sum_{i=1}^{n}(y_i - x_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})^4}{\left(n^{-1}\sum_{i=1}^{n}(y_i - x_i^{\mathrm{T}}\hat{\boldsymbol{\beta}})^2\right)^2}.$$

Having given estimator $\hat{a}$ of a, the degree of freedom parameter ν can be estimated by

$$\hat{\nu} = \frac{2(2\hat{a}-3)}{\hat{a}-3}.$$

Appendix E: The conditional variance of the MVT-AR(p) process

Let $\boldsymbol{Y}_{n+1} = \begin{bmatrix} \boldsymbol{Y}_n \\ y_{n+1} \end{bmatrix} \sim \mathrm{Mt}_{n+1}\left(\boldsymbol{\mu}_{n+1}, \sigma^2\boldsymbol{\Gamma}_{n+1}, \nu\right)$ and let $\boldsymbol{\mu}_{n+1}$ and $\boldsymbol{\Gamma}_{n+1}$ be partitioned as

$$\boldsymbol{\mu}_{n+1} = \begin{bmatrix} \boldsymbol{\mu}_n \\ \mu_{n+1} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\Gamma}_{n+1} = \begin{bmatrix} \boldsymbol{\Gamma}_n & \boldsymbol{\gamma}_n \\ \boldsymbol{\gamma}_n^{\mathrm{T}} & \gamma_{n+1} \end{bmatrix}.$$

Define $\mu_{2\cdot 1} = \mu_{n+1} + \boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}(\boldsymbol{Y}_n - \boldsymbol{\mu}_n)$ and $\sigma_{22\cdot 1} = \sigma^2(\gamma_{n+1} - \boldsymbol{\gamma}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_n) = \sigma^2$. By Proposition 2,

$$y_{n+1}|\boldsymbol{Y}_n \sim t\left(\mu_{2\cdot 1},\ \frac{\sigma^2\nu + \boldsymbol{Y}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{Y}_n}{n+\nu},\ n+\nu\right).$$

Then,

$$\sigma_{n+1}^2 = \mathrm{Var}(y_{n+1}|\boldsymbol{Y}_n) = \frac{\sigma^2\nu + \boldsymbol{Y}_n^{\mathrm{T}}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{Y}_n}{n+\nu-2}.$$

Consider the inverse of a partition of $\boldsymbol{\Gamma}_n$:

$$\boldsymbol{\Gamma}_n^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}_{n-1} & \boldsymbol{\gamma}_{n-1} \\ \boldsymbol{\gamma}_{n-1}^{\mathrm{T}} & \gamma_n \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}_{n-1}^{-1} + \kappa a a^{\top} & -\kappa a \\ -\kappa a^{\top} & \kappa \end{bmatrix},$$

where $a = \mathbf{\Gamma}_{n-1}^{-1}\gamma_{n-1}$ and $\kappa = (\gamma_n - \gamma_{n-1}^{\mathrm{T}}\mathbf{\Gamma}_{n-1}^{-1}\gamma_{n-1})^{-1}$. Letting $\boldsymbol{Y}_{n-1} = (Y_{n-1}, \ldots, Y_1)$, we have

$$\boldsymbol{Y}_n^{\mathrm{T}}\mathbf{\Gamma}_n^{-1}\boldsymbol{Y}_n = \boldsymbol{Y}_{n-1}^{\mathrm{T}}\mathbf{\Gamma}_{n-1}^{-1}\boldsymbol{Y}_{n-1} + \kappa(a^{\mathrm{T}}\boldsymbol{Y}_{n-1} - Y_n)^2.$$

After some algebra, we obtain the recursion

$$\begin{aligned}
\sigma_{n+1}^2 &= \frac{\sigma^2\nu + \boldsymbol{Y}_n^{\mathrm{T}}\mathbf{\Gamma}_n^{-1}\boldsymbol{Y}_n}{n+\nu-2} \\
&= \frac{\sigma^2\nu + \boldsymbol{Y}_{n-1}^{\mathrm{T}}\mathbf{\Gamma}_{n-1}^{-1}\boldsymbol{Y}_{n-1} + \kappa(a^{\mathrm{T}}\boldsymbol{Y}_{n-1} - Y_n)^2}{n+\nu-2} \\
&= \frac{\sigma^2\nu + \boldsymbol{Y}_{n-1}^{\mathrm{T}}\mathbf{\Gamma}_{n-1}^{-1}\boldsymbol{Y}_{n-1}}{n+\nu-2} + \frac{\kappa(a^{\top}\boldsymbol{Y}_{n-1} - Y_n)^2}{n+\nu-2} \\
&= \frac{n+\nu-3}{n+\nu-2}\sigma_n^2 + \frac{\kappa(a^{\top}\boldsymbol{Y}_{n-1} - Y_n)^2}{n+\nu-2} \\
&= \frac{n+\nu-3}{n+\nu-2}\sigma_n^2 + \frac{\kappa}{n+\nu-2}(a^{\top}\boldsymbol{Y}_{n-1} - Y_n)^2.
\end{aligned}$$

Thus, this form is explicitly more general than the GARCH(n,1) model.

References

Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis* (Wiley, 3rd. ed.).

Arellano-Valle, R.B., Galea-Rojas, M., Zuazola P.I. (2000), Bayesian sensitivity analysis in elliptical linear regression models. *J Statist. Plan. Inference* **86**, 175-199.

Barndorff-Nielsen, O. E., Schou, G. (1973), On the reparameterisation of autoregressive models by partial autocorrelations. *J. Multivariate. Anal.* **3**, 408-419.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-27.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1994), *Time Series Analysis Forecasting and Control* (Holden-Day, San Francisco, 3rd ed.).

Brooks, S. P. and Gelman, A. (1998), General methods for monitoring convergence of iterative simulations. *J. Comp. Graph. Statist.* **7**, 434-455.

Chen, C.W.S., Lee, J.C., Lee, S.Y., Niu, W.F. (2004), Bayesian estimation for time series regressions improved with exact likelihoods, *J. Statist. Comput. Simulation* **74**, 727 - 740

Chib, S. (1993), Bayes regression with autoregressive errors: A Gibbs sampling approach. *J. Econometrics.* **58**, 275-294.

Chib, S., Osiewalski, J., Steel M.F.J. (1991), Posterior inference on degrees of freedom parameters in multivariate-t regression models. *Econom Lett.* **37**, 391-397.

Chib, S., Tiwari, R.C., JAMMLAMADAKA (1988), Bayes prediction in regression with elliptical errors. *J. Econometrics.* **38**, 349-360.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2004). *Bayesian Data Analysis* (Chapman & Hall, New York, 2nd ed.)

Kim, H.M., Mallick B.K. (2003), A note on Bayesian spatial prediction using the elliptical distribution. *Statist. Probab. Lett.* **64**, 271-276.

Ljung, G.M., Box, G.E.P. (1980), Analysis of variance with autocorrelated observations. *Scand. J. Statist.* **7**, 172-180.

Monahan, J.F. (1984), A note on enforcing stationarity in autoregressive moving average models. *Biometrika* **71**, 403-404.

Lee, J.C., Wang, R.S., Lin, T.I. (2004), On the inverse of the autocorrelation matrix for an AR(p) process. *J. Chinese Statist. Assoc.* **42**, 81-89.

McCulloch, R.E., Tsay R.S. (1994), Bayesian analysis of autoregressive time series via the Gibbs sampler. *J. Time Ser. Anal.* **15**, 235-250.

Osiewalski, J. (1991), A note on Bayesian inference in a regression model with elliptical errors. *J. Econometrics* **48**, 183-193.

Osiewalski, J., Steel M.F.J. (1993), Robust Bayesian inference in elliptical models. *J. Econometrics* **57**, 345-363.

Singh, R.S. (1988), Estimation of error variance in linear regression models with errors having multivariate student-t distribution with unknown degrees of freedom. *Econom Lett.* **27**, 47-53.

Tarami, B., Pourahmadi M. (2003), Multi-variate t autoregressions: innovations, prediction variances and exact likelihood equation. *J. Time Ser. Anal.* **24**, 739-754.

Zellner, A. (1976), Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *J. Amer. Statist. Assoc.* **71**, 400-405.