Hitotsubashi University

Global COE Hi-Stat Discussion Paper Series 248

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Dirichlet Prior for Estimating Unknown Regression Error Heteroscedasticity

Hiroaki Chigira
Tsunemasa Shiba

October 2012

Hi-Stat Discussion Paper

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Dirichlet Prior for Estimating Unknown Regression Error Heteroscedasticity

Hiroaki Chigira*and Tsunemasa Shiba†

October 11, 2012

Abstract

We propose a Bayesian procedure to estimate heteroscedastic variances of the regression error term $\boldsymbol{\omega}$, when the form of heteroscedasticity is *unknown.* The prior information on $\boldsymbol{\omega}$ is based on a Dirichlet distribution, and in the Markov Chain Monte Carlo sampling, its proposal density parameters' information is elicited from the well-known Eicker–White Heteroscedasticity Consistent Variance–Covariance Matrix Estimator. We present a numerical example to show that our scheme works.

key words

Dirichlet prior, Eicker–White HCCM, informative prior pdf's, MCMC,

JEL Classification

C11, C13

*Department of Economics, Tohoku University, 27-1, Kawauchi, Aoba-ku, Sendai-shi 980-8576 Japan. e-mail: hchigira@econ.tohoku.ac.jp Tel & fax: +81-22-795-6313.

†Graduate School of Finance, Accounting &Law, Waseda University, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, e-mail: tshiba@waseda.jp and Department of Economics, Hitotsubashi University, Kunitachi-shi, Tokyo 186-8601 Japan. e-mail: tshiba@stat.hit-u.ac.jp

Dirichlet Prior for Estimating Unknown Regression Error Heteroscedasticity

10/2012

Hiroaki Chigira
Tsunemasa Shiba

1 Introduction

In this paper, we shall propose an approach to estimating heteroscedastic parameters of regression error variances that are of unknown form, using Dirichlet prior pdf in a Bayesian inference. As pointed out by Amemiya (1985, p.199), the crucial $\boldsymbol{\omega}$ vector[1] cannot be consistently estimated because as the number of parameters increases, the sample size also increases at the same rate, leading to the lack of identifiability of $\boldsymbol{\omega}$. In asymptotics framework, Eicker (1963) and White (1980) independently developed a well-known consistent variance-covariance matrix estimator ("HCCM" hereafter) for the OLS regression coefficient estimator. The methodology we propose in this paper is a Bayesian that uses information obtained from the HCCM, in terms of a proposal density of a Metropolis-Hastings ("M-H" hereafter) algorithm in Markov Chain Monte Carlo simulation. The lack of identifiability of $\boldsymbol{\omega}$ poses no problem. For one thing, as in Amemiya (1985) we use an orthogonal regression that circumvents possible underidentifiability of $\boldsymbol{\omega}$, and we shall explain this method in detail later[2]. Second, we impose a prior on $\boldsymbol{\omega}$ so that the vector becomes identifiable in a Bayesian context.

[1]The $\boldsymbol{\omega}$ vector has in its elements, all the *normalized* diagonal elements of variance-covariance matrix of the regression error term. The normalization rule for the matrix is given just below equation (1).

[2]Amemiya *op cit* proposes and uses an orthogonal regression to obtain a better performing GLS.

The trend in the HCCM literature seems to be how to improve the finite sample performance of tests of the linear restriction(s) on the coefficient vector, *e.g.,* Long and Ervin(2000) and Godfrey (2006), among others. We note that our focus in this paper is in the direct estimation of the elements of $\boldsymbol{\omega}$. There are papers that deal with statistical inferences of regression coefficients, when the scedastic function of the error term is unconstrained. Robinson (1987), for example, assumes it to be a function of regressors, and derives an GLS estimator that is more efficient than the existing ones. Our Bayesian estimation of heteroscedasticity should sharpen posterior density of regression coefficient vector $\boldsymbol{\beta}$ and/or lead to a better predictive density. It may also lead to more efficient estimator of $\boldsymbol{\beta}$ in terms of asymptotic theory framework as well.

We need to discuss the direct estimation of the $\boldsymbol{\omega}$ vector. In financial returns, $\boldsymbol{\omega}$ is nothing but the volatility. In order to access an option pricing, what we need to do first is to come up with a reasonable estimate of volatility. Our estimation of $\boldsymbol{\omega}$ needs no parametric model for the volatility process such as the GARCH model, since we use information obtained from the HCCM estimation, in our MCMC simulation. If we wish to estimate a volatility process in time series data nonparametrically, what we usually do is to calculate a historical volatility series. But this is just a descriptive statistic without a theoretical background. Moreover, when it comes to cross section data, historical volatility calculation breaks down for obvious reasons. Our Bayesian method, on the other hand, should provide a good deal of theoretical support for cross sectional data.

Our strategy to estimate the $\boldsymbol{\omega}$ vector is Bayesian. After assuming a usual prior density for the parameters in the regression model, we are able to write down a joint posterior density. The usual parameters such as regression coefficients, $\boldsymbol{\beta}$, may be easily simulated using the Gibbs sampler scheme. It is in the simulation of the elements of $\boldsymbol{\omega}$ that we use the HCCM. We use results from HCCM to form the candidate density in the M-H algorithm.

The rest of this paper is organized as follows. In section 2, we set our regression model. Prior pdf's are assumed here, and the joint posterior pdf is derived. Section 3 starts out with our Bayesian MCMC calculation by a Gibbs sampler. We propose to use the Eicker–White result to simulate $\boldsymbol{\omega}$ by a M-H scheme. Our numerical illustration is very much limited. Section 5 concludes.

2 Model and Assumptions

2.1 Likelihood

We first give a heteroscedastic NLR (normal linear regression) model, as follows:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \quad (i = 1, \ldots, n) \tag{1}$$

where $y_i \sim$ dependent variable, $\boldsymbol{x}_i \sim K \times 1$ non stochastic explanatory variables, $\boldsymbol{\beta} \sim K \times 1$ coefficients, and the properties of regression error term u will be given below. This equation (1), with the assumption on the disturbance term, u_i,

$$u_i|\omega_i, \sigma^2 \ \sim N(0, \sigma^2\omega_i) \tag{2}$$

our single likelihood function for y_i has the following distribution

$$y_i| \ \boldsymbol{x}_i, \boldsymbol{\beta}, \omega_i, \sigma^2 \ \sim N(\boldsymbol{x}_i'\boldsymbol{\beta}, \ \sigma^2\omega_i) \ , \tag{3}$$

where $\omega_i \;\; (i = 1, \ldots, n)$ is the heteroscedastic variance parameter for u_i.

We may switch around the following two notations: $\omega_i = \dfrac{1}{\lambda_i}$, where λ_i refers to the precision parameter. Geweke (1993, 2005) in particular, uses the λ_i notation. Greenberg (2008) also.

Let us now present (1) in an matrix form.

$$\boldsymbol{y} \;=\; \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \;, \tag{4}$$

where $\boldsymbol{y} \sim n \times 1$ of y_i's, $\boldsymbol{X} \sim n \times K$ matrix of stacked up $\boldsymbol{x}_i'$, $\boldsymbol{u} \sim N_n(\boldsymbol{0},\, \sigma^2\boldsymbol{\Omega})$, and $\boldsymbol{\Omega} = diag(\boldsymbol{\omega}) = diag(\omega_1, \ldots, \omega_n) \sim n \times n$ with $\sum_{i=1}^{n}\omega_i = tr(\boldsymbol{\Omega}) = n$. We may note that $tr(\boldsymbol{\Omega}) = n$ restriction is often employed in heteroscedasticity literature, *e.g.,* Greene (2012, p 308). Our likelihood function for the entire sample of size n given (3) becomes

$$\ell(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta}) \propto \sigma^{-n}(\prod_{i=1}^{n} \omega_i^{\frac{-1}{2}}) \exp\left(\frac{-1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) \;, \tag{5}$$

where we noted $|\sigma^2\boldsymbol{\Omega}|^{\frac{-1}{2}} \;=\; \sigma^{-n}\prod_{i=1}^{n}\omega_i^{\frac{-1}{2}}$. On completing squares on $\boldsymbol{\beta}$ above, we obtain another likelihood expression as

$$\ell(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta}) \propto \sigma^{-n}(\prod_{i=1}^{n} \omega_i^{\frac{-1}{2}}) \exp\left(\frac{-1}{2\sigma^2}(\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right), \tag{6}$$

where $\nu = n - K$, νs^2 is the sum of squared residuals from the regression of $\tilde{\boldsymbol{y}}$ on $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{y}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{y}$, $\tilde{\boldsymbol{X}} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}$, and $\hat{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$ is the GLS estimator of $\boldsymbol{\beta}$. (6) turns out to be useful for simulating $\boldsymbol{\beta}$ since it is in a Multivariate Normal form in $\boldsymbol{\beta}$.

2.2 Prior to Posterior

Let our parameter vector be $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2, \boldsymbol{\omega}')' \sim (K + 1 + n) \times 1$. Our assumption on $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}) \;\propto\; \pi(\boldsymbol{\beta})\,\pi(\sigma^2)\,\pi(\boldsymbol{\omega})\,, \tag{7}$$

where the three arguments of $\boldsymbol{\theta}$ are independent in $\pi(\boldsymbol{\theta})$. We postulate a set of priors for $\boldsymbol{\theta}$ starting with $\boldsymbol{\beta}$ and σ^2 as follows:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \boldsymbol{B}_0), \quad \sigma^2 \sim \mathcal{IG}(\frac{\alpha_0}{2}, \frac{\delta_0}{2}), \tag{8}$$

where $\boldsymbol{\beta}_0 \sim K \times 1$, $\boldsymbol{B}_0 \sim K \times K$, α_0 and δ_0 are hyper parameters in the prior pdf's that are assumed to be *known*, and $\mathcal{IG}(\cdot)$ denotes an inverted gamma distribution.

It should be noted that our prior for $\boldsymbol{\omega}$, needs to satisfy the $tr(\boldsymbol{\Omega}) = \sum_i \omega_i = n$ restriction. The best suited prior to this regard, is obviously Dirichlet with its hyper parameter values the same for all $i = 1, \ldots, n$. This way, we may effectively represent unknown heteroscedasticity structure and the needed restriction that the elements add up to one, at the same time. If we employed Gamma prior, *e.g.,* Geweke (1993, 2005) and Greenberg (2008), among others, then we are in effect imposing a certain structure in the heteroscedasticity.

Since a Dirichlet has the property that its elements add up to one, not n, we cannot place a Dirichlet prior directly on $\boldsymbol{\omega}$. Instead, we shall assume

$$\tilde{\boldsymbol{\omega}} \sim D(\boldsymbol{\eta}), \tag{9}$$

where $\tilde{\boldsymbol{\omega}} = \dfrac{1}{n}\boldsymbol{\omega}$, and $D(\boldsymbol{\eta})$ denotes a Dirichlet distribution with a parameter vector $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)' \sim n \times 1$. The values of $\boldsymbol{\eta}$ will be given later in this paper.

Note that this assumption on $\tilde{\boldsymbol{\omega}}$ implies $tr(\tilde{\boldsymbol{\Omega}}) = 1$, thus $tr(\boldsymbol{\Omega}) = n$. If we make a transformation from $\tilde{\boldsymbol{\omega}}$ vector to $\boldsymbol{\omega}$ vector, we should arrive at our prior distribution on $\boldsymbol{\omega}$ that resembles to $D(\boldsymbol{\eta})$ aside from normalizing constant.

$$\pi(\boldsymbol{\omega}) \;=\; \frac{1}{n}\frac{\Gamma\left(\sum_i \eta_i\right)}{\prod_{j=1}^{n}\Gamma(\eta_j)}\prod_{h=1}^{n}\left(\frac{\omega_h}{n}\right)^{\eta_h-1} \;=\; n^{n-1-\sum_h \eta_h}\frac{\Gamma\left(\sum_i \eta_i\right)}{\prod_{j=1}^{n}\Gamma(\eta_j)}\prod_{\ell=1}^{n}\omega_\ell{}^{\eta_\ell-1}, \quad (10)$$

thus its kernel is given by

$$\pi(\boldsymbol{\omega}) \;\propto\; \prod_{i=1}^{n}\omega_i^{\eta_i-1}. \quad (11)$$

Given (8) and (11), our joint prior for $\boldsymbol{\theta}$ becomes

$$\pi(\boldsymbol{\theta}) \propto \exp\left(\frac{-1}{2}q_\beta\right)\left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_0}{2}+1}\exp\left(\frac{-\delta_0}{2\sigma^2}\right)\prod_{h=1}^{n}\tilde{\omega}_h^{\eta_h-1}, \quad (12)$$

where $q_\beta = (\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{B}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$. Finally our joint posterior of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X})$, is obtained by combining (6) and (12) as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X}) \;\propto\; \ell(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

to obtain

$$\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X}) \propto \sigma^{-n}(\prod_{i=1}^{n}\omega_i^{\frac{-1}{2}})\exp\left(\frac{-\psi}{2\sigma^2}\right)\exp\left(\frac{-1}{2}q_\beta\right)\left(\sigma^2\right)^{-(\frac{\alpha_0}{2}+1)}\exp\left(\frac{-\delta_0}{2\sigma^2}\right)\prod_{h=1}^{n}\tilde{\omega}_h^{\eta_h-1}, \quad (13)$$

where $\psi = \nu s^2 + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})$. Note that so far as $\boldsymbol{\theta}$ is concerned, ψ depends on $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$, while q_β depends on $\boldsymbol{\beta}$.

2.3 What is the difference?

We, now, point out a salient feature of our model that distinguishes it from a class of heteroscedastic NLR model of Geweke(1996, 2005), Greenberg (2008), among others. In an earlier paper, Chigira and Shiba (2012), we showed that a heteroscedastic NLR model $u_i|\ \lambda_i, \sigma^2 \sim N(0,\ \sigma^2/\lambda_i)$ with a set of nonconjugate mutually independent priors on $\boldsymbol{\theta}$ usually boils down to $u_i|\ \sigma^2 \sim$ unknown distribution, although it resembles to a Student-t[3] . If we further assume a special prior on λ_i, *viz.* a Gamma with an identical shape and scale parameters, $\nu = \nu_1 = \nu_2$, then the resultant $u_i|\ \sigma^2$ distribution becomes a fat-tailed homoscedastic Student-t with the parameters $(\nu, 0, \sigma^2)$. We, here note that in both cases, *viz.* two ν_i parameters' case and a single ν parameter case, Bayesian analysis may be carried out. In particular, single ν case does not require any MCMC simulation.

The above peculiar result is sometimes construed as a model with Normal heteroscedastic error term is equivalent to a model with homoscedastic fat-tailed Student-t error term. We caution the reader that this interpretation is valid only when we assume a very special prior pdf on the precision parameter, λ_i. This is something akin to obtaining an analytically manageable posterior (in the present case, homoscedastic model) using a set of natural conjugate priors (in the present case, Gamma prior with identical parameters). If there is a compelling need for such a peculiar prior on λ_i, then the above equivalence seems to be of a great value between:

Normally distributed error ==> heteroscedasticity

Student-t fat-tail error ==> homoscedasticity.

What would the effect be if there is only one parameter value $\nu_1 = \nu_2 = \nu$ in the Gamma distribution? This is answered in the Appendix below. Such Gamma random variable would have $E(x) = 1$ always, and its pdf becomes

[3] Only in this subsection, $\boldsymbol{\lambda}$ replaces $\boldsymbol{\omega}$ in the definition of the entire parameter vector, $\boldsymbol{\theta}$.

concentrated abound it as ν gets large. We may thus conclude that the one parameter Gamma distribution assumption is rather peculiar.

In view of the above conclusion, we suggest that we depart from $\lambda_i \sim$ Gamma distribution assumption, and adopt more reasonable prior for u_i heteroscedasticity. Notice that if the Gamma prior $G(\frac{\nu_1}{2}, \ \frac{\nu_2}{2})$, on λ_i is assumed, then n λ_i's are assumed to be generated from one single Gamma. This is in effect assuming a particular structure on the heteroscedasticity of u_i's. If the interest of the Bayesian analysis, is in finding the parameter of the structure, $\nu = \nu_1 = \nu_2$, then the Gamma distribution assumption may be justified. But if we are interested in estimating each $\sigma_i^2 = \sigma^2/\lambda_i = \sigma^2\omega_i$ then we need something other than the Gamma assumption. As we developed in the previous subsection, we employ a Dirichlet prior to this end.

Dirichlet prior on λ_i or ω_i is suitable for Bayesian estimation of heteroscedastic variance parameters that have unknown structure, on two grounds. First, it is free of restrictions from the small number of parameters that governs the entire shape of the prior pdf. In Geweke's Gamma pdf prior for $\boldsymbol{\lambda} \sim n \times 1$, ν is the only parameter of the distribution. If the prior pdf of the $\boldsymbol{\lambda}$ vector were to be of unknown structure, it should have n parameters. Secondly, Dirichlet distributed random variables, x_i's, satisfy the $\sum_{i=1}^{n} x_i = 1$ constraint by construction. This is a welcome constraint to our setup, where $\sum \omega_i = n$ restriction, needs to be satisfied *a priori.*

3 MCMC Simulation of $\boldsymbol{\theta}$

We use notations such as "$\boldsymbol{\theta}_{-\vartheta}$." For instance, "$\boldsymbol{\theta}_{-\boldsymbol{\beta}}$" implies $\boldsymbol{\theta}_{-\boldsymbol{\beta}} = (\sigma^2, \boldsymbol{\omega}')' \sim (1+n) \times 1$ and so on.

3.1 Gibbs Sampler for $\boldsymbol{\beta}$ and σ^2

As shown in the two remarks below, tractable fully conditional posteriors of $\boldsymbol{\beta}$ and σ^2 may be obtained, thus making Gibbs sampler applicable. On the other hand we need to implement an independence chain Metropolis-Hastings algorithm for simulating $\boldsymbol{\omega}$.

Remark 1. Fully conditional posterior of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}|\, \boldsymbol{\theta}_{-\beta},\ \boldsymbol{y}, \boldsymbol{X}\ \sim N_K(\boldsymbol{\beta}_1,\ \sigma^{-2}\, \boldsymbol{B}_1)\,, \tag{14}$$

where $\boldsymbol{B}_1 = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + (\sigma^{-2}\boldsymbol{B}_0)^{-1})^{-1}$, $\boldsymbol{\beta}_1 = \boldsymbol{B}_1\boldsymbol{\varphi}$, and $\boldsymbol{\varphi} = \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}} + (\sigma^{-2}\boldsymbol{B}_0)^{-1}\boldsymbol{\beta}_0$.

Proof From the joint posterior (13), conditional posterior of $\boldsymbol{\beta}$ becomes

$$\pi(\boldsymbol{\beta}|\ \boldsymbol{\theta}_{-\beta}, \boldsymbol{y}, \boldsymbol{X})\ \propto\ \exp(\left(\frac{-\psi}{2\sigma^2}\right) \exp\left(\frac{-1}{2} q_\beta\right) \propto\ \exp(\frac{-1}{2\sigma^2}\, A_\beta)\,,$$

where $A_\beta = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\sigma^{-2}\boldsymbol{B}_0)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. On completing squares for $\boldsymbol{\beta}$, A_β becomes

$$A_\beta = (\boldsymbol{\beta} - \boldsymbol{\beta}_1)'\boldsymbol{B}_1^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_1) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'[(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1} + \sigma^{-2}\boldsymbol{B}_0]^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\,.$$

Hence,

$$\pi(\boldsymbol{\beta}|\, \boldsymbol{\theta}_{-\beta},\ \boldsymbol{y}, \boldsymbol{X})\ \propto \exp\left(-\frac{1}{2}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_1)'(\sigma^{-2}\boldsymbol{B}_1)^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_1)\right) \tag{15}$$

The right hand side of (15) may be used to simulate $\boldsymbol{\beta}$, however, there is a simpler set of expressions to that effect. Let

$$\tilde{\boldsymbol{I}}_K = (\boldsymbol{\iota} \otimes \boldsymbol{I}_K) \sim K \times 2K, \quad \boldsymbol{b} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \boldsymbol{\beta}_0 \end{pmatrix} \sim 2K \times 1,$$

and $\boldsymbol{Q} = diag(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}},\, (\sigma^{-2}\boldsymbol{B}_0)^{-1}) \sim 2K \times 2K$ block diagonal, then we have

$$\boldsymbol{B}_1^{-1} = \tilde{\boldsymbol{I}}_K\boldsymbol{Q}\tilde{\boldsymbol{I}}, \quad \boldsymbol{\varphi} = \tilde{\boldsymbol{I}}_K\boldsymbol{Q}\boldsymbol{b}, \quad \boldsymbol{\beta}_1 = \boldsymbol{B}_1\boldsymbol{\varphi}\ .$$

Remark 2. Fully conditional posterior of σ^2 is given by

$$\sigma^2 \mid \boldsymbol{\theta}_{-\sigma^2},\, \boldsymbol{y},\, \boldsymbol{X} \;\sim \mathcal{IG}\left(\frac{n+\alpha_0}{2},\; \frac{\psi+\delta_0}{2}\right) \tag{16}$$

Proof From the joint posterior (13), conditional posterior of σ^2 becomes

$$\begin{aligned}\pi(\sigma^2 |\, \boldsymbol{\theta}_{-\sigma^2},\, \boldsymbol{y},\, \boldsymbol{X}) \;&\propto \sigma^{-n} \exp\left(-\frac{\psi}{2\sigma^2}\right)(\sigma^2)^{-\left(\frac{\alpha_0}{2}+1\right)} \exp\left(\frac{-\delta_0}{2\sigma^2}\right)\\ &\propto (\sigma^2)^{-\left(\frac{n+\alpha_0}{2}+1\right)} \exp\left(-\frac{(\psi+\delta_0)}{2\sigma^2}\right) \;.\end{aligned}$$

3.2 Independence Chain for $\boldsymbol{\omega}$

We now turn to $\boldsymbol{\omega}$ simulation. From the joint posterior, (13), we have

$$\pi(\boldsymbol{\omega} |\, \boldsymbol{\theta}_{-\omega},\, \boldsymbol{y}, \boldsymbol{X}) \;\propto\; (\prod_{i=1}^{n} \omega_i^{(\eta_i^*-1)}) \exp\left(\frac{-\psi}{2\sigma^2}\right)\,, \tag{17}$$

where $\eta_i^* = \eta_i - \frac{1}{2} > 0$ for $i = 1, .., n$ in order for $\boldsymbol{\eta}^*$ vector to make sense as a Dirichlet parameter. Let

$$A_\omega = \prod_{i=1}^{n} \omega_i^{(\eta_i^*-1)} \quad \text{and} \quad B_\omega = \exp\left(\frac{-\psi}{2\sigma^2}\right)$$

hence $\pi(\boldsymbol{\omega} |\, \boldsymbol{\theta}_{-\omega},\, \boldsymbol{y}, \boldsymbol{X}) = \; A_\omega\, B_\omega$. Obviously A_ω is a kernel of $D(\boldsymbol{\eta}^*)$. On the other hand B_ω certainly looks like a $N_n(\boldsymbol{X\beta},\, \sigma^2\boldsymbol{\Omega})$, however, as a kernel of $\boldsymbol{\omega}$, B_ω is not of any known form.

We shall use an Independence Chain M-H simulator for $\boldsymbol{\omega}$. Since B_ω is not going to give any clue for a proposal density, we use information contained in $A_\omega \sim$ Dirichlet distribution, for our proposal density. Particular value of the parameter vector in the proposal density, will be discussed later. In the following, we shall first give an outline of our M-H strategy.

3.2.1 Acceptance Probability of M-H on $\boldsymbol{\omega}$

Let the "rth" current value in the chain be $\boldsymbol{\omega}^{(r)}$, and suppose that we need to decide whether to accept $\boldsymbol{\omega}^{(\prime)}$ or not. Then the acceptance probability of accepting $\boldsymbol{\omega}^{(\prime)}$, given $\boldsymbol{\omega}^{(r)}$ would be

$$\alpha(\boldsymbol{\omega}^{(r)},\, \boldsymbol{\omega}^{(\prime)}) = \min\left(1,\, \frac{\pi(\boldsymbol{\omega}^{(\prime)},\, \boldsymbol{\theta}_{-\omega}|\, \boldsymbol{y}, \boldsymbol{X})}{\pi(\boldsymbol{\omega}^{(r)},\, \boldsymbol{\theta}_{-\omega}|\, \boldsymbol{y}, \boldsymbol{X})}\, \frac{f(\boldsymbol{\omega}^{(r)})}{f(\boldsymbol{\omega}^{(\prime)})}\right),$$

where $f(\boldsymbol{\omega})$ represents our proposal density for $\boldsymbol{\omega}$, and it is explained in below. We first take up the ratio of posterior densities in the acceptance probability. Noting that arguments other than $\boldsymbol{\omega}$ cancels out, it becomes

$$\frac{\pi(\boldsymbol{\omega}^{(\prime)},\, \boldsymbol{\theta}_{-\omega}|\, \boldsymbol{y}, \boldsymbol{X})}{\pi(\boldsymbol{\omega}^{(r)},\, \boldsymbol{\theta}_{-\omega}|\, \boldsymbol{y}, \boldsymbol{X})} = \frac{\prod_i \left(\omega_i^{(\prime)}\right)^{\eta_i^*-1}}{\prod_j \left(\omega_j^{(r)}\right)^{\eta_j^*-1}}\, \frac{\exp\left(\dfrac{-\psi^{(\prime)}}{2\sigma^2}\right)}{\exp\left(\dfrac{-\psi^{(r)}}{2\sigma^2}\right)}, \tag{18}$$

where $\psi^{(\prime)} = (\boldsymbol{y} - \boldsymbol{X\beta})'\left(\boldsymbol{\Omega}^{(\prime)}\right)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$, $\boldsymbol{\Omega}^{(\prime)} = diag(\boldsymbol{\omega}^{(\prime)})$ and this is not to be confused with a transpose of $\boldsymbol{\Omega}$. Likewise, $\psi^{(r)} = (\boldsymbol{y} - \boldsymbol{X\beta})'\left(\boldsymbol{\Omega}^{(r)}\right)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$, $\boldsymbol{\Omega}^{(r)} = diag(\boldsymbol{\omega}^{(r)})$. Suppose that we employ $D(\boldsymbol{\eta})$ as our proposal[4] , then the ratio of proposal densities above becomes,

$$\frac{f(\boldsymbol{\omega}^{(r)})}{f(\boldsymbol{\omega}^{(\prime)})} = \frac{\pi(\boldsymbol{\omega}^{(r)})}{\pi(\boldsymbol{\omega}^{(\prime)})} = \frac{\prod_i \left(\omega_i^{(r)}\right)^{\eta_i-1}}{\prod_j \left(\omega_j^{(\prime)}\right)^{\eta_j-1}}. \tag{19}$$

Results of (18) and (19) may be put together to produce

$$\alpha(\boldsymbol{\omega}^{(r)},\, \boldsymbol{\omega}^{(\prime)}) = \min\left(1,\, \frac{\exp\left(\dfrac{-\psi^{(\prime)}}{2\sigma^2}\right)}{\exp\left(\dfrac{-\psi^{(r)}}{2\sigma^2}\right)}\, \frac{\prod_i \left(\omega_i^{(r)}\right)^{\eta_i-\eta_i^*}}{\prod_j \left(\omega_j^{(\prime)}\right)^{\eta_j-\eta_j^*}}\right). \tag{20}$$

3.2.2 Proposal Density Parameter

So far, we have just said that our proposal density is Dirichlet with a known parameter vector. Let this parameter vector be "$\boldsymbol{\eta}^\circ$." In the numerical experiments below, we have let $\boldsymbol{\eta}^\circ = c\,\hat{\boldsymbol{\eta}}$, where "$c$" is a tuning constant,

[4]The details of our proposal density are given later.

and $\hat{\boldsymbol{\eta}}$ is obtained from information that White's HCCM (Heteroscedasticity Consistent Covariance Matrix) estimator provides.

Let us briefly describe how we obtain $\hat{\boldsymbol{\eta}}$.[5] We first regress $\boldsymbol{y}$ on $\boldsymbol{X}$ by the OLS to obtain estimated residual vector $\boldsymbol{e}_{ols}$. We then use it to construct White's HCCM estimator, $\hat{\boldsymbol{H}}$. Let a vector obtained from $\hat{\boldsymbol{H}}$ be $\hat{\boldsymbol{h}}$, where $\hat{\boldsymbol{h}} = vech(\hat{\boldsymbol{H}}) \sim K' \times n$ and $K' = \frac{1}{2}K(K+1)$. As a regressor matrix to $\hat{\boldsymbol{h}}$, consider $\boldsymbol{\mathcal{X}}_n = [vech(\boldsymbol{x_1 x_1'}), \ldots, vech(\boldsymbol{x_n x_n'})] \sim K' \times n$. Regression of $\hat{\boldsymbol{h}}$ on $\boldsymbol{\mathcal{X}}_n$, *i.e.*,

$$\hat{\boldsymbol{h}} = \boldsymbol{\mathcal{X}}_n \sigma^2 \boldsymbol{\omega} + \boldsymbol{v}\,, \tag{21}$$

where $\boldsymbol{v}$ is an appropriate error term vector, will yield an estimator of $\boldsymbol{\omega}$, when $K' > n$ and $n > K$. When these conditions are not met, we may augment $\boldsymbol{X}$ by $\boldsymbol{W} \sim n \times K_W$ such that $\boldsymbol{W}'(\boldsymbol{y}, \boldsymbol{X}) = \boldsymbol{0}$. Since $\boldsymbol{W}$ is orthogonal to both $\boldsymbol{y}$ and $\boldsymbol{X}$, $\boldsymbol{e}_{ols}$ would be the same as before, even if we used the augmented new $\boldsymbol{X}$. In summary, we may always estimate $\boldsymbol{\omega}$ given data, $\boldsymbol{y}$ and $\boldsymbol{X}$, using (21). Letting the result from (21) be $\hat{\boldsymbol{\omega}}_{ols}$, our $\hat{\boldsymbol{\eta}}$ becomes $\hat{\boldsymbol{\omega}}_{ols}$ with some restrictions imposed on it. We shall discuss such restrictions in the paragraphs below.

Since setting proposal parameter values would be heavily dependent on data, we need to outline our numerical experiments that is given in our earlier paper (Chigira and Shiba (2009)). We set our sample size, n, to be 50. $\hat{\boldsymbol{\omega}}_{ols}$ is used to set our proposal density parameter, since we suspect that White's HCCM should contain information on $Var(\boldsymbol{u})$. In our setting $\sum \omega_i$ is restricted to be $n = 50$. Note that $\hat{\boldsymbol{\omega}}_{ols}$ need not satisfy this condition,

[5] We refer our earlier discussion paper, Chigira and Shiba (2009) for the details.

however[6]. What matters, is the relative magnitude of $\hat{\omega}_{ols\,i}$'s for $i = 1, .., n$. Hence, although "c" in $\boldsymbol{\eta}^{\circ} = c\,\hat{\boldsymbol{\eta}}$, may be used to adjust the overall level of $\boldsymbol{\eta}^{\circ}$, relative magnitude of $\hat{\boldsymbol{\eta}}$ is crucial for conveying HCCM information to proposal density.

Letting $\omega^{max} = \max(\hat{\boldsymbol{\omega}}_{ols})$ and $\omega^{min} = \min(\hat{\boldsymbol{\omega}}_{ols})$, we may set bounds for $\hat{\boldsymbol{\eta}}$ as $\max(\hat{\boldsymbol{\eta}}) = \omega^{max}$ and $\min(\hat{\boldsymbol{\eta}}) = (\omega^{max}/n)$, say. This certainly sets $\max(\hat{\boldsymbol{\eta}})/\min(\hat{\boldsymbol{\eta}}) = n$ the sample size, but the relative magnitude of $\hat{\boldsymbol{\omega}}_{ols} < \min(\hat{\boldsymbol{\eta}})$ is ignored. To rectify this situation, we add $d > 0$ to all $\hat{\omega}_{ols\,i}$ values. It can be shown that adding $d > 0$ to both the numerator and to the denominator, makes the ratio smaller:

$$\frac{\omega^{max}}{\omega^{min}} > \frac{\omega^{max} + d}{\omega^{min} + d}$$

but the ordering of values of $\hat{\boldsymbol{\omega}}_{ols} < \min(\hat{\boldsymbol{\eta}})$ is now preserved. In summary, we set

$$\hat{\eta}_i = \hat{\omega}_{ols\,i} + d$$

in practice. The actual value of "d" is arbitrary. For instance, in our numerical experiment, $n = 50$ and $d = 1$ yielded $\max(\hat{\boldsymbol{\eta}})/\min(\hat{\boldsymbol{\eta}})$ ratio of $9 << n = 50$.

4 Numerical Experiments

In the numerical experiments in this section, we employed the same Data Generating Process (DGP) as in our earlier paper. As to our prior, we used

$$\boldsymbol{\beta} \sim N_K(\boldsymbol{\beta}_0, \boldsymbol{B}_0), \text{ where } \boldsymbol{\beta}_0 = \boldsymbol{0} \text{ and } \boldsymbol{B}_0 = 10^4 \boldsymbol{I}_K.$$
$$\sigma^2 \sim \mathcal{IG}\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right), \text{ where } \alpha_0 = \delta_0 = 10^{-2}.$$

[6]Surprisingly, in our numerical experiments, $tr(\hat{\boldsymbol{\Omega}}_{ols})$ is "60.12". Fair to say that this is close enough to $n = 50$.

Prior Dirichlet parameters and proposal Dirichlet parameters are specified as follows, where prior parameter vector $\boldsymbol{\eta} = k\,\boldsymbol{\iota}_n$ and $k \sim$ scalar, $\boldsymbol{\iota}_n \sim n \times 1$ vector of all one's.

	Prior $\boldsymbol{\eta}$	$\hat{\boldsymbol{\eta}}$ in Proposal $\boldsymbol{\eta}^\circ = c\,\hat{\boldsymbol{\eta}}$
Case 1	$k = 10^{-4}$	true $\boldsymbol{\omega}$ (as in the DGP)
Cases 2 to 6	five different k's	HCCM based $\hat{\boldsymbol{\eta}}$

c in $\boldsymbol{\eta}^\circ$ is given in below. k in Cases 2 to 6 were $k = 10^{-4}$, 10^{-2}, 10^{0}, 10, 50.

Case 1: In this case, we did not need to make bounds restriction on $\hat{\boldsymbol{\eta}}$, since the ratio of maximum to minimum of $\hat{\omega}_{ols}$'s was less than n; actually 3.26 and 0.0747 yielding $43.7 < n = 50$. We let $c = 36.5$ and the acceptance rate in M-H was 68.94%. Figures 1 and 2 give $n = 50$ different posterior means of $\sigma^2\omega_i$. Figure 1, in particular, demonstrates that our Bayesian estimation procedure works, if appropriate proposal density were found. A goodness of fit measure was computed as :

$$D_{true} \;=\; (\sigma^2\boldsymbol{\omega} - \widehat{\sigma^2\boldsymbol{\omega}})'(\sigma^2\boldsymbol{\omega} - \widehat{\sigma^2\boldsymbol{\omega}}) = 0.71\,.$$

Cases 2 to 6: In these cases, $\max(\hat{\boldsymbol{\omega}}) = 7.97$ and $\min(\hat{\boldsymbol{\omega}}) = 1.157 \times 10^{-4}$, thus the ratio was $6.88 \times 10^4 >> n = 50$. We imposed a set of bounds restrictions. Adjustment scalar to be added to each $\hat{\omega}_i$ was set to be 0.162. For cases 2 to 5, $c = 14.9$ yielded acceptance rates in the range 88 to 89%, and the results seem to resemble each other. For $h = 50$, however, different c values brought quite unstable results and acceptance rate often approached to 100%. Perhaps this fact indicates that h values ought to be small.

Figures 4 to 6 are $h = 1$ case. Figure 4 seems to indicate that our posterior means tracks the true, given values, reasonably well. The goodness of fit measure was

$$D_{HCCM} \;=\; (\sigma^2\boldsymbol{\omega} - \widehat{\sigma^2\boldsymbol{\omega}})'(\sigma^2\boldsymbol{\omega} - \widehat{\sigma^2\boldsymbol{\omega}}) = 57.08\,.$$

which is quite larger than D_{true}. The absolute vale, 57.08, in itself does not render any easy interpretation. We have, thus, computed different measures for comparison purposes. First, without the bounds constraints in the proposal parameter $\hat{\boldsymbol{\eta}}$, we have

$$D_{noBoundsHCCM} \;=\; (\sigma^2\boldsymbol{\omega} - \widehat{\sigma^2\boldsymbol{\omega}})'(\sigma^2\boldsymbol{\omega} - \widehat{\sigma^2\boldsymbol{\omega}}) = 72.07\,.$$

This shows that the bounds constraints are effective. Next, let $\hat{\sigma}^2 = \boldsymbol{e}'_{ols}\boldsymbol{e}_{ols}/(n-K)$ the OLS σ^2 estimator, and $\hat{\omega}_{ols}$ be the HCCM based $\hat{\boldsymbol{h}}$ regression estimated coefficient. Then

$$D_{olsHCCM} \;=\; (\sigma^2\boldsymbol{\omega} - \hat{\sigma}^2\hat{\boldsymbol{\omega}}_{ols})'(\sigma^2\boldsymbol{\omega} - \hat{\sigma}^2\hat{\boldsymbol{\omega}}_{ols}) = 86.51\,.$$

This shows that our Bayesian procedure makes a considerable improvement over the $\hat{\boldsymbol{h}}$ regression method.

5 Concluding Remarks

In this paper, we proposed a Bayesian method to estimate regression error heteroscedasticity structure that is unknown. "Unknown" in the sense that no structure is assumed. It is well known that when Gamma priors with a particular set of hyper parameter values are assumed on the precision parameters of heteroscedastic regression error term , then this leads to a homoscedastic Student-t regression error term. We pointed out that assuming such priors, are in effect, imposing an unwanted structure in the heteroscedasticity. We have, thus, proposed to use a Dirichlet prior with equal hyper parameter values. This should represent we “know nothing” status about the structure of heteroscedasticity.

We, on the other hand, believe that the Eicker-White HCCM (Heteroscedasticity Consistent Covariance Matrix estimator) should provide valuable information about the heteroscedasticity, although derived from a sampling theory point of view. In empirical analysis, regression equation is bound to be misspecified. HCCM, in essence, draws heteroscedasticity information connecting with regressors. Our idea is to use this HCCM information in the proposal distribution in the Independence sampler. We showed that this approach is reasonably successful in a numerical experiment.

Finally we may present two tasks that need our attention. The first one is concerned with a comparison between our Dirichlet prior approach, and Geweke (1993) type Gamma prior modelling. In the Gamma prior type modelling, their interest is mainly in inferences of $\boldsymbol{\beta}$ vector and σ^2. Despite this fact, it is still possible to compute posterior means of the heteroscedastic parameters in their approach, and compare them with that of our approach.

The second one is about our use of the HCCM information. We certainly need to make sure that this information improves our result in numerical experiments. As to the design of experiments, we may result to the many DGP’s of HCCM papers.

References

[1] Amemiya, T., 1985, *Advanced Econometrics,* Harvard University Press.

[2] Chigira, H. and T. Shiba, 2012, "Heteroskedastic NLR Model and Its Student-t Interpretation," *A note prepared for presentation at the Econometrics Workshop, January 13*, Hitotsubashi University.

[3] Chigira, H. and T. Shiba, 2009, "Bayesian Estimation of Unknown Regression Error Heteroscedasticity," *Hi-Stat D.P. 051*, Hitotsubashi University.

[4] Eicker, F., 1963, Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of the Mathematical Statistics,* 34, 447–456.

[5] Greene, W.H., 2012, *Econometric Analysis, 7th ed.,* Pearson.

[6] Greenberg, E., 2008, *Introduction to Bayesian Econometrics,* Cambridge University Press.

[7] Geweke, J., 1993, "Bayesian treatment of the independent Student-t linear model," *Journal of Applied Econometrics,* 8, Supplement (December), S19-S40.

[8] Geweke, J., 2005, *Contemporary Bayesian Econometrics and Statistics,* Wiley.

[9] Godfrey, L.G., 2006, Tests for regression models with heteroskedasticity of unknown form. *Computational Statistics & Data Analysis,* 50, 2715–2733.

[10] Koop, G., 2003, *Bayesian Econometrics,* Wiley.

[11] White, H., 1980, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica,* 48, 817–838.

[12] Long, J.S. and L.H. Ervin, 2000, Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician,* 54, 3, 217–224.

[13] Robinson, P.M., 1987, Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica,* 55, 4, 875–891.

Appendix: Figures

Figure 1: Posterior Means of the 50 ω_i's: true value

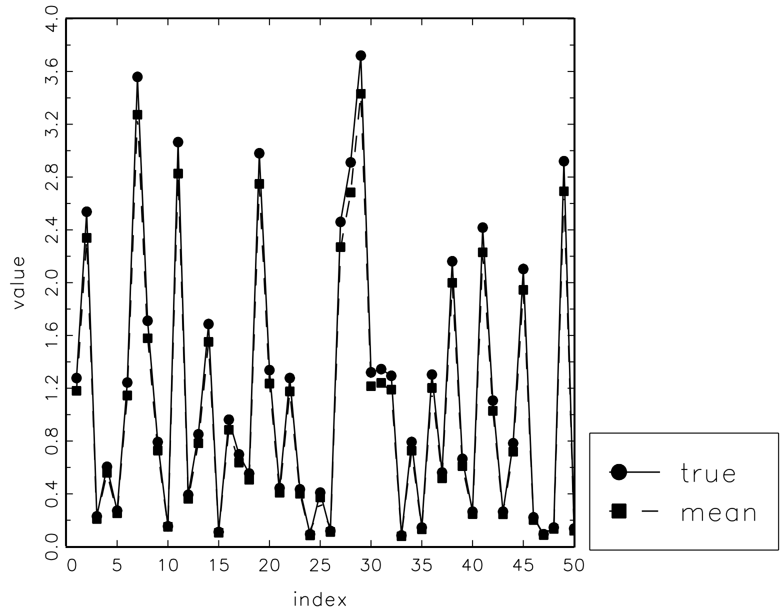

Figure 2: Posterior Std's of the 50 ω_i's: true value

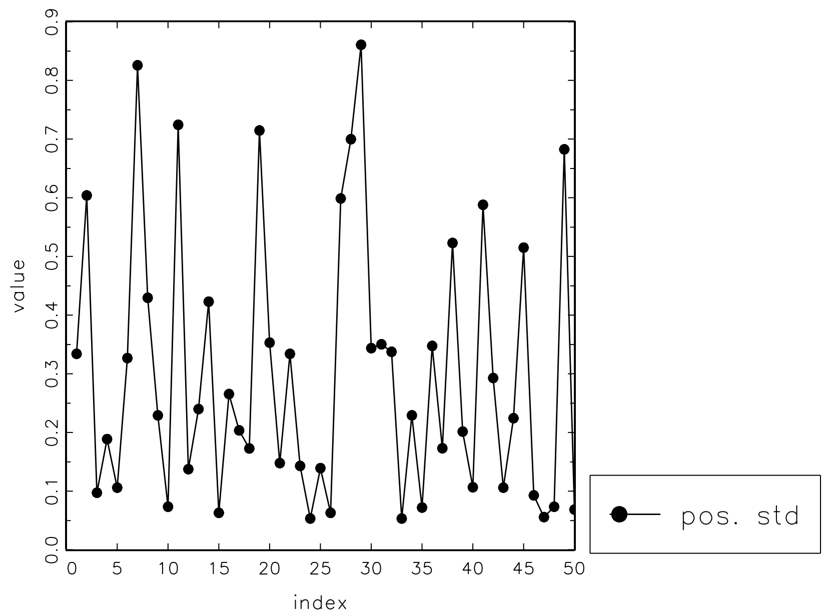

Figure 3: Posterior Density of ω_1 and ω_2: true value

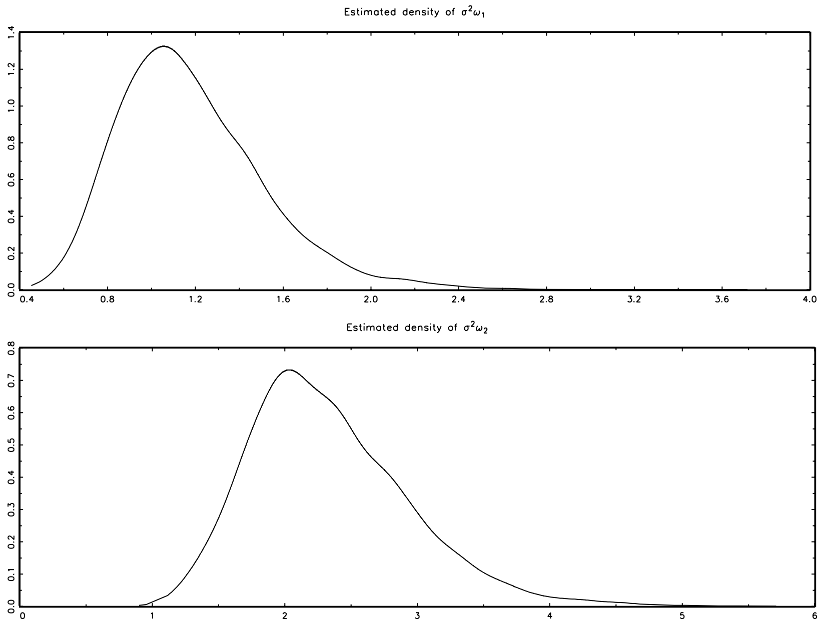

Figure 4: Posterior Means of the 50 ω_i's: HCCM

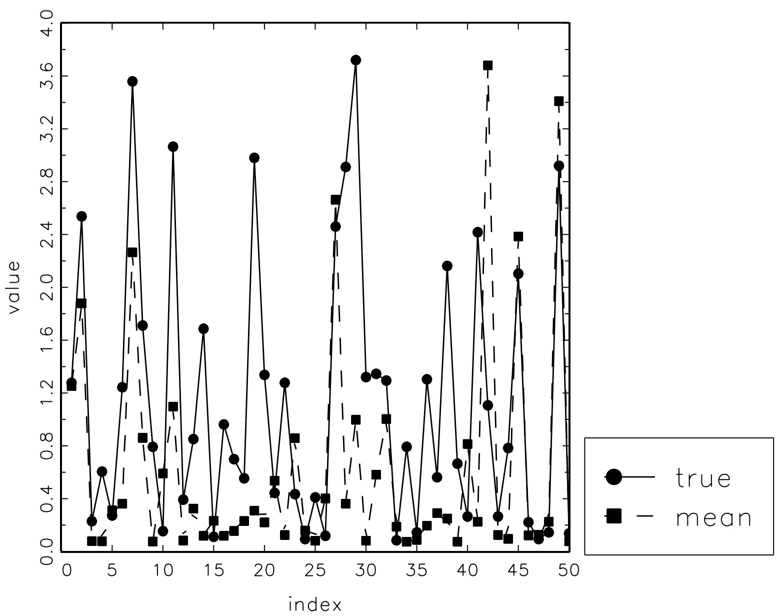

Figure 5: Posterior Std's of the 50 ω_i's: HCCM

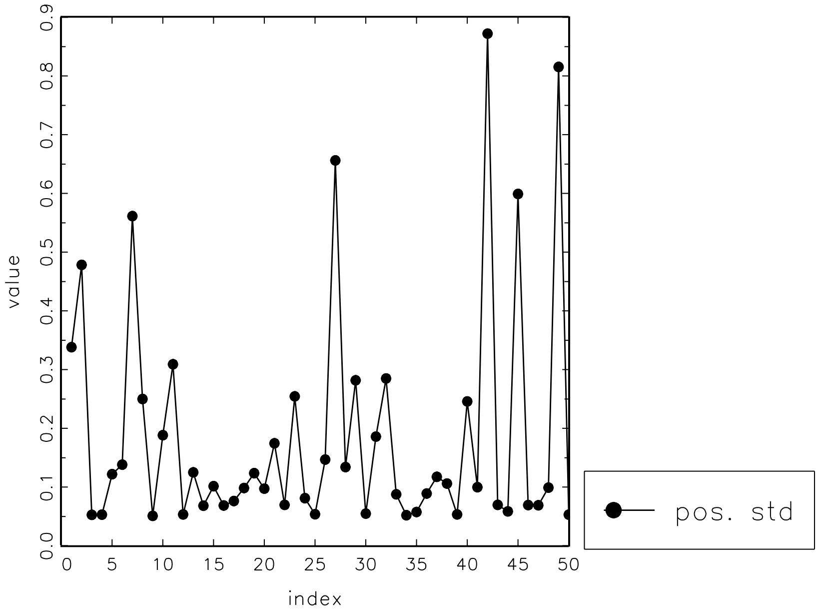

Figure 6: Posterior Density of ω_1 and ω_2: HCCM