DSSR

Discussion Paper No. 51

DIRICHLET PRIOR FOR ESTIMATING UNKNOWN REGRESSION ERROR HETEROSKEDASTICITY

Hiroaki Chigira
Tsunemasa Shiba

December 24, 2015

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

DIRICHLET PRIOR FOR ESTIMATING UNKNOWN REGRESSION ERROR HETEROSKEDASTICITY

Hiroaki Chigira* **and Tsunemasa Shiba†**

12/2015

Abstract

We propose a Bayesian procedure to estimate heteroskedastic variances of the regression error term $\boldsymbol{\omega}$, when the form of heteroskedasticity 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 Heteroskedasticity Consistent Variance–Covariance Matrix Estimator. We present an emprical 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

†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.

Dirichlet Prior for Estimating Unknown Regression Error Heteroskedasticity

Hiroaki Chigira and Tsunemasa Shiba

1 Introduction

We propose a Bayesian approach to estimate heteroskedastic parameters of regression error variances that are of unknown form, using Dirichlet prior pdf. As Amemiya (1985, p.199) points out, 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 $\boldsymbol{\omega}$. Eicker (1963) and White (1980) independently developed a well-known consistent variance-covariance matrix estimator ("HCCM" hereafter) for the OLS regression coefficient estimator. We use HCCM information to formulate proposal density of a Metropolis-Hastings ("M-H" hereafter) algorithm in Markov Chain Monte Carlo simulation. Unidentifiability of $\boldsymbol{\omega}$ poses no problem to us. As in Amemiya (1985, p.199) we use an orthogonal regression that circumvents possible underidentifiability of $\boldsymbol{\omega}$, this problem. Also, a Dirichlet prior on $\boldsymbol{\omega}$ should make it identifiable in a Bayesian context.

The trend in the HCCM literature seems to be in the ways to improve 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. 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 skedastic function of the error term is unconstrained. Robinson (1987) 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 $\boldsymbol{\omega}$ will help to sharpen posterior density of $\boldsymbol{\beta}$ 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.

In Bayesian econometrics, starting with a seminal work by Geweke (1993) there is a homoskedastic Student-t regression model derived from normal linear regression ("NLR" hereafter) with a particular set of Gamma priors for heteroskedasticity parameters. This model has been introduced in such books as Koop (2003), Geweke(2005) and Greenberg (2013), among others, and is now a popular Bayesian model. We will compare our model to this model using the Deviance Information Criterion ("DIC" hereafter).

We shall discuss estimation of $\boldsymbol{\omega}$, where ω_i is the ith volatility. In order to access an option pricing, we first need to come up with a reasonable estimate of volatility. Our estimation of $\boldsymbol{\omega}$ needs

[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 (4).

no parametric model for volatility process such as the GARCH model, since we only use HCCM information. 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.

After assuming a usual prior density for the parameters in the regression model, we may obtain a joint posterior density. The usual parameters, *e.g.,* $(\boldsymbol{\beta}, \sigma^2)$, may be easily simulated using the Gibbs sampling. It is in the simulation of the elements of $\boldsymbol{\omega}$ that we use the HCCM. We use results from HCCM to form the proposal 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. In sections 4 numerical illustrations to compare our methodology to homoskedastic t distiruted error term model, are given. We us DIC to this effect.

2 Model and Assumptions

2.1 Likelihood

Let an NLR model with heteroskedastic error term be

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

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 be

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

Our single likelihood function for y_i has the following normal distribution

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

We may use the following two notations: $\omega_i \;=\; \dfrac{1}{\lambda_i}$, where λ_i is the precision. Geweke (1993) uses $\omega_{_i}$, whereas Geweke (2005) and Greenberg (2013) use $\lambda_{_i}$. (1) in an matrix form becomes

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

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 $\sum_{i=1}^{n} \omega_i = tr(\boldsymbol{\Omega}) = n$. Note that $tr(\boldsymbol{\Omega}) = n$ restriction is often employed in heteroskedasticity literature, *e.g.,* Greene (2012, p 308). The likelihood function for the whole sample 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

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

where $\nu = n - K$, $\nu s^2 \sim$ 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 and Posterior

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2, \boldsymbol{\omega}')' \sim (K + 1 + n) \times 1$ be our parameter vector. We assume prior distribution on $\boldsymbol{\theta}$ as

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

where the three arguments of $\boldsymbol{\theta}$ are independent. Individual prior for $\boldsymbol{\beta}$ and σ^2 are

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

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.

We note that $\boldsymbol{\omega}$ is an n dimensional continuous multivariate random variable vector that satisfies 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 heteroskedasticity structure and satisfy $tr(\boldsymbol{\Omega}) = n$ restriction. 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 heteroskedasticity.

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

$$\tilde{\boldsymbol{\omega}} \sim D(\boldsymbol{\eta}_0), \tag{9}$$

where $\tilde{\boldsymbol{\omega}} = \dfrac{1}{n}\boldsymbol{\omega}$, and $D(\boldsymbol{\eta}_0)$ denotes a Dirichlet distribution with a parameter vector $\boldsymbol{\eta}_0 = (\eta_{01}, \ldots, \eta_{0n})' \sim n \times 1$. The values of $\boldsymbol{\eta}_0$ will be given later in this paper. The assumption on $\tilde{\boldsymbol{\omega}}$ is $tr(\tilde{\boldsymbol{\Omega}}) = 1$, thus $tr(\boldsymbol{\Omega}) = n$. If we make a transformation from $\tilde{\boldsymbol{\omega}}$ vector to $\boldsymbol{\omega}$ vector, we arrive at our prior distribution on $\boldsymbol{\omega}$ that resembles to Dirichlet aside from normalizing constant:

$$\pi(\boldsymbol{\omega}) \;=\; \frac{1}{n}\,\frac{\Gamma\left(\sum_i \eta_{0i}\right)}{\prod_{j=1}^{n}\Gamma(\eta_{0j})}\prod_{h=1}^{n}\left(\frac{\omega_h}{n}\right)^{\eta_{0h}-1} \;=\; n^{\,n-1-\tilde{\eta}_0}\;\frac{\Gamma\left(\sum_i \eta_{0i}\right)}{\prod_{j=1}^{n}\Gamma(\eta_{0j})}\prod_{\ell=1}^{n}{\omega_\ell}^{\eta_{0\ell}-1}, \tag{10}$$

where $\tilde{\eta}_0 = \sum_{h=1}^{n}\eta_{0h}$. Hence, its kernel is given by

$$\pi(\boldsymbol{\omega}) \;\propto\; \prod_{i=1}^{n}\omega_i^{\eta_{0i}-1}. \tag{11}$$

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

$$\pi(\boldsymbol{\theta}) \propto\; \exp\left(\frac{-1}{2}q_\beta\right)\left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_0}{2}+1}\exp\left(\frac{-\delta_0}{2\sigma^2}\right)\prod_{h=1}^{n}\tilde{\omega}_h^{\eta_{0h}-1}, \tag{12}$$

where $q_\beta = (\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{B}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$. Finally, the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X})$, 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

$$\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X}) \propto \sigma^{-n}(\prod_{i=1}^{n}\omega_i^{\frac{-1}{2}})\exp\left(\frac{-\psi}{2\sigma^2}\right)\exp\left(\frac{-1}{2}q_\beta\right)\left(\sigma^2\right)^{-(\frac{\alpha_0}{2}+1)}\exp\left(\frac{-\delta_0}{2\sigma^2}\right)\prod_{h=1}^{n}\tilde{\omega}_h^{\eta_{0h}-1}, \tag{13}$$

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

2.3 Student-t homoskedasticity model and our model

The main purpose of our model is in Bayesian estimation of the n elements in $\boldsymbol{\omega}$ without assuming a prior structure for it. On the other hand, Geweke's (1993) Student-t homoskedasticity model, also given in Geweke (2005) and Greenberg (2012), among others, is primarily concerned with estimating $\boldsymbol{\beta}$ coefficient vector in heteroskedastic NLR model.

Suppose a heteroskedastic regression disturbance $u_i|\,\lambda_i, \sigma^2 \sim N(0,\; \sigma^2/\lambda_i)$. If we assume a Gamma with an identical shape and scale parameters, $\nu = \nu_1 = \nu_2$ on a prior for λ_i , then the resultant $u_i|\,\lambda_i, \sigma^2$ distribution becomes a fat-tailed homoskedastic Student-t with the parameters $(\nu, 0, \sigma^2)$.

If there is a compelling need for such Gamma prior on λ_i, then the above equivalence of (1) NLR with heteroskedasticity, and (2) homoskedastic Student-t should be of a great value. What would the effect be if there is only one parameter value $\nu_1 = \nu_2 = \nu$ in the Gamma distribution? Such

Gamma random variable would have $E(x) = 1$ always, and its pdf becomes concentrated abround it as ν gets large. We may thus conclude that the single 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 heteroskedasticity. Notice, however, that even if a Gamma prior $G(\frac{\nu_1}{2},\ \frac{\nu_2}{2})$, on λ_i is assumed, then n λ_i's are, again, generated from one single Gamma. This is in effect assuming a particular structure on the heteroskedasticity 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 a scalar 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 heteroskedastic variance parameters that have unknown structure, on two grounds. First, it is free of restrictions from the small number of parameters that govern 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 are continuous and 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 "$\boldsymbol{\theta}_{-\vartheta}$" to denote $\boldsymbol{\theta}$ less $\boldsymbol{\vartheta}$ hereafter. For instance, "$\boldsymbol{\theta}_{-\boldsymbol{\beta}}$" implies $\boldsymbol{\theta}_{-\boldsymbol{\beta}} = (\sigma^2, \boldsymbol{\omega}')' \sim (1+n) \times 1$.

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. On the other hand we need to implement an Independence Chain M-H 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 $\boldsymbol{\beta}$ from a multivariate normal.

Remark 2. Fully conditional posterior of σ^2 is in Inverted Gamma:

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

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

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

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

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

$$\pi(\boldsymbol{\omega} | \, \boldsymbol{\theta}_{-\omega}, \, \boldsymbol{y}, \boldsymbol{X}) \, \propto \, (\prod_{i=1}^{n} \omega_i^{(\eta_{0i}^* - 1)}) \exp\left(\frac{-\psi}{2\sigma^2} \right) , \tag{17}$$

where $\eta_{0i}^* = \eta_{0i} - \dfrac{1}{2} > 0$ for $i = 1, .., n$ in order for $\boldsymbol{\eta_0^*}$ vector to make sense as a Dirichlet parameter. Let

$$A_\omega = \prod_{i=1}^{n} \omega_i^{(\eta_{0i}^* - 1)} \quad \text{and} \quad B_\omega = \exp\left(\frac{-\psi}{2\sigma^2} \right)$$

hence $\pi(\boldsymbol{\omega} | \, \boldsymbol{\theta}_{-\omega}, \, \boldsymbol{y}, \boldsymbol{X}) = \; A_\omega \, B_\omega$. Obviously A_ω is a kernel of $D(\boldsymbol{\eta_0^*})$. 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 M-H Acceptance Probability of $\boldsymbol{\omega}$

Let $\boldsymbol{\omega}^{(r)}$ be the "rth" current value in the chain. Then the acceptance probability of $\boldsymbol{\omega}^{(\prime)}$ 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

$$\frac{\pi(\boldsymbol{\omega}^{(\prime)}, \boldsymbol{\theta}_{-\omega}|\, \boldsymbol{y}, \boldsymbol{X})}{\pi(\boldsymbol{\omega}^{(r)}, \boldsymbol{\theta}_{-\omega}|\, \boldsymbol{y}, \boldsymbol{X})} = \frac{\prod_i \left(\omega_i^{(\prime)}\right)^{\eta_{0i}^*-1}}{\prod_j \left(\omega_j^{(r)}\right)^{\eta_{0j}^*-1}} \frac{\exp\left(\frac{-\psi^{(\prime)}}{2\sigma^2}\right)}{\exp\left(\frac{-\psi^{(r)}}{2\sigma^2}\right)}, \tag{18}$$

where $\psi^{(\prime)} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \left(\boldsymbol{\Omega}^{(\prime)}\right)^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, $\boldsymbol{\Omega}^{(\prime)} = diag(\boldsymbol{\omega}^{(\prime)})$ and this is not to be confused with a transpose of $\boldsymbol{\Omega}$. Likewise, $\psi^{(r)} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \left(\boldsymbol{\Omega}^{(r)}\right)^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, $\boldsymbol{\Omega}^{(r)} = diag(\boldsymbol{\omega}^{(r)})$. Suppose that we employ $D(\boldsymbol{\eta}_p)$ as our proposal[2] , then the ratio of proposal densities above becomes

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

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

$$\alpha(\boldsymbol{\omega}^{(r)}, \boldsymbol{\omega}^{(\prime)}) = \min\left(1, \frac{\exp\left(\frac{-\psi^{(\prime)}}{2\sigma^2}\right)}{\exp\left(\frac{-\psi^{(r)}}{2\sigma^2}\right)} \frac{\prod_i \left(\omega_i^{(r)}\right)^{\eta_i-\eta_{0i}^*}}{\prod_j \left(\omega_j^{(\prime)}\right)^{\eta_j-\eta_{0j}^*}}\right). \tag{20}$$

3.2.2 Proposal Density Parameter

In this subsection, we outline how we specify the Dirichlet proposal density that uses White's HCCM information. Notice that our prior for $\boldsymbol{\omega} = n\,\tilde{\boldsymbol{\omega}}$ resembles to a Dirichlet, where $\tilde{\boldsymbol{\omega}} \sim D(\boldsymbol{\eta_0})$. We let the parameter vector of the Dirichlet proposal density be $\boldsymbol{\eta}_p = c\,\hat{\boldsymbol{\eta}}$, where c is a scalar tunning parameter, and information on $\hat{\boldsymbol{\eta}}$ is to be extracted from a regression to be explained in below. We briefly discuss how we obtain $\hat{\boldsymbol{\eta}}$.

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 the vector obtained from $\hat{\boldsymbol{H}}$ be $\hat{\boldsymbol{h}}$, where $\hat{\boldsymbol{h}} = vech(\hat{\boldsymbol{H}}) \sim K' \times$

[2]The details of our proposal density are given in the next subsection.

1 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 some error term vector, will yield an estimator of $\boldsymbol{\omega}$, when $K' > n$ and $n > K$. When these conditions are not met, we may always augment $\boldsymbol{X}$ by $\boldsymbol{W} \sim n \times K_W$ such that $\boldsymbol{W}'(\boldsymbol{y},\, \boldsymbol{X}) = \boldsymbol{0}$. This is essentially finding an orthogonal comlement of the matrix $(\boldsymbol{y},\, \boldsymbol{X})$. Using such software as *R* and *GAUSS*, among others, we can easily find $\boldsymbol{W}$. Actually we only need $\boldsymbol{W} \sim n \times K_{_W}$ such that $K_{_W}$ meets the next two conditions: $\frac{1}{2}(K + K_{_W})(K + K_{_W} + 1) > n$ and $n > (K + K_{_W})$. Since $\boldsymbol{W}$ is orthogonal to both $\boldsymbol{y}$ and $\boldsymbol{X}$, OLS estimated regression residual from (21), $\boldsymbol{e}_{ols}$, is the same whether we used $\boldsymbol{X}$ or $\boldsymbol{W}$ augmented $(\boldsymbol{X},\, \boldsymbol{W})$ as the regressor matrix. In summary, we may always estimate $\boldsymbol{\omega}$ given data, $\boldsymbol{y}$ and $\boldsymbol{X}$, using (21). Let the resultant estimator of $\boldsymbol{\omega}$ from (21) be $\hat{\boldsymbol{\omega}}_{ols}$.

It turns out, however, that we cannot simply set $\hat{\boldsymbol{\omega}}_{ols}$ equal to $\hat{\boldsymbol{\eta}}$. Our preliminary investigations revealed that there are cases where the variability of $\hat{\boldsymbol{\omega}}_{ols}$ is too big for it to be $\hat{\boldsymbol{\eta}}$. A variability measure of $\hat{\boldsymbol{\omega}}_{_{ols}}$ could be $\mathcal{R} = \dfrac{\max(\hat{\omega}_{_{ols\,i}})}{\min(\hat{\omega}_{_{ols\,i}})}$, where $\hat{\omega}_{_{ols\,i}}$ is the ith element of $\hat{\boldsymbol{\omega}}$. We aim to obtain a moderate $\mathcal{R}$ value, while preserving the ranking, the order or the relative magnitude of the elements of $\hat{\boldsymbol{\omega}}_{_{ols}}$ that HCCM information conveys. To this effect, we introduce an additional scalar tunning parameter $d \geq 0$ to be added to $\hat{\omega}_{_{ols\,i}}$ as $\hat{\eta}_i = (\hat{\omega}_{_{ols\,i}} + d)$, where $\hat{\eta}_i$ is the ith element of $\hat{\eta}$. Define the new ratio to be $f(d) = \dfrac{\max(\hat{\omega}_{_{ols\,i}}) + d}{\min(\hat{\omega}_{_{ols\,i}}) + d}$. This ratio, $f(d)$ is a function of $d \geq 0$, and has the property such that

$$\mathcal{R} \geq f(d) \geq 1 \text{ since } f'(d) < 0,\; \lim_{d\to\infty} f(d) = 1,\; \text{and } \lim_{d\to 0} f(d) = \mathcal{R}.$$

Thus by an appropriate choice of d, we may make the variability of $\hat{\boldsymbol{\omega}}_{_{ols}}$ to be any desirable level. A suggested value for d is to set $f(d) = n$, *i.e.,* sample size n dependent. Solving the equation for d, we obtain

$$d = \frac{\max(\hat{\omega}_{_{ols\,i}}) - n\, \min(\hat{\omega}_{_{ols\,i}})}{n-1} \approx \frac{1}{n} \max(\hat{\omega}_{_{ols\,i}}) - \min(\hat{\omega}_{_{ols\,i}}),$$

which is a reasonable value.

In summary our Dirichlet proposal density parameter $\boldsymbol{\eta}_p = c\,\hat{\boldsymbol{\eta}}$ is set to be $\hat{\eta}_i = \hat{\omega}_{_{ols\,i}} + d$ with c and d tunning parameters. One proposed value for d is given in the preceding paragraph. If it happens that $\mathcal{R} \leq n$, then set $d = 0$ and $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\omega}}_{_{ols}}$ would do the job.

4 Empirical Investigation: Japanese Stock Returns

In this section, using Japanese stock dataset, we present a comparison of our Dirichlet prior method, Student-t homoskedasticity model (Geweke(1993)'s), and the usual homoskedastic NLR model, in terms of DIC.

4.1 One Factor Model and Dataset

We carry out a two-step time series to cross section regressions, in a way similar to the Fama-Macbeth procedure (see *e.g.,* Cochrane (2001 p.244)). We used data on the daily stock prices of fifty Japanese pharmaceutical/ biomedical companies. To obtain an excess return series, we used 10-year Japanese Government Bond (JGB) rate for the risk free rate. For the market return, we used TOPIX. The sample period is from May 6, 2005 to April 28, 2006, hence the sample size is 245 in total. We obtained stock return data and the JGB data from Yahoo Finance and Nikko Financial Intelligence web site, respectively.

We begin with an one factor return generating equation:

$$\boldsymbol{R} = (\boldsymbol{\iota}_T \ \boldsymbol{f}) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\varepsilon}, \tag{22}$$

where $\boldsymbol{R}$= $(\boldsymbol{R}_1 \cdots \boldsymbol{R}_N) \sim T \times N$ is a T period excess returns for N firms, $\boldsymbol{\iota}_T \sim$ vector of one's, $\boldsymbol{f}$ is a $T \times 1$ vector of one factor, $\boldsymbol{\alpha}$=$(\alpha_1, \ldots, \alpha_N) \sim 1 \times N$ vector of constants, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_N) \sim 1 \times N$ is a vector of beta's, $\boldsymbol{\varepsilon}$=$(\boldsymbol{\varepsilon}_1 \cdots \boldsymbol{\varepsilon}_N) \sim T \times N$ matrix of error terms, N is the number of stocks, and T is the time series sample size. Equation (22) is just a set of N time series regressions. We obtain an OLSE of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}} \sim 1 \times N$ from equation (22). Define sample mean of $\boldsymbol{R}$ to be an N dimensional vector $\bar{\boldsymbol{R}}$, we then obtain a cross sectional regression model

$$\bar{\boldsymbol{R}} = \hat{\boldsymbol{\beta}}' \phi + \boldsymbol{u}, \tag{23}$$

where $\bar{\boldsymbol{R}} = \frac{1}{T} \boldsymbol{R}' \boldsymbol{\iota}_T = (\bar{R}_1 \cdots \bar{R}_N)' \sim N \times 1$ vector of average excess returns, $\phi \sim 1 \times 1$ scalar is a risk premium associated with the factor $\boldsymbol{f}$, $\boldsymbol{u} \sim N \times 1$ is a vector of pricing errors. Equation (23) is the one factor type CAPM *without* an intercept term given in Cochrane *op.cit.*, p.235, among others[3].

[3]This specification is found *e.g.,* in Cochrane *op.cit.* equation 12.10. We have regressed with an intercept term, and the OLSE of it is 0.001 (0.054) and the slope estimate is 0.109 (0.069), where the figures inside the parentheses are estimated standard errors. Without an intercept term, the slope estimate is 0.110 (0.025), and there is very little difference whether we include an intercept term or not.

In this section so far, we have used a set of notations that are common in empirical finance, and in this paper equation (23) corresponds to equation (1). We need to clarify the notational correspondences between the ones used so far in the current section, and in the previous subsections. The correspondences are: $\bar{\boldsymbol{R}} \sim N \times 1$ corresponds to $\boldsymbol{y}$ in (1), $\hat{\boldsymbol{\beta}}' \sim N \times 1$ corresponds to $\boldsymbol{X}$ in (1), $\boldsymbol{u} \sim N \times 1$ corresponds to $\boldsymbol{\varepsilon}$ in (1), $\lambda \sim$ scalar corresponds to $\boldsymbol{\beta}$ in (1), the number of factor in (23) is one and it corresponds to K in (1), N the number of stocks corresponds to n in (1).

4.2 The Three Models Compared

The three models we compare are

(i) homoskedastic NLR model,

(ii) Geweke's (1993) model, and

(iii) our model in this paper.

We designed the common parameters of the three models to be the same. For instance, all are based on NLR of equation (23), with mutually independent priors:

$$\phi \sim N(\beta_0, B_0), \quad \sigma^2 \sim \mathcal{IG}\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \tag{24}$$

where we set $\alpha_0 = \delta_0 = 10^{-2}$, $\beta_0 = 0$, and $B_0 = 10^4$. The posterior pdf is given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) = \pi(\phi)\pi(\sigma^2)\pi(\boldsymbol{\omega})\pi(\boldsymbol{y}|\boldsymbol{\omega}, \boldsymbol{X}), \tag{25}$$

where $\pi(\phi)$ and $\pi(\sigma^2)$ are given in (24), while $\pi(\boldsymbol{\omega})$ depends on the model, and $\pi(\boldsymbol{y}|\boldsymbol{\omega}, \boldsymbol{X})$ is a multivariate normal based likelihood given in (6)[4] . In all three MCMC similations, burn-in is set to be 10,000 while one apart simulated values are taken out of 8,000, *i.e.* total of 4,000 simulated values are obtained.

Let us first discuss model **(i)**, homoskedastic NLR Model. Homoskedasticity in terms of $\pi(\boldsymbol{\omega})$ is represented by $\boldsymbol{\omega} = \boldsymbol{\iota}_n$, *i.e.,* all elements to be one. The model is now a homoskedastic NLR, however, MCMC is needed since our prior for σ^2 is an informative inverted gamma.

We next take up model **(ii)**, Geweke (1993)'s model. This is an NLR model with heteroskedasticity. As we stated in section 2.3, Geweke (1993) converted this model to Student-t homoskedastic model. Unlike our inverted gamma σ^2 prior in (24), he used noninformative σ^2 prior, however. The same

[4] $\boldsymbol{\beta}$ in (6) should be replaced by ϕ.

inverted gamma σ^2 prior as in (24), is used in Koop (section 6.4, 2003). Hence, we decided to follow Koop (*op. cit.*)'s MCMC sampler for model **(ii).** Joint posterior of Geweke (1993)'s model is still (25) except for $\pi(\boldsymbol{\omega})$. This is specified with a single parameter ν_0 inverted gamma density. We use two different ν_0 values: 25 and 5. $\nu_0 = 5$ should be interpreted as an indication of larger heteroskedasticity compared to $\nu_0 = 25$.[5] So far as the regression coefficients $\boldsymbol{\beta}$ and the error term variance scale parameter σ^2 are concerned, MCMC procedures are the same as homoskedastic NLR model stated in the above paragraph.

4.2.1 Comparison of the Models: DIC and Posterior Evaluation

Since Spiegelhalter *et al.* (2002) first proposed DIC it has become one of the most frequently used model comparison criterion for Bayesians.[6] Let a single parameter of interest be θ, then the deviance is defined to be $D(\theta) = -2\log(\text{likelihood})$. The posterior mean deviance, $\overline{D(\theta)}$, is defined to be $D(\theta)$ evaluated under posterior measure, thus $\overline{D(\theta)} = E_\theta[D(\theta)|\boldsymbol{y}]$, where $\boldsymbol{y}$ denotes data. The penalty term for DIC, p_D, is defined to be $p_D = \overline{D(\theta)} - D(\bar{\theta})$, where $\bar{\theta}$ denotes posterior mean of θ. Then DIC is given by

$$DIC = \overline{D(\theta)} + p_D \,. \tag{26}$$

Note that p_D is often interpreted to be the effective number of parameters, hence a measure of model complexity.

The following shows $\overline{D(\theta)}$, p_D and DIC of the three models.

Table 1. DIC Compared

	homoskedastic NLR	Geweke ($\nu_0 = 25$)	Geweke ($\nu_0 = 5$)	Our model
$\overline{D(\theta)}$	-52.975	-79.049	-81.598	-114.956
p_D	1.924	3.070	2.438	2.907
DIC	-51.051	-75.980	-79.160	-112.050

Since the null model in (24) essentially contains two parameters, σ^2 and ϕ, p_D should approximately two. We observe that homoskedastic NLR has $p_D \approx 2$ while Geweke (1993)'s model and ours have $p_D \approx 3$. As regards to the goodness of fit measure, *i.e.,* $\overline{D(\theta)}$, our model has the best measure while

[5] See Koop (2003, p.129) for setting ν_0.
[6] See Spiegelhalter *et al.* (2014) for an excellent survey, pros and cons of DIC.

the other two have lesser $\overline{D(\theta)}$. In summary, our model has the best DIC with reasonable $p_{_D}$ values in the models compared.

We next examine three model's parameter $\boldsymbol{\theta}$ estimation result. Numerical results, except for $\boldsymbol{\omega}$ are given in Table 2. Homoskedastic NLR has the smallest ϕ and the largest σ^2 of all the models considered. This may indicate a tradeoff between ϕ and σ^2 due to misspecification of homoskedasticity, since other models' ϕ and σ^2 are more or less the same. Let estimated parameter value over its standard deviation be "t-value." We observe that our model's two parameters presented in Table 2, have the largest t-value. This fact may imply GLS-like treatment of our model is most appropriate for the dataset.

Table 2. Posterior Mean Compared

	homoskedastic NLR	Geweke ($\nu_{_0} = 25$)	Geweke ($\nu_{_0} = 5$)	Our model
ϕ	0.111 (0.026)	0.138 (0.025)	0.141 (0.024)	0.118 (0.013)
σ^2	0.021 (0.004)	0.010 (0.004)	0.008 (0.003)	0.014 (0.003)
ν	$--$	4.967 (5.174)	3.349 (1.735)	$--$

"$(\cdot)$" indicates posterior standard deviation.

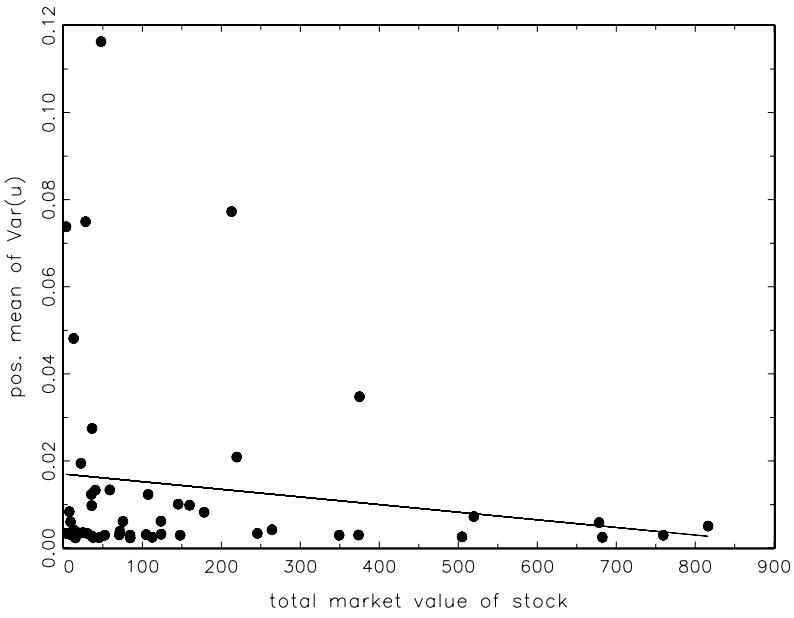
4.3 $\boldsymbol{\omega}$: Posterior Mean and Posterior Density

We may now present 50 posterior mean's of our remaining parameter, $\boldsymbol{\omega}$, *i.e.,* $E_{\boldsymbol{\theta}}(\boldsymbol{\omega}\,|\text{data}) \sim 50 \times 1$. Since we do not know the true heteroskedasticity of our data, however, presenting 50 posterior means would not contribute to our understanding of unknown skedastic structure.

"Size effect" hypothesis that indicates an inverse relationship between volatility to size of a company, may be an appropriate hypothesis to be dealt with. In the U.S. and world wide, starting with a seminal paper by Banz (1981) many observed "size effect" that is the smaller the company is the higher its return[7]. This phenomenon could be naively associated to the mean-variance efficiency to yield a thesis that says smaller companies are expected to be more riskier, *i.e.,* the smaller the size, the larger the mean and volatility of returns. Berk (1997) among others examined the so-called *size effect* and proposed that a size of a company should not be measured by the market value of its equity

[7]For the size effect in Japan, see *e.g.,* Chan and Chen (1991), among others.

Figure 1: Volatility versus Size: $\sigma^2\omega_i$'s and MVE



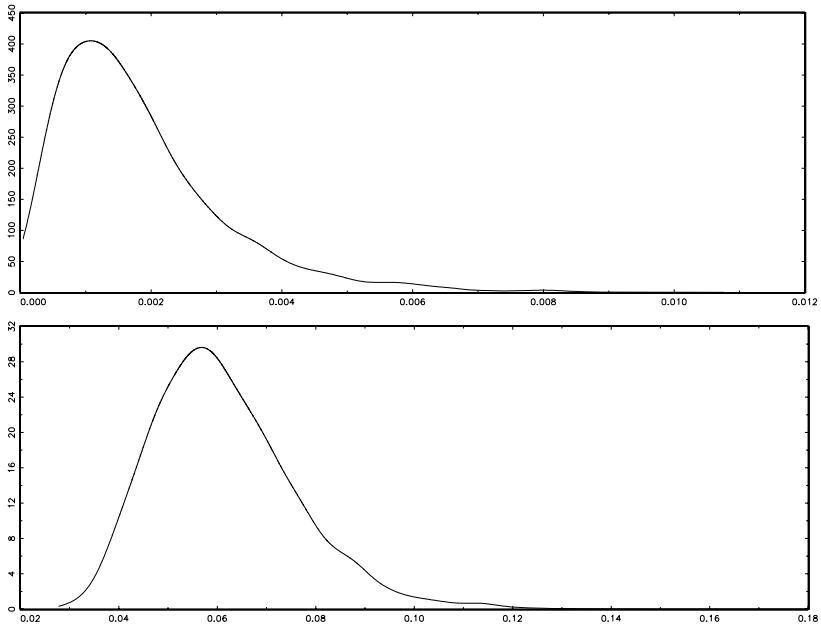
(MVE) but some other variables such as sales. In this section, we intend to investigate whether larger (smaller) size companies have smaller (larger) ω_i's.

For the fifty stock data, we now collected the market value of its equity (MVE) data (in one billion yen)[8] . We then drew a graph with the MVE on the horizontal axis, and posterior mean of $\sigma^2\omega_i$'s from our model on the vertical axis. This is shown in Figure 1. The downward sloping solid line in Figure 1, is the OLS estimated line. This figure clearly shows the larger the MVE, the smaller the volatility as measured by $\sigma^2\omega_i$. In summary, we have confirmed that $E_{\boldsymbol{\theta}}(\boldsymbol{\omega}\,|\text{data})$ from our model, gives reasonable values.

We need to see each simulated posterior density of ω_i has a shape that is reasonable as a density of variance, *e.g.,* gamma density. To this end we selected two stocks (i) that has large MVE and small $\sigma^2\omega_i$, and (ii) that has small MVE and large $\sigma^2\omega_i$, to see what the marginal posterior pdf's of $\sigma^2\omega_i$ of these companies look like. Specifically, we chose Taisho Pharmaceutical Co., Ltd. for (i), and Site Support Institute Co., Ltd. for (ii). They are given in Figures 2 below. Notice that the two charts have different horizontal axis scale. The smaller MVE stock has very large volatility (see the lower chart) compared to the that of the larger MVE stock (see the upper chart). The two pdf's have quite

[8]Berk (1997) among others, examined the so-called *size effect* and proposed that a size of a company should not be measured by MVE but some other variables such as sales.

Figure 2: Marginal Posterior pdf of $\sigma^2\omega_i$ for Taisho Phamaceutical Co. and Site Support Institute



reasonable shapes. We conclude that our Bayesian estimation of volatility supports the view that the smaller the size of the stock, the larger is the volatility.

5 Concluding Remarks

In this paper, we proposed a Bayesian method to estimate regression error heteroskedasticity structure that is unknown. "Unknown" in the sense that no structure is assumed. Geweke (1993)'s model is such that when Gamma priors with a particular set of hyper parameter values are assumed on the precision parameters of heteroskedastic regression error term, then this leads to a homoskedastic Student-t regression error term. We pointed out that assuming such priors, are in effect, imposing an unwanted structure in the heteroskedasticity. 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 heteroskedasticity.

We, on the other hand, believe that the Eicker-White HCCM should provide valuable information about the heteroskedasticity, although derived from a sampling theory point of view. In empirical analysis, regression equation is bound to be misspecified. HCCM, in essence, draws heteroskedasticity 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.

Finally we compared homoskedastic NLR, Geweke (1993)'s model, and our model in terms of DIC, posterior mean significance (we used "t-value" to this effect), and posterior pdf's. All the exercises indicate that our model is definitely capable of drawing unknown heteroskedasticity structure.

References

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

[2] Banz, R.W., 1981, The relationship between return and market value of common stocks. *Journal of Financial Economics,* 9, 3–18.

[3] Berk, J.B., 1997, Does size really matter? *Financial Analysts Journal,* 53, 12–18.

[4] Chan, K.C. and Nai-Fu Chen, 1991, Structural and return characteristics of small and large firms. *Journal of Finance,* 46, 1467–84.

[5] Cochrane, J.H., 2001, *Asset Pricing,* Princeton University Press.

[6] 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.

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

[8] Greenberg, E., 2013, *Introduction to Bayesian Econometrics, 2nd ed.* Cambridge University Press.

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

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

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

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

[13] Lancaster, T., 2004, *An Introduction to Modern Bayesian Econometrics,* Blackwell.

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

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

[16] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A., 2002, Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society Series B.*, 64, 583–639.

[17] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A., 2014, Deviance information measure: after 12 years (with discussion), *Journal of the Royal Statistical Society Series B.*, 76, part 3, 485–493.

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