Global COE Hi-Stat Discussion Paper Series 051

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

Bayesian Estimation of Unknown Regression Error Heteroscedasticity

Hiroaki Chigira
Tsunemasa Shiba

March 2009
(Revised: April 2009)

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

Bayesian Estimation of Unknown Regression Error Heteroscedasticity

Hiroaki Chigira*and Tsunemasa Shiba†

April 17, 2009

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 elicited from the well-known Eicker–White Heteroscedasticity Consistent Variance–Covariance Matrix Estimator. Markov Chain Monte Carlo algorithm is used to simulate posterior pdf's of the unknown elements of $\boldsymbol{\omega}$. In addition to the numerical examples, we present an empirical investigation of the stock prices of Japanese pharmaceutical and biomedical companies to demonstrate usefulness of the proposed method.

key words
Eicker–White HCCM, orthogonal regressors, informative prior pdf's, MCMC, stock return variance

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.

†Department of Economics, Hitotsubashi University, Kunitachi-shi, Tokyo 186-8601 Japan. e-mail: tshiba@stat.hit-u.ac.jp Tel & fax: +81-42-580-8786

1 Introduction

We propose a Bayesian procedure to estimate heteroscedastic variances of the regression error term, when the form of heteroscedasticity is unknown. 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 candidate 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.

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 $\boldsymbol{\omega}$ vector. 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

[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.

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 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 and an empirical investigation of the stock prices of Japanese pharmaceutical and biomedical companies, are given in section 4. Section 5 concludes.

2 The Model and the Joint Posterior pdf

Suppose a heteroscedastic regression model,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{X} \sim n \times K$ is a purely exogenous variable matrix, $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\boldsymbol{\Omega})$, $\boldsymbol{\Omega} = \mathrm{diag}(\boldsymbol{\omega}) = \mathrm{diag}(\omega_1, \ldots, \omega_n)$, $\boldsymbol{\omega} \subset \Re^n_+$, $\mathrm{tr}(\boldsymbol{\Omega}) = \sum_{i=1}^n \omega_i = n$, $\sigma^2 > 0$, and n is the sample size[3]. Let the column vector $\boldsymbol{\theta}$ be all the parameters in the model, *i.e.,* $\boldsymbol{\theta}' = (\boldsymbol{\beta}' \ \sigma^2 \ \boldsymbol{\omega}') \sim 1 \times (K + 1 + n)$. The likelihood function, then, becomes

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{X}) \propto |\boldsymbol{\Omega}|^{-1/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right). \tag{2}$$

Assuming that the components of $\boldsymbol{\theta}$ are independent *a priori*, and specifically the prior of $\boldsymbol{\beta}$ is diffuse, we obtain the following prior pdf under the $\mathrm{tr}(\boldsymbol{\Omega}) = n$ assumption:

$$p(\boldsymbol{\theta}) \propto |\sigma^2 \ \boldsymbol{\Omega}|^{-1/2}\mathbf{1}_{[\underline{\omega}, \overline{\omega}]}(\boldsymbol{\omega}), \tag{3}$$

where "$\mathbf{1}_{[\underline{\omega}, \overline{\omega}]}(\boldsymbol{\omega})$" designates an indicator function such that it takes on the value one if an element of the $\boldsymbol{\omega}$ vector is contained in the range $[\underline{\omega}, \overline{\omega}]$, and zero otherwise.

We shall discuss how we obtained the upper bound $\overline{\omega}$ and the lower bound $\underline{\omega}$. Our scaling assumption of $\boldsymbol{\omega}$ is given by $\sum_{i=1}^n \omega_i = n$. This means

[3] "$\mathrm{tr}(\boldsymbol{\Omega}) = n$" is a technical assumption that is often emplyed for identifiability purpose in regression models with heteroscedastic error term. See, *e.g.,* Greene (2008, p.158).

that the sample mean of $\boldsymbol{\omega}$ is one, *i.e.,* $\bar{\omega} = 1$. Consider a rather extreme case that ten percent of ω_i's have ten times the magnitude of $\bar{\omega} = 1$, *i.e.,* 10% of elements of $\boldsymbol{\omega}$ is $\omega_i = 10$. This leaves the rest of ninety percent of ω_j's, zeros only, *i.e.,* $\omega_j = 0$ for $j \neq i$, where the i belong to the first ten percent of the entire n. Hence, along with the assumption $\omega_i > 0$ for $i = 1, \ldots, n$, the fact that $\bar{\omega} = 1$ encourages us to set a certain range for each ω_i that depends on the sample size, n. In the rest of this paper, our sample size in simulation and also empirical investigation, is fifty. We may, therefore, safely set $(\underline{\omega}, \overline{\omega}) = (0.025,\ \ 12.5)$, say. What does this imply? Consider an example of the first four observations taking up the bulk of the entire volatility. That is $\sum_{i=1}^{4} \omega_i = 46$ while $\sum_{i=5}^{50} \omega_i = 4$. If this is the case, our $\underline{\omega}$ and $\overline{\omega}$ may be given by

$$\frac{4}{46} = 0.087 > 0.025 = \underline{\omega} \quad \text{and} \quad \frac{46}{4} = 11.5 < 12.5 = \overline{\omega}.$$

The above implies that even in this extreme case, our lower bound, $\underline{\omega}$, is far smaller than the average of the latter forty six ω_i's, while our upper bound, $\overline{\omega}$, far exceeds the average of the first four ω_i's. We emphasize that this data dependent prior on the bounds of ω_i for $i = 1, \ldots, n$, are relatively easy to set, since the only information needed to form it is n. We also emphasize that our experiments show insensitivity of MCMC draws to the bounds, in the sense that the results are robust to the values of approximately ten percent changes in the values of $(\underline{\omega}, \overline{\omega})$. Our prior pdf for $\boldsymbol{\theta}$ in this paper is given by

$$p(\boldsymbol{\theta}) \propto \sigma^{-1}\ |\boldsymbol{\Omega}|^{-1/2} \mathbf{1}_{[0.025,\ \ 12.5]}(\boldsymbol{\omega}), \tag{4}$$

and this replaces equation (3). Note that this specification of $(\underline{\omega}, \overline{\omega})$ amounts to

$$(\underline{\omega}, \overline{\omega}) = (0.025/50, \quad 12.5/50) = (0.0005, \quad 0.25)$$

that is $\underline{\omega} = 0.05$ percent of n, while $\overline{\omega} = 25$ percent of n. This is the rule of thumb that we recommend. When implementing this "rule of thumb," around ten percent of each bound, should be explored.

Our joint posterior pdf is proportional to the product of equation (2) times (4):

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{X}) \propto |\boldsymbol{\Omega}|^{-1}\sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) \mathbf{1}_{[0.025,\ 12.5]}(\boldsymbol{\omega}) \tag{5}$$

3 Posterior Simulation: MCMC

We shall simulate $\boldsymbol{\beta}$ and σ using a Gibbs sampler, and then combine a Metropolis-Hastings ("MH" hereafter) algorithm to simulate the entire $\boldsymbol{\theta}$.

3.1 Simulating $\boldsymbol{\beta}$

Let $\boldsymbol{\theta}_{-\beta}$ represent a vector $\boldsymbol{\theta}$ without $\boldsymbol{\beta}$. Then the Gibbs sampler for $\boldsymbol{\beta}$ starts out with

$$\boldsymbol{\beta}^{(1)}|\boldsymbol{\theta}^{(0)}_{-\beta} \sim p(\boldsymbol{\beta}|\boldsymbol{\theta}^{(0)}_{-\beta}, \boldsymbol{y}, \boldsymbol{X}),$$

where the superscript r in $\boldsymbol{\beta}^{(r)}$ denotes the rth simulated $\boldsymbol{\beta}$ (thus $r = 0$ is the initinal value) for $r = 1, \ldots, R$, and R is the number of replications in the MCMC. The conditional probability in the right hand side represents a fully conditional posterior pdf of $\boldsymbol{\beta}$, where a notation "$A \sim p(\cdot)$" is used to indicate that a value A is generated from $p(\cdot)$ pdf. We use the OLS estimated residual vector $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ to obtain $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-K)$, and this is used as the initial value of σ^2. Next, note that it is pretty reasonable to set the initial values of $\boldsymbol{\omega}$, $\boldsymbol{\omega}^{(0)}$, as $\boldsymbol{\omega}^{(0)} = \boldsymbol{\iota}_n$, where $\boldsymbol{\iota}_n \sim n \times 1$ is a vector of all one's. This comes from the scaling assumption of $\boldsymbol{\omega}$, *i.e.,* $\bar{\omega} = 1$. As a result, our initial value vector becomes $\boldsymbol{\theta}^{(0)}_{-\beta} = (s^2\ \boldsymbol{\iota}'_n)'$. Notice that the full conditional distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\theta}_{-\boldsymbol{\beta}}$ is

$$\boldsymbol{\beta}|(\boldsymbol{\theta}_{-\boldsymbol{\beta}}, \boldsymbol{y}, \boldsymbol{X}) \sim \mathrm{N}\left(\hat{\boldsymbol{\beta}},\ \sigma^{-2}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\right).$$

Then, using $\boldsymbol{\theta}_{-\boldsymbol{\beta}}^{(0)}$, we may generate $\boldsymbol{\beta}^{(1)}$ from a multivariate normal conditional density:

$$\boldsymbol{\beta}^{(1)} \sim \mathrm{N}\left(\hat{\boldsymbol{\beta}},\ s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right),$$

where it turns out that the diagonal elements of the initial value of $\boldsymbol{\Omega}$ is simply $\mathrm{diag}(\boldsymbol{\iota}_n) = \boldsymbol{I}_n$. In $r \geq 2$ iterations, above $\boldsymbol{\beta}^{(1)}$ generation scheme is replaced by

$$\boldsymbol{\beta}^{(r)} \sim \mathrm{N}\left(\boldsymbol{\beta}^{(r-1)},\ \sigma^{(r-1)^2}(\boldsymbol{X}'\boldsymbol{\Omega}^{(r-1)^{-1}}\boldsymbol{X})^{-1}\right), \tag{6}$$

however.

3.2 Simulating σ

We, next, turn to generation of σ. The conditional posterior pdf of σ becomes

$$p(\sigma|\boldsymbol{\theta}_{-\boldsymbol{\sigma^2}}, \boldsymbol{y}, \boldsymbol{X}) \propto \sigma^{-n-1}\exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}}{2}/\sigma^2\right), \tag{7}$$

where $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. This is the kernel of an inverse gamma pdf, $\mathcal{IG}\left(\frac{n}{2}, \frac{\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}}{2}\right)$[4]. The initial value of $\boldsymbol{\theta}_{-\boldsymbol{\sigma^2}}$ becomes $\boldsymbol{\theta}_{-\boldsymbol{\sigma^2}}^{(0)} = (\boldsymbol{\beta}^{(1)'}\ \boldsymbol{\iota}_n)'$, hence $\sigma^{(1)}$ is generated as follows:

$$\sigma^{(1)}|\boldsymbol{\theta}_{-\boldsymbol{\sigma^2}}^{(0)} \sim \mathcal{IG}\left(\frac{n}{2}, \frac{\boldsymbol{\varepsilon}^{(1)'}\boldsymbol{\varepsilon}^{(1)}}{2}\right),$$

where $\boldsymbol{\varepsilon}^{(1)} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(1)}$. For $r \geq 2$ simulations, the above is replaced by

$$\sigma^{(r)}|\boldsymbol{\theta}_{-\boldsymbol{\sigma^2}}^{(r-1)} \sim \mathcal{IG}\left(\frac{n}{2}, \frac{\boldsymbol{\varepsilon}^{(r)'}\boldsymbol{\Omega}^{(r-1)^{-1}}\boldsymbol{\varepsilon}^{(r)}}{2}\right). \tag{8}$$

[4]We denote $f(x|\alpha,\beta) \propto x^{-(\alpha+1)}\exp(-\beta/x)\mathbf{1}_{[0,+\infty)}(x)$ distribution as $\mathcal{IG}(\alpha,\beta)$, inverse gamma with parameters α and β. Such x has $\mathrm{E}(x) = \beta/(\alpha-1)$ for $\alpha > 1$, and $\mathrm{Var}(x) = \beta^2/((\alpha-1)^2(\alpha-2))$ for $\alpha > 2$. See, *e.g.,* Robert and Casella (2004, p.582). Our fully conditional σ pdf is given in equation (7), however, has σ^2 in the denominator in the exponential function. To match the inverse gamma expression to the above, make a transformation of random variables from x to σ and we obtain $f(\sigma|\alpha,\beta) \propto \sigma^{-(2\alpha+1)}\exp(-\beta/\sigma^2)\mathbf{1}_{[0,+\infty)}(\sigma)$. $\mathcal{IG}(\cdot)$ expression in the text, thus, follows.

3.3 Simulating $\boldsymbol{\omega}$ Using HCCM Information

3.3.1 Conditional pdf of ω_i

Let us now derive conditional distribution, $\omega_i|\boldsymbol{\theta}_{-\omega_i}, \boldsymbol{y}, \boldsymbol{X}$. It is easy to see that the conditional distribution of $\boldsymbol{\omega}$ is in the form,

$$p(\boldsymbol{\omega}|\boldsymbol{\theta}_{-\boldsymbol{\omega}}, \boldsymbol{y}, \boldsymbol{X}) \propto |\boldsymbol{\Omega}|^{-1} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\right) \mathbf{1}_{[0.025,\ 12.5]}(\boldsymbol{\omega}).$$

Inside the exponential function in the above, $\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$ becomes $\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} = \sum_{t=1}^{n}\left(\frac{\varepsilon_i}{\sqrt{\omega_i}}\right)^2 = \boldsymbol{\varepsilon}^{*'}\boldsymbol{\varepsilon}^*$, where $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)' \sim n \times 1$ and $\varepsilon_i^* = \varepsilon_i/\sqrt{\omega_i}$, since $\boldsymbol{\Omega} = \mathrm{diag}(\boldsymbol{\omega})$. Note also $|\boldsymbol{\Omega}|^{-1} = \prod_{t=1}^{n}\omega_i^{-1}$, we then have for a particular ω_i,

$$p(\omega_i|\boldsymbol{\theta}_{-\omega_i}, \boldsymbol{y}, \boldsymbol{X}) \propto \omega_i^{-1}\exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}/\omega_i\right) \mathbf{1}_{[0.025,\ 12.5]}(\omega_i),$$

where $\varepsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ and $\boldsymbol{x}_i$ is the ith row of $\boldsymbol{X}$. The above is not a tractable pdf for two reasons, however. First, if we did not have a bounds constraint, $[\underline{\omega},\ \overline{\omega}]$, this function cannot be integrated, *i.e.,* $\int_0^\infty \omega_i^{-1}\exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}/\omega_i\right) d\omega_i$ is not finite. Secondly, even if we let the first parameter of the inverse gamma pdf to be some positive value $n_o > 1$[5], it would be awfully difficult to simulate a truncated inverse gamma random variable that has parameters, $\mathcal{IG}(n_o, \frac{\varepsilon_i^2}{2\sigma^2})$, in a Gibbs scheme, *i.e.,* always accepting whatever is generated.

3.3.2 HCCM Information to be used in M-H

Let us, for a moment, depart from the problem of generating ω_i, and see what the sampling theory has to tell us about the heteroscedastic $\boldsymbol{\varepsilon}$. An OLS estimated residual vector, $\boldsymbol{e} = (e_1, \ldots, e_n)'$, is already defined for equation (1). Eicker (1963) and White (1980)'s result may be used to form a sandwitch variance-covariance matrix estimator of $\hat{\boldsymbol{\beta}}$, HCCM, that

[5] We note that setting such $n_o > 0$, when it is not, is quite arbitrary. To deal with this possibly unidentified conditional posterior pdf, we propose to set bounds on the $\boldsymbol{\omega}$ vector, and employ an M-H scheme, as discussed in section 3.3.3, below. Instead of this avenue of approach, we might assume a prior on the first parameter in the inserve gamma pdf, to turn our model to a hierarchical Bayes model.

is consistent: $\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\hat{\boldsymbol{H}}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$, where $\hat{\boldsymbol{H}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' e_i^2$ and $\boldsymbol{X}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \sim K \times n$. Under the usual assumptions such as a full rank matrix $\lim \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} \equiv \boldsymbol{Q} < \infty$, it is well known that

$$\hat{\boldsymbol{H}} \xrightarrow{p} \boldsymbol{H} = \lim \frac{\sigma^2}{n}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X} = \lim \frac{\sigma^2}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\omega_i \tag{9}$$

holds. We may, hence, postulate a multivariate regression model of the following:

$$\hat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{V}, \tag{10}$$

where $\boldsymbol{V} \sim K \times K$ is an error term matrix that is constrained to make both $\boldsymbol{H}$ and $\hat{\boldsymbol{H}} > 0$. [6] After some manipulations (see the appendix to this paper), the above becomes

$$\hat{\boldsymbol{h}} = \boldsymbol{\mathcal{X}}_n\sigma^2\boldsymbol{\omega} + \boldsymbol{v}, \tag{11}$$

where $\hat{\boldsymbol{h}} = \text{vech}(\hat{\boldsymbol{H}}) \sim K' \times 1$, $\boldsymbol{\mathcal{X}}_n = [\text{vech}(\boldsymbol{x}_1\boldsymbol{x}_1') \ldots \text{vech}(\boldsymbol{x}_n\boldsymbol{x}_n')] \sim K' \times n$, $\boldsymbol{\omega} = (\omega_1 \ldots \omega_n)' \sim n \times 1$, and $\boldsymbol{\omega} \subset \Re_+^n$, $\boldsymbol{v} = \text{vech}(\boldsymbol{V}) \sim K' \times 1$, and $K' \equiv \frac{1}{2}K(K+1)$. This is a regression of K' dimension $\hat{\boldsymbol{h}}$ variable on $K' \times n$ explanatory variable matrix $\boldsymbol{\mathcal{X}}_n$. Hence the degrees of freedom, here, is df $= K' - n$, and under the usual situations such "df" could even be negative, *i.e.*, the $\hat{\boldsymbol{h}}$ regression is unestimable. We have now found that the $\boldsymbol{\omega}$ parameters in equation (11) are identifiable only when the $\boldsymbol{\beta}$ vector in equation (1) is identifiable. To restate,

> the $\boldsymbol{\omega}$ vector is identifiable if the inequalities $K' > n$ and $n > K$ simultaneously hold.

In a typical case like $n = 50$ and $K = 3$, $\boldsymbol{\omega}$ is unidentifiable since $K' = 6$, however, take a case like $n = 20$ and $K = 10$ then $K' = 55$ which makes $\boldsymbol{\omega}$ to be identifiable. But this is, indeed, a rare case. $\boldsymbol{\omega}$ may become identifiable

[6] We use the notation "$\boldsymbol{A} > \boldsymbol{0}$" to denote that the matrix $\boldsymbol{A}$ is positive definite.

if the column dimension of $\boldsymbol{X}$ increases, then, the "sample size" in equation (11), *i.e.,* K', should also increase. What we need to do is to increase K' without altering $\boldsymbol{e}$, estimated regression residual vector from equation (1), that constitutes a part of dependent variable in equation (11).

Amemiya (1983) proposed "partially generalized least squares" estimation method to improve upon efficiency over OLSE. We may use his idea to circumvent the problem that we stated in the previous paragraph. Let the regressor matrix in equation (1) to be

$$\boldsymbol{Z} = (\boldsymbol{X} \;\; \boldsymbol{W}) \sim n \times (K + K_W), \tag{12}$$

where we require that $\boldsymbol{W}$ is orthogonal to *all* the variables in equation (1), *i.e.,* $\boldsymbol{W}'(\boldsymbol{y} \quad \boldsymbol{X}) = \boldsymbol{0} \sim K_W \times (K+1)$ [7]. It is easy to see that $\boldsymbol{b}$ and $\boldsymbol{e}$ from equation (1) remain the same even after we regress $\boldsymbol{y}$ on $\boldsymbol{Z}$ instead of $\boldsymbol{X}$ alone. We now note that the the identifiability condition given in the previous paragraph becomes $K'' > n$ and $n > K$, where $K'' \equiv (K + K_W)(K + K_W + 1)/2$. We could now increase K_W to the extent that $K'' > n$ condition is satisfied[8]. In summary, we now have the control over the number of additional orthogonal variables, K_W, so that the above two conditions are met, while the $\boldsymbol{e}$ vector is unchanged.

3.3.3 HCCM Information Accomodated in Candidate Density

We have thus shown that using $\boldsymbol{Z}$ instead of $\boldsymbol{X}$ as the regressor matrix in equation (1), the $\boldsymbol{\omega}$ vector is always identifiable. Let the OLSE of $\sigma^2\boldsymbol{\omega}$ be $\sigma^2\hat{\boldsymbol{\omega}}$. "$\sigma^2$" has no hat on it since in our Bayesian MCMC, σ^2 is already generated in the previous step by the Gibbs sampler scheme. We turn to an

[7]It is well known that the OLSE of $\boldsymbol{\beta}$ for the transformed regression, $\boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}'\boldsymbol{\epsilon}$, is identical to the OLSE for the original equation. Incidentally, Amemiya *op.cit.* showed that the GLS estimator for the transformed model is more efficient than the OLSE.

[8]For instance, take the previous numerical case of $n = 50$ and $K = 3$. If we set $7 < K_W < 47$ then the conditions are satisfied

M-H algorithm, to generate ω_i. For this purpose, we use a truncated normal as our candidate generating density:

$$\omega_i \sim N(\hat{\omega}_i,\ \tilde{c} \times \hat{\omega}_i^2)\mathbf{1}_{\omega_i \in (.25,\ 12.5)}, \tag{13}$$

for $i = 1, \ldots, n$ where $\tilde{c}$ is a constant that has been determined in the following way. In the footnote that discusses inverse gamma distribution (footnote 7), we see that $\mathrm{Var}(x) = \mathrm{E}(x)^2(\alpha - 2)^{-1}$ for $x \sim \mathcal{IG}(\alpha, \beta)$. Hence, if we were to let α a little over "2" so that the variance would not vanish, then almost any value of $\tilde{c}$ suffices. The value ought to be such that the average acceptance rate in the independence chain M-H to be 20% to 50%. The value we came up with is $\tilde{c} = 15$ after some experiments, and the acceptance rate for such candidate generating density function, in our experiment, is 33%, which is reasonable[9]. For $r \geq 2$ sequence, equation (13) has been emplyed as the candidate generating density.

3.4 Hybrid of Gibbs Sampler and M-H

Let us describe our MCMC sequence in its entirety. For the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}'\ \sigma\ \boldsymbol{\omega}')'$, we set the initial value $\boldsymbol{\theta}^{(0)} = (\hat{\boldsymbol{\beta}}'\ s^2\ \boldsymbol{\iota}_n')'$ to generate the first round $\boldsymbol{\theta}^{(1)}$. Note here that we first draw $\boldsymbol{\beta}^{(1)}$, hence $\hat{\boldsymbol{\beta}}$ is not really needed in $\boldsymbol{\theta}^{(0)}$. What we actually need is $\boldsymbol{\theta}^{(0)}_{-\boldsymbol{\beta}}$. The sequence of our MCMC continues in the way specified by equations (6), (8) and (13) for $r \geq 2$, combining the Gibbs sampler and M-H. The details of monitoring and diagnostic checking concerning this MCMC sequence, is given in the next section.

[9] In M-H, an acceptance rate between 20% and 50% is regarded as reasonable, *e.g.,* see Koop (2003, p.98). In fact in a series of experiments, we found that the average acceptance rate of 21% when $\tilde{c} = 10$, 49% when $\tilde{c} = 25$, and 58% when $\tilde{c} = 35$. We may add that despite these differences in the average acceptance rates, however, the results are fairly robust.

4 Numerical and Empirical Investigation

We first generate a data set to make sure that our Bayesian procedure works. After that, we move on to use a Japanese stock price data set to show that our Bayesian method can be used to estimate stock return volatilities without assuming *any* structure on them *a priori.*

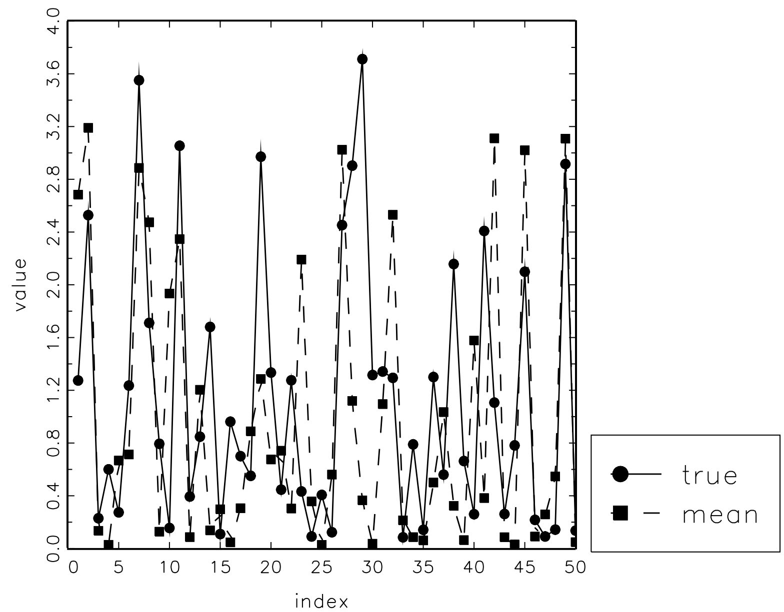
4.1 Numerical Illustration

For the generation purpose, we let the parameters of our model be $\boldsymbol{\beta} = (1, 3, -2)'$; $\sigma^2 = 10$; sample size $n = 50$; $\boldsymbol{X} = (\boldsymbol{\iota}_n, \boldsymbol{x}_2, \boldsymbol{x}_3) \sim n \times 3$, where $\boldsymbol{\iota}_n$ is an $n \times 1$ vector of all one's, $\boldsymbol{x}_2 \sim N(3, I_n)$, and $\boldsymbol{x}_3 \sim N(6, 2^2 I_n)$. As to $\boldsymbol{\omega}$, we need to specify total of 50 elements. We generated ω_i's using an uniform random number generator. Assigned ω_i's will be given in a chart (Figure 1) that compares them to the Bayesian posterior mean values, later. Using these prespecified set of parameter values, we generated data on $\boldsymbol{y}$. This is used to generate $\boldsymbol{W}$ matrix to form $\boldsymbol{Z} = (\boldsymbol{X}\ \boldsymbol{W})$[10]. By regressing $\boldsymbol{y}$ on $\boldsymbol{Z}$, we obatin one set of $\boldsymbol{e} \sim n \times 1$, and hence a $\hat{\boldsymbol{h}} \sim K'' \times 1$ vector. We now have one sample of $\boldsymbol{y}$, $\boldsymbol{X}$ and $\hat{\boldsymbol{h}}$.

Let us introduce the following notation to combine σ^2 and $\boldsymbol{\omega}$: $\sigma^2 \boldsymbol{\omega} \equiv \underset{\sim}{\boldsymbol{\omega}} \sim n \times 1$. Assigned or true vector in the numerical illustration is denoted by putting "A" as $\underset{\sim}{\boldsymbol{\omega}}^A \sim n \times 1$. In the MCMC simulation, ith simulated posterior of $\sigma^2 \omega_i$ is denoted as $\underset{\sim}{\omega}_i$, where $i = 1, \ldots, n$. Actually, there are R of $\underset{\sim}{\omega}_i$ for each i, hence this is a vector but we will ignore "r" subscript since it is too cumbersome to attach it, where $r = 1, \ldots, R$ and R is the number of replications in the MCMC. In the vector form, this becomes $\underset{\sim}{\boldsymbol{\omega}} \sim n \times 1$. Now, the Bayes estimator or the posterior mean of $\underset{\sim}{\omega}_i$ is given by $\hat{\underset{\sim}{\omega}}_i$, and as

[10]In our *GAUSS* programming, we used the "NULL($\cdot$)" command to obtain $\boldsymbol{W} \sim n \times K_W$ that is orthogonal to $\boldsymbol{X}$. NULL command uses the QR decomposition to obtain the orthogonal complement of $\boldsymbol{X}$.

Figure 1: Fifty Posterior Means $\hat{\underset{\sim}{\boldsymbol{\omega}}}$ versus $\underset{\sim}{\boldsymbol{\omega}}^A$

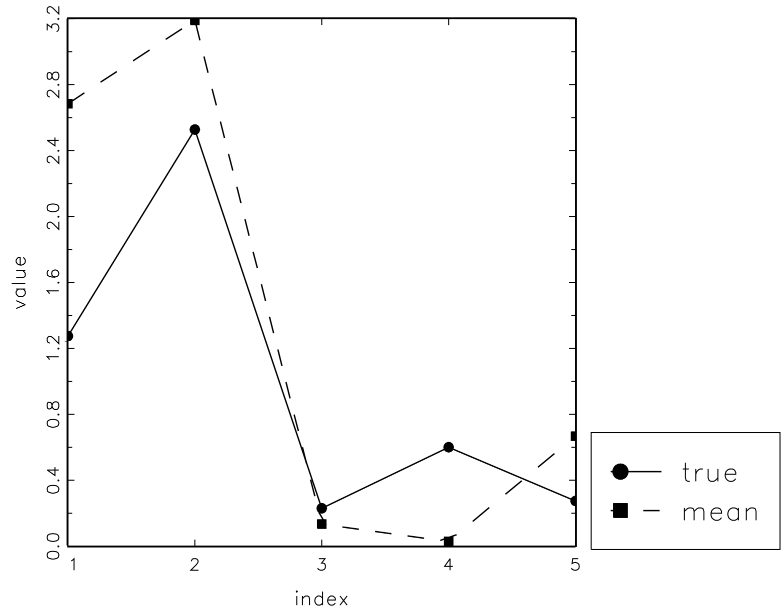


a vector $\hat{\underset{\sim}{\boldsymbol{\omega}}} \sim n \times 1$.

In the MCMC computation, we drew total of 30,000 $\boldsymbol{\theta}$'s. After some experiments, using such diagnostics devices as Geweke's (1992), we decided to throw away the first 10,000 simulated values as burn-in's, and took every fifth drawn values after that to lessen the possible serial correlation inherent in MCMC. The average acceptance rate in the M-H step was 33%. We now have a set of $20,000/5 = 4,000$ simulated values for each of $\underset{\sim}{\omega_i}$'s $(i = 1, \ldots, 50)$ available. Since giving 50 posterior pdf's of $\underset{\sim}{\omega_i}$'s would not contribute to the understanding of our procedure, we decided to present posterior means of the fifty $\underset{\sim}{\omega_i}$'s instead, and this is given in Figure 1.

Fifty assigned values are given using a solid line, and posterior means using a dotted line. We see that there are some discrepancies between the true (or given) value $\underset{\sim}{\boldsymbol{\omega}}^A$ and the posterior mean value, $\hat{\underset{\sim}{\boldsymbol{\omega}}}$, however, mostly they are close. Their closeness may be observed more clearly for the first

Figure 2: First Five $\underset{\sim}{\hat{\omega}}_1$ to $\underset{\sim}{\hat{\omega}}_5$



five $\underset{\sim}{\hat{\omega}}_i$'s of Figure 1, in Figure 2. In particular, posterior means $\underset{\sim}{\hat{\omega}}_1$ to $\underset{\sim}{\hat{\omega}}_3$ are seen to follow randomly assigned values $\underset{\sim}{\omega}_1^A$ to $\underset{\sim}{\omega}_3^A$ very closely.

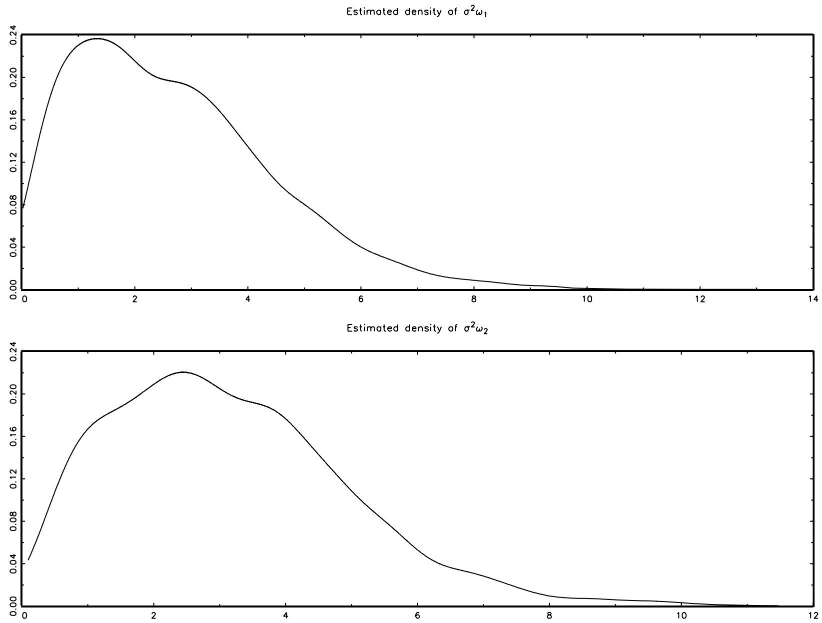
Encouraged by this observation, we computed a measure of closeness of a vector $\boldsymbol{\eta}$ to the assigned value vector $\underset{\sim}{\boldsymbol{\omega}}$, SS_η, by

$$SS_{\boldsymbol{\eta}} = (\underset{\sim}{\boldsymbol{\omega}} - \boldsymbol{\eta})'(\underset{\sim}{\boldsymbol{\omega}} - \boldsymbol{\eta}).$$

We compared $SS_{\underset{\sim}{\hat{\boldsymbol{\omega}}}}$ and $SS_{\hat{\boldsymbol{\sigma}}^2}$, where $\hat{\boldsymbol{\sigma}}^2 = \hat{\sigma}^2\boldsymbol{\iota}$ and $\hat{\sigma}$ is the OLS estimator of σ assuming homoscedasticity, *i.e.,* $\hat{\sigma}^2 = (\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}})/(n-K)$. The results are $SS_{\underset{\sim}{\hat{\boldsymbol{\omega}}}} = 43.04 < SS_{\hat{\boldsymbol{\sigma}}^2} = 57.61$ and this shows that our Bayesian estimates $\underset{\sim}{\hat{\boldsymbol{\omega}}}$ indeed performed better compared to the benchmark of $\hat{\sigma}^2\boldsymbol{\iota}$ that assumes homoscedasticity.

We depicted the marginal posterior pdf's of the first two simulated $\underset{\sim}{\omega}_i$'s in Figure 3. We see that the two marginal posterior pdf's seem to be equipped with typical characteristics of a variance pdf, *i.e.,* a gamma like shape with a hump close to the origin and skewed to the right. We emphasize that the

Figure 3: First Two Marginal Posterior pdf's : $\underset{\sim}{\omega}_1$ and $\underset{\sim}{\omega}_2$



shapes have been obtained without any prior shape information on them. This fact may indicate an appropriateness of our MCMC simulator.

We next examine the posterior standard deviations of the simulated $\underset{\sim}{\omega}_i$'s. The 50 posterior standard deviations are given in Figure 5. They do fluctuate. For instance, at $i = 2$, as indicated in the previous paragraph, posterior mean is about 2.5, and from this chart, *i.e.,* Figure 3, standard deviation, 1.8, which is not small at all. But if we compare Figure 4 to the previous Figure 1, we notice a tendency that posterior means and standard errors, move in the same direction.

4.2 Empirical Investigation: Japanese Stock Returns

4.2.1 Theoretical Model

In this section, we use a panel data set to show that our procedure works well in practice. Saying "panel data set" may be far stretched, however, since we do not implement, for example, random effects model. We carry

Figure 4: Standard Deviations of $\underset{\sim}{\boldsymbol{\omega}}$

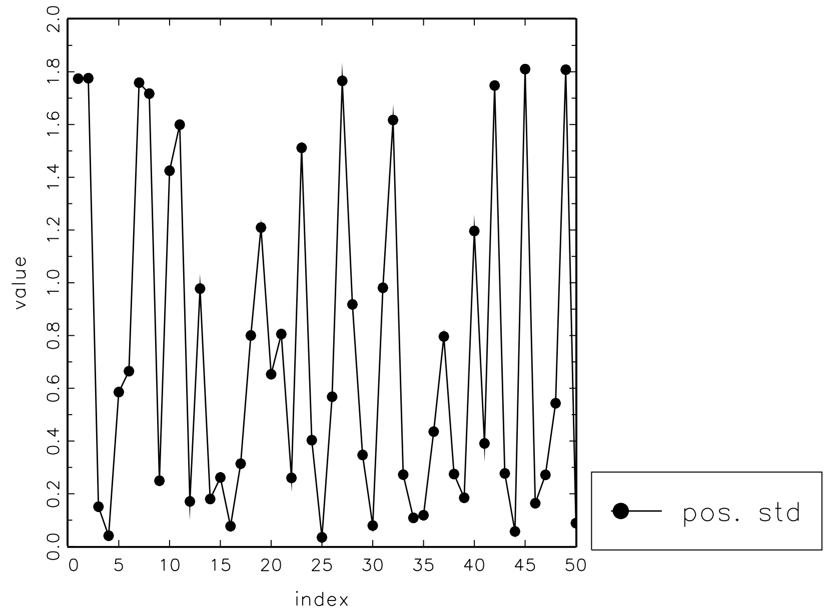
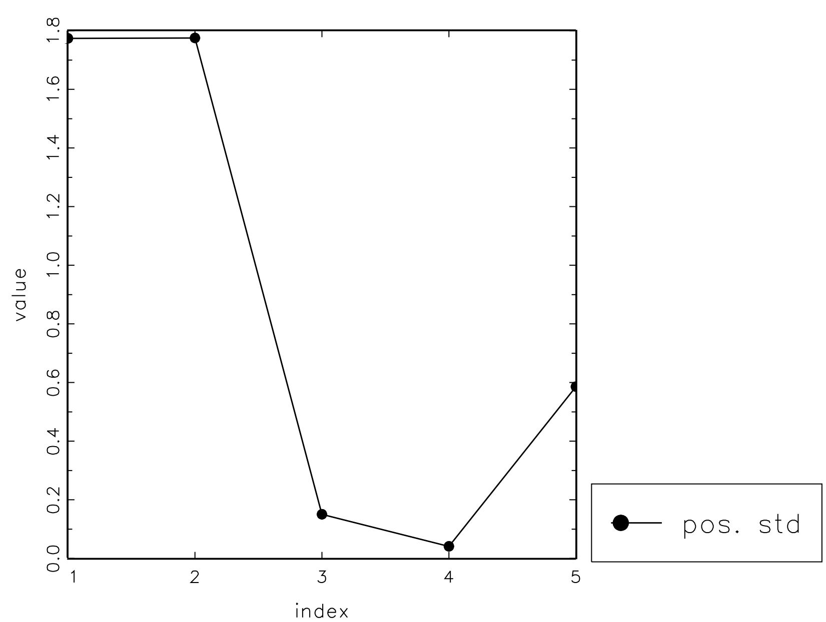


Figure 5: First Five Standard Deviations of $\underset{\sim}{\omega}_1$ to $\underset{\sim}{\omega}_5$



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 venture capital 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 a multifactor multivariate return generating equation:

$$\boldsymbol{R} = (\boldsymbol{\iota}_T \ \boldsymbol{F}) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{B} \end{pmatrix} + \boldsymbol{\varepsilon}, \tag{14}$$

where
$\boldsymbol{R} = (\boldsymbol{R}_1 \cdots \boldsymbol{R}_N) \sim T \times N$ is a T period excess returns for N firms,
$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N) \sim 1 \times N$ vector of constants,
$\boldsymbol{B} = (\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_N) \sim F \times N$ is a matrix of beta's,
$\boldsymbol{F} = (\boldsymbol{f}_1 \cdots \boldsymbol{f}_F) \sim T \times F$ is a matrix of F factors,
$\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1 \cdots \boldsymbol{\varepsilon}_N) \sim T \times N$ matrix of error terms,

N is the number of stocks, and T is the time series sample size. When $F = 1$ the above reduces to a single factor return generating equation:

$$\boldsymbol{R} = (\boldsymbol{\iota}_T \ \boldsymbol{f}) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\varepsilon}, \tag{15}$$

where
$\boldsymbol{F} = \boldsymbol{f}$ is a $T \times 1$ vector of one factor,
$\boldsymbol{B} = \boldsymbol{\beta}$ is a $1 \times N$ vector of beta's.

Equation (15) 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 (15). Define sample mean of $\boldsymbol{R}$ to be an N dimensional vector $\bar{\boldsymbol{R}}$, we then obtain a cross sectional regression model

$$\bar{\boldsymbol{R}} = \hat{\boldsymbol{\beta}}' \lambda + \boldsymbol{u}, \tag{16}$$

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,
$\lambda \sim 1 \times 1$ scalar is a risk premium associated with the factor $\boldsymbol{f}$,
$\boldsymbol{u} \sim N \times 1$ is a vector of pricing errors.

Equation (16) is the one factor type CAPM *without* an intercept term given in Cochrane *op.cit.*, p.235, among others[11].

In this section so far, we have used a set of notations that are common in empirical finance, and in this paper it corresponds to equation (1). We, thus, need to clarify the notational correspondences between the ones used so far in the current section, and in the previous subsections. The correspondences are given in below.

$\bar{\boldsymbol{R}} \sim N \times 1$	(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)),
$F = 1$ is the number of factors	(corresponds to K in (1)),
N the number of stocks	(corresponds to n in (1)).

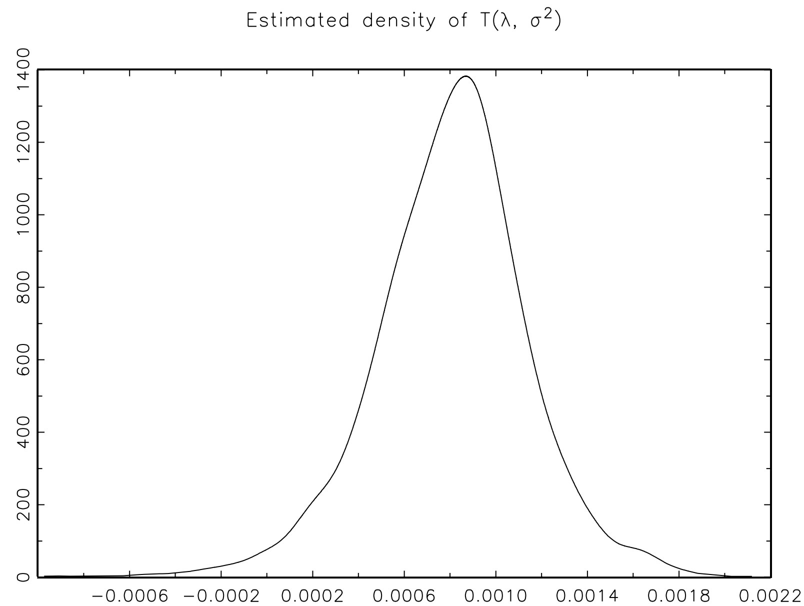
4.2.2 Bayesian Test of Hetoroscedasticity

Since we do not know whether our data set exhibits a heteroscedasticity or not, we carried out a Bayesian test of heteroscedasticity *a la* Lancaster(1994, p.149). In equation (1) or equivalently in the present case equation (16), under the null hypothesis of homogeneity, H_0, we have $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$. Hence, letting $\varepsilon_i = \bar{R}_i - \hat{\beta}_i \lambda$, we have $\mathrm{E}(\varepsilon_i^2 - \sigma^2)^2 = \mathrm{E}(\varepsilon_i^4 - \sigma^2 \varepsilon_i^2 + \sigma^4) = 2\sigma^4$, under H_0, since our normality assumption on ε_i calls for $\mathrm{E}(\varepsilon_i^4) = 3\sigma^4$. Lancaster *op.cit.*, thus, proposes $T(\lambda, \sigma^2) = \left(\frac{\sum_{i=1}^{N} (\hat{\varepsilon}_i^2 - \sigma^2)^2}{N} \right) - 2\sigma^4$, where $\hat{\varepsilon}_i = \bar{R}_i - \hat{\beta}_i \hat{\lambda}$ and $\hat{\lambda}$ is an MCMC generated value, to be used in a Highest Posterior Density Interval (HPDI) test. This test would reject H_0 if the mass of the T statistic distribution does not contain "zero," *i.e.,* homoscedasticity[12].

[11] 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 0.003 (0.048) and the slope estimate is 0.104 (0.062), where the figures inside the parentheses are estimated standard errors. Without an intercept term, the slope estimate is 0.108 (0.024), and there is very little difference whether we include an intercept term or not.

[12] Lancaster *op. cit.* does not call T, a test statistic, however. He rightfully regards T as a "checking" device. For one thing, the empirical distribution obtained from simulated

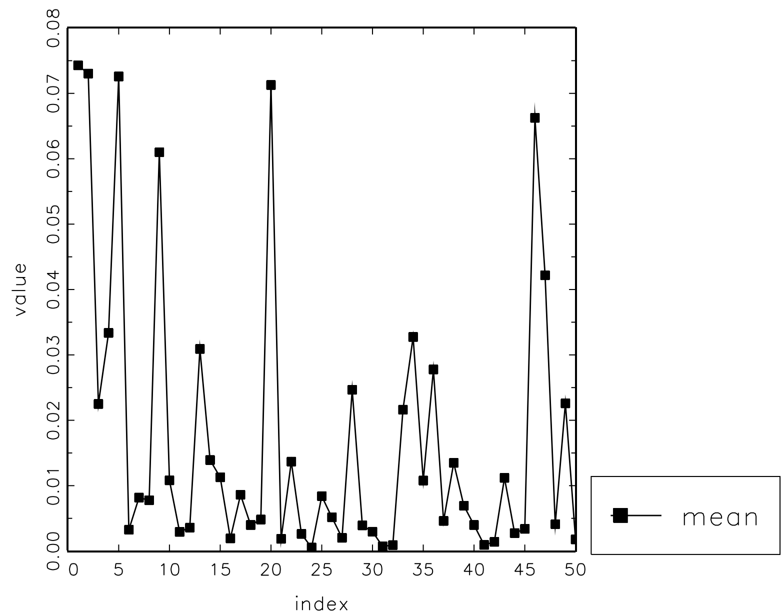
Figure 6: Marginal Posterior pdf of the T Statistic



It is easy to draw λ and σ from their respective marginal posterior pdf, since each pdf boils down to an univariate normal and an $\mathcal{IG}$, respectively. We will not discuss the details of the MCMC procedure involved in simulating T under H_0. Simulated T statistic is shown in Fig. 4. Our calculations show that for the distribution of T, $P(T < 0) = 0.016$, which is smaller than the usual significance level even if we multiply it by 2, *i.e.*, $0.016 \times 2 = 3.2\% < 5\%$, say. We would safely reject H_0. We also carried out several well-known sampling theory heteroscedasticity tests. They all reject H_0.

T is *not* the marginal posterior pdf under H_0. In other words, we (or Lancaster *op cit.* too) did not compute the posterior pdf of the T statistic. Secondly, in a sampling theory framework, obviously $\text{plim}T = 0$ under H_0 as $N \to \infty$. This implies that whether $T > 0$ or $T < 0$ is not known under the alternative hypothesis as $N \to \infty$. It seems like we ought to take a $100(1-\alpha)\%$ HPDI of the T statistic, and reject H_0 if it does not contain zero, where α is a significance level. This is in essence, what we did.

Figure 7: Posterior Means : $\hat{\underset{\sim}{\omega}}_i$'s



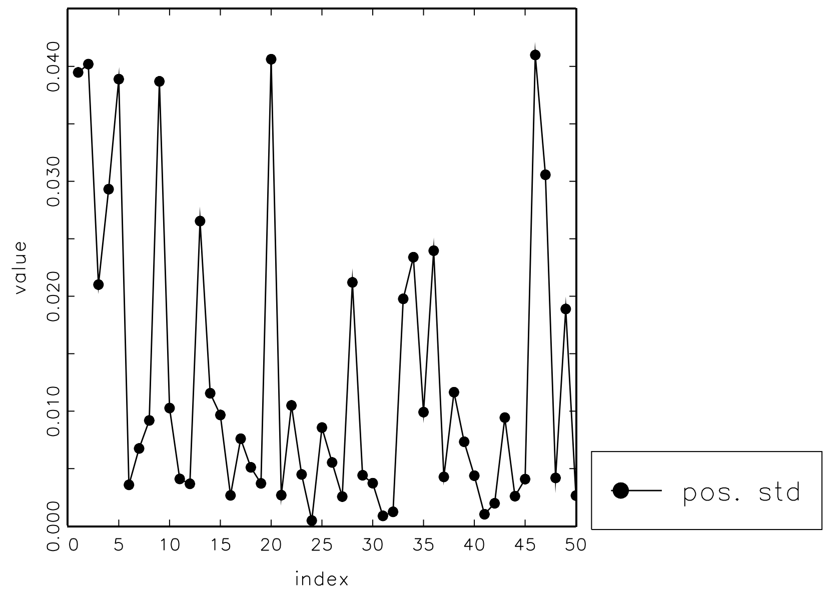
4.2.3 Empirical Estimation of $\underset{\sim}{\omega}_i$'s

The MCMC simulation procedure, here, is exactly the same as the one discussed in the numerical illustration of section 4.1. Posterior means and standard deviations are given, respectively, in Figures 7 and 8. Since we have confirmed that our method works in section 4.1, we need to make sure that our results may be reasonable as an empirical investigation.

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[13]. 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

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

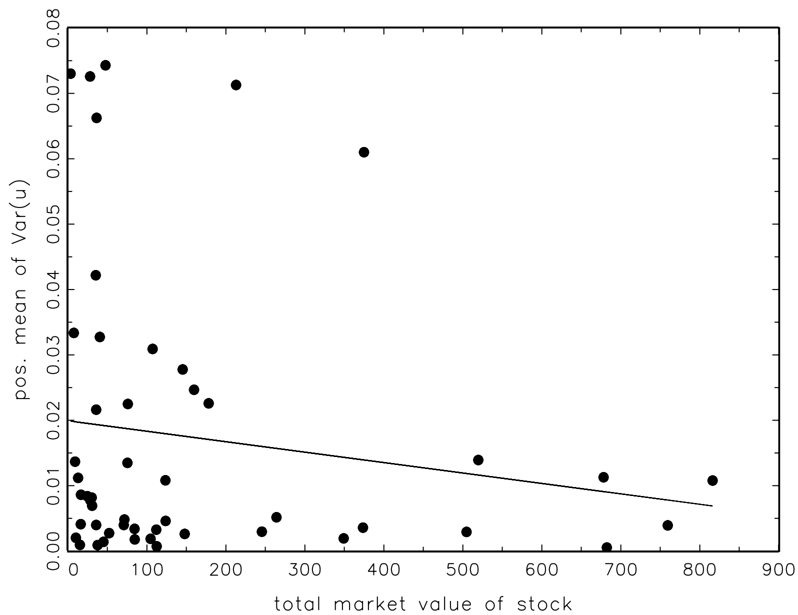
Figure 8: Standard Deviations of $\underset{\sim}{\boldsymbol{\omega}}$



by the market value of its equity (MVE) but some other variables such as sales. In this section, we intend to investigate whether larger (smaller) size companies have smaller (larger) $\underset{\sim}{\omega}_i$'s.

For the fifty stock data, we now collected MVE data from Yahoo Finance (in one billion yen). We then drew a graph with the MVE on the horizontal axis, and the $\underset{\sim}{\hat{\omega}}_i$'s on the vertical axis. This is shown in Figure 9. The downward sloping solid line in Figure 9, is the OLS estimated line. This figure clearly shows the larger the MVE, the smaller the volatility as measured by $\underset{\sim}{\hat{\omega}}_i$. We further selected two stocks (1) that has large MVE and small $\underset{\sim}{\hat{\omega}}_i$, and (2) that has small MVE and large $\underset{\sim}{\hat{\omega}}_i$, to see what the marginal posterior pdf's, $\underset{\sim}{\omega}_i$, of these companies look like. Specifically, we chose Taisho Pharmaceutical Co., Ltd. for (1), and Site Support Institute Co., Ltd. for (2). They are given in Figures 10 below. Notice that the two charts have different horizontal axis scale. The smaller MVE stock has very

Figure 9: Volatility versus Size: $\underset{\sim}{\hat{\omega}}_i$'s and MVE



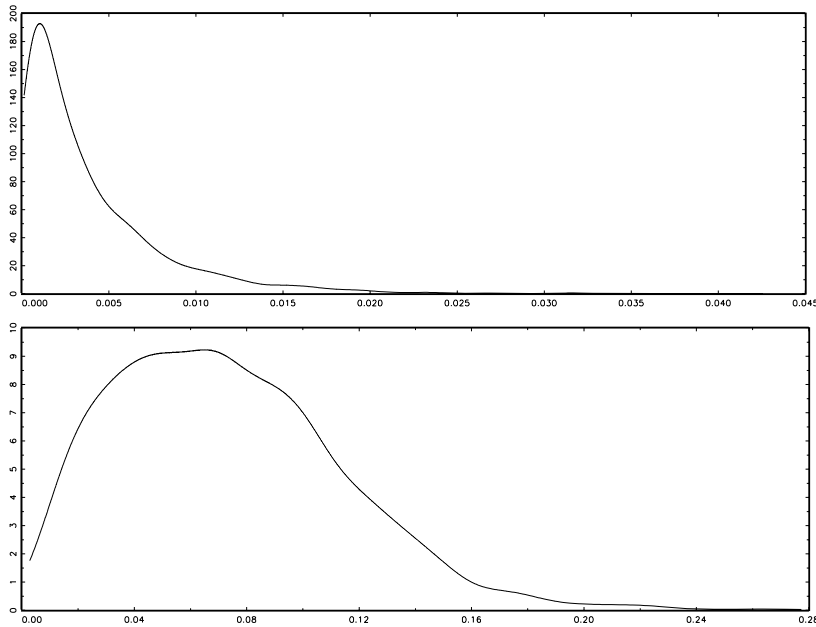
large volatility (see the lower chart) compared to the that of the larger MVE stock (see the upper chart). The two pdf's have quite reasonable shapes. We conclude that our Bayesian estimation of volatility supports the view that the smaller the size of the stock, the larger is the volatility.

5 Concluding Remarks

In this paper, we proposed a fully Bayesian method to estimate regression error term scedastic function $\boldsymbol{\omega}$ *without* assuming any structure on it. One unique feature of our methodology is in using the Eicker-White result in a candidate generating function in our M-H step of MCMC. We have supplied a numerical example, and also an empirical example to show that our method works reasonably. Finally, as demonstrated in our empirical research, obtaining $\boldsymbol{W}$ such that $\boldsymbol{W}'(\boldsymbol{y}\ \boldsymbol{X}) = \boldsymbol{0}$ is no problem since $\boldsymbol{W}$ is not found as an empirical data but rather computer generated data.

One possible arbitrariness might remain in this paper. That is, a reader

Figure 10: Marginal Posterior pdf $\underset{\sim}{\omega}_i$ for Taisho Phamaceutical Co. and Site Support Institute



might feel uneasy about our setting bounds for $\boldsymbol{\omega}$ prior pdf. As we suggested in footnote 8, we may adopt a hierarchical Bayes model.

Acknowledgements
This research is supported in part by the Ministry of Education and Culture, Grant in Aid #19530178, and also by the Global COE Program, " Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi- Stat). " We thank Hiroki Tsurumi, Takeshi Amemiya, and Toshiaki Watanabe for incisive comments and suggestions for earlier drafts of this paper. We are responsible for all errors that inadvertently remain in this paper, however. Earlier versions of this paper has been presented at the 2008 Far Eastern Econometric Society Meeting in Singapore Management University as well as in Econometrics Workshop, Department of Economics, Hitotsubashi University.

References

[1] Amemiya, T., 1983, Partially generalized least squares and two-stage least squares estimators. *Journal of Econometrics,* 23, 275–283.

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

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

[4] Belsley, D.A., 2002, An investigation of an unbiased correction for heteroskedasticity and the effects of misspecifying the skedastic function. *Journal of Economic dynamics and Control,* 26, 1379–1396.

[5] Berk, J.B., 1997, Does size really matter? *Financial Analysts Journal,* 53, no.5, 12–18.

[6] Cameron, C.A. and P.K. Trivedi, 2005, *Microeconometrics: Methods and Applications,* Cambridge University Press.

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

[8] Chang, E.C. and S. Dong, 2006, Idiosyncratic volatility, fundamentals, and institutional herding: Evidence from the Japanese stock market. *Pacific-Basin Finance Journal,* 14, 135–154.

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

[10] 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.

[11] Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. in Bernardo, J., Berger, J., Dawid A., and Smith, A. (eds.), *Basyesian Statistics,* Volume 4, pp.641–649, Clarendon Press.

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

[13] Greene, W.H., 2008, *Econometric Analysis, 6th ed.,* Prentice-Hall.

[14] Koop, G., 2003, *Bayesian Econometrics,* John Wiley & Sons.

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

[16] Lütkepohl, H., 1996, *Handbook of Matrices,* John Wiley & Sons.

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

[18] Robert, C.P. and G. Casella, 2004, *Monte Carlo Statistical Methods,* 2nd ed., Springer.

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

[20] Wooldridge, J.M. 2002, *Econometric Analysis of Cross Section and Panel Data,* MIT Press.

Appendix : Derivation of equation (11)

$\boldsymbol{H}$ is defined by $\boldsymbol{H} = \boldsymbol{X'\Omega X}$. Our purpose in this appendix is to derive equation (9) in the text, *i.e.*, $\boldsymbol{h} \equiv \text{vech}(\boldsymbol{H}) = \boldsymbol{\mathcal{X}}_n\boldsymbol{\omega}$. To this effect, we note the following series of lemmas.

Lemma 1 . $\text{vec}(\boldsymbol{H}) = (\boldsymbol{X}' \otimes \boldsymbol{X}')\text{vec}(\boldsymbol{\Omega}) \sim K^2 \times 1$.

proof : Just use the usual $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B})$ formula. See, *e.g.,* Lütkepohl (1996, p.97).

Lemma 2 . $\boldsymbol{X}' \otimes \boldsymbol{X}' = [(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1) \dots (\boldsymbol{x}_1 \otimes \boldsymbol{x}_n)] \dots [(\boldsymbol{x}_n \otimes \boldsymbol{x}_1) \dots (\boldsymbol{x}_n \otimes \boldsymbol{x}_n)] \sim K^2 \times n^2$.

proof : Note $\boldsymbol{X}' = (\boldsymbol{x}_1 \dots \boldsymbol{x}_n) \sim K \times n$, and the result follows.

Lemma 3 . $\text{vec}(\boldsymbol{H}) = [(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1) \dots (\boldsymbol{x}_n \otimes \boldsymbol{x}_n)]\boldsymbol{\omega}$, where $\boldsymbol{\omega} = (\omega_1 \dots \omega_n)' \sim n \times 1$.

proof : Note

$$\text{vec}(\boldsymbol{\Omega}) = \text{vec}(\text{diag}(\boldsymbol{\omega})) = (\omega_1\, 0 \cdots 0 \vdots 0\, \omega_2\, 0 \cdots 0 \vdots \cdots \vdots \cdots \omega_n)' \sim n^2 \times 1,$$

where $\boldsymbol{\omega}' = (\omega_1, \dots, \omega_n)$, "diag($\boldsymbol{\omega}$)" denotes a diagonal matrix with the vector $\boldsymbol{\omega}$ in its diagonal. Note that $\text{vec}(\boldsymbol{\Omega})$ is a vector of size $n^2 \times 1$, and picks up the columns $(\boldsymbol{x}_t \otimes \boldsymbol{x}_t)$ only for $t = 1, \dots, n$ from $\boldsymbol{X}' \otimes \boldsymbol{X}'$. Lemmas 1 and 2, in conjunction with the above proves the lemma.

Lemma 4 . $\boldsymbol{x}_t \otimes \boldsymbol{x}_t = \text{vec}(\boldsymbol{x}_t\boldsymbol{x}_t') \sim K^2 \times 1$.

proof : Use the $\text{vec}(\boldsymbol{AB}) = (\boldsymbol{B}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{I}_p)$ relationship, where p is the column dimension of $\boldsymbol{A}$ (see *e.g.,* Lütkepohl (1996, p.97)). In the present case $p = 1$, hence, $\text{vec}(\boldsymbol{I}_p)$ is equal to a scalar one.

We now define a matrix that converts a $\text{vec}(\boldsymbol{A})$ type vector to a $\text{vech}(\boldsymbol{A})$ vector.

Definition of $\boldsymbol{L}_{K'}$ matrix : Let $K' \equiv K + (K-1) + \dots + 1 = \frac{1}{2}K(K+1)$ and define

$$\boldsymbol{L}_{K'} = \begin{bmatrix} \boldsymbol{I}_K & & & & 0 \\ & \begin{matrix} 0 \\ \vdots \end{matrix}\ \boldsymbol{I}_{k-1} & & & \\ & & \begin{matrix} 0 & 0 \\ \vdots & \vdots \end{matrix}\ \boldsymbol{I}_{k-2} & & \\ & & & \cdots\cdots & \\ 0 & & & & \\ & & & & 0 \ \cdots \ 0 \ \ 1 \end{bmatrix} \sim K' \times K^2.$$

In essence, the diagonal submatrices of $\boldsymbol{L}_{K'}$ shrink from $\boldsymbol{I}_K, \boldsymbol{I}_{K-1} \dots \boldsymbol{I}_{K-j} \dots$ to $\boldsymbol{I}_{K-(K-1)} = 1$ but each submatrix is preceded by a zero matrix $\mathbf{0}_j$, where $\mathbf{0}_j \sim (K-j) \times j$. Such $\boldsymbol{L}_{K'}$ converts vec($\boldsymbol{A}$) to vech($\boldsymbol{A}$) for a $K \times K$ square matrix $\boldsymbol{A}$ as follows.

Lemma 5 . Let $\boldsymbol{A} \sim K \times K$ square, then $\text{vech}(\boldsymbol{A}) = \boldsymbol{L}_{K'}\text{vec}(\boldsymbol{A}) \sim K' \times 1$.

proof : Let $\boldsymbol{A} = (\boldsymbol{a}_1 \dots \boldsymbol{a}_K) \sim K \times K$ and $\boldsymbol{a}_j \sim K \times 1$. Also let, for instance,

$$\boldsymbol{a}_{j(-1,2)} = (a_{3j} \ a_{4j} \ \cdots \ a_{Kj})' \sim (K-2) \times 1,$$

i.e., $\boldsymbol{a}_{j(-1,2)}$ deletes the first two elements from $\boldsymbol{a}_j$. Then,

$$\boldsymbol{L}_{K'}\text{vec}(\boldsymbol{A}) = \boldsymbol{L}_{K'} \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_K \end{pmatrix} = \boldsymbol{L}_{K'} \begin{pmatrix} \boldsymbol{a}_1 \\ \cdots\cdots \\ 0 \\ \boldsymbol{a}_{2(-1)} \\ \cdots\cdots \\ 0 \\ 0 \\ \boldsymbol{a}_{3(-1,2)} \\ \cdots\cdots \\ \vdots \\ a_{KK} \end{pmatrix} = \text{vech}(\boldsymbol{A}).$$

Lemma 6 . $\text{vech}(\boldsymbol{x}_t\boldsymbol{x}_t') = \boldsymbol{L}_{K'}\text{vec}(\boldsymbol{x}_t\boldsymbol{x}_t')$.

proof : Omitted.

We are now, in a position to derive the $\text{vech}(\boldsymbol{H}) = \boldsymbol{\mathcal{X}}_n\boldsymbol{\omega}$ relationship. Use Lemmas 2 to 4 on the Lemma 1 relationship to obtain

$$\text{vec}(\boldsymbol{H}) = [\text{vec}(\boldsymbol{x}_1\boldsymbol{x}_1') \dots \text{vec}(\boldsymbol{x}_n\boldsymbol{x}_n')]\boldsymbol{\omega}.$$

Further, use Lemmas 5 and 6 to obtain the following.

$$\begin{aligned} \text{vech}(\boldsymbol{H}) &= \boldsymbol{L}_{K'}\text{vec}(\boldsymbol{H})\boldsymbol{\omega} \\ &= [\boldsymbol{L}_{K'}\text{vec}(\boldsymbol{x}_1\boldsymbol{x}_1') \dots \boldsymbol{L}_{K'}\text{vec}(\boldsymbol{x}_n\boldsymbol{x}_n')]\boldsymbol{\omega} \\ &= \boldsymbol{\mathcal{X}}_n\boldsymbol{\omega}. \end{aligned}$$

where $\boldsymbol{\mathcal{X}}_n = [\text{vech}(\boldsymbol{x}_1\boldsymbol{x}_1') \dots \text{vech}(\boldsymbol{x}_n\boldsymbol{x}_n')] \sim K' \times n$.