



Discussion Paper Series

No.185

**Bayesian Estimation of
Unknown Heteroscedastic Variances**

Hiroaki Chigira
Tsunemasa Shiba

September 2006

**Hitotsubashi University Research Unit
for Statistical Analysis in Social Sciences**
A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
<http://hi-stat.ier.hit-u.ac.jp/>

Bayesian Estimation of Unknown Heteroscedastic Variances*

Hiroaki Chigira[†] and Tsunemasa Shiba[‡]

September 3, 2006

Abstract

We propose a Bayesian procedure to estimate possibly heteroscedastic variances of the regression error term, *without assuming any structure* on them. What we propose in this paper, may be construed as a *Conditional Bayesian* procedure that is conditioned upon the HCCM obtained from the OLS estimation of the original regression model. After we obtain the Eicker–White HCCM, we set up a Bayesian model and use an MCMC to simulate posterior pdf’s of heteroscedastic variances whose structures are unknown. In addition to the numerical examples, we present an empirical investigation on the stock prices of Japanese pharmaceutical and biomedical companies.

key words

Eicker–White HCCM, orthogonal regressors, conditional Bayesian, MCMC, stock return variance estimation

*We thank Hiroki Tsurumi of Rutgers University, and Takeshi Amemiya of Stanford University for critical comments and suggestions. We are responsible for all errors that inadvertently remain in this manuscript, however.

[†]Department of Economics, Hitotsubashi University, Kunitachi-shi, Tokyo 186 Japan.
e-mail: hchigira@mercury.ne.jp

[‡]Department of Economics, Hitotsubashi University, Kunitachi-shi, Tokyo 186 Japan.
e-mail: tshiba@stat.hit-u.ac.jp

1 Introduction

We propose a Bayesian procedure to estimate possibly heteroscedastic variances of the regression error term, *without assuming any structure* on them. Our focus is on the direct *estimation* of the diagonal elements of regression error term variance-covariance matrix, ω vector.¹ In a sampling theory asymptotics framework, Eicker (1963) and White (1980) developed a well-known consistent variance-covariance matrix estimator for the OLS regression coefficient estimator (“HCCM” for short hereafter). What we propose in this paper, may be construed as a *Conditional Bayesian* procedure that is conditioned upon the HCCM obtained from the OLS estimation of the original regression model. We note that our focus in this paper is *not* in statistical inferences of the regression coefficient vector, however, our results may be used to obtain a regression coefficient estimator that is relatively more efficient,² and to construct better performing tests of restrictions on the regression coefficient vector.

Our strategy to estimate the ω vector is in two stages. In the first stage, we obtain the HCCM from the estimation results of the original regression model using the OLS. Since we do not assume any *a priori* structure on ω , the resultant method is nonparametric in its nature up to this point. In the second stage, viewing the HCCM as a regression model, and assuming appropriate informative prior pdf’s, we simulate ω as a set of posterior random variables in a Bayesian model using an MCMC (Markov Chain Monte Carlo) method. In this sense, our entire method may be termed *Conditional Bayesian*³.

The current trend in the HCCM literature seems to focus on improving the finite sample performance of tests on the linear restriction(s) on the coefficient vector, *e.g.*, Long and Ervin(2000) and Godfrey (2006), among others.

The import of direct estimation of the ω vector needs few discussions. For instance, ω in stock return in finance, is nothing but the volatility.⁴ 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 ω needs no parametric model for the volatility process such as the GARCH model. Since our method depends on the OLS estimation of a regression model in the first stage to obtain an HCCM, and the fact that the HCCM estimator is nonparametric in its nature, our method is nonparametric as well. If we wish to estimate a volatility process in time series data without assuming any structure on it, what we usually do is to calculate a historical volatility series.

¹ ω vector is the diagonal elements of the regression error term variance-covariance matrix. It is formally defined just below equation (4).

²For instance, Robinson (1987) assuming error term scedastic function to be a function of regressors, thus the scedastic function is an unknown form, derived a GLS estimator that is more efficient than the existing ones.

³We are not aware of any paper that uses the term “Conditional Bayesian,” in the way we use it in this paper.

⁴Actually, in finance “standard deviation” instead of “variance”, is equivalent to the term “volatility”. Hence, to be precise, we should say that “elementwise square root of ω is volatility.”

But this is just a descriptive statistic without any theoretical background. Moreover, when it comes to cross section data, historical volatility calculation breaks down for obvious reasons. Our conditional Bayesian method, on the other hand, should provide a good deal of theoretical support for cross sectional data.

The rest of this paper is organized as follows. In section 2, we set our regression model, and use the Eicker–White result to formulate the HCCM regression model of the second stage conditional Bayesian inference. We make several remarks for this model in the section; all proofs are relegated to the appendix. Section 3 starts out with our Bayesian setup, that is followed by some numerical and empirical results. Our empirical investigation is on the stock prices of Japanese pharmaceutical and biomedical companies. Section 4 concludes.

2 The Model

For a heteroscedastic regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{X} \sim n \times K$ is a purely exogenous variable matrix, $\boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma^2\boldsymbol{\Omega})$, $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\omega}) = \text{diag}(\omega_1, \dots, \omega_n)$, $\text{tr}(\boldsymbol{\Omega}) = n$, σ^2 is a scalar and n is a sample size, a well-known consistent variance estimator of the $\boldsymbol{\beta}$ vector, when it is estimated by the OLS, is available.⁵ Let the OLSE (OLS estimator) of $\boldsymbol{\beta}$ be \mathbf{b} . Using \mathbf{b} we obtain the OLS estimated residual vector, $\mathbf{e} = (e_1, \dots, e_n)'$ as $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$. Eicker and White’s result may be used to form a sandwich variance-covariance matrix estimator of \mathbf{b} that is consistent:

$$\left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} \hat{\mathbf{H}} \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1},$$

where $\hat{\mathbf{H}} = \frac{1}{n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' e_t^2$ and $\mathbf{X}' = (\mathbf{x}_1, \dots, \mathbf{x}_n) \sim K \times n$. Under the usual assumptions such as a full rank matrix $\lim \frac{1}{n} \mathbf{X}'\mathbf{X} \equiv \mathbf{Q} < \infty$, it is well known that

$$\hat{\mathbf{H}} \xrightarrow{p} \mathbf{H} = \lim \frac{\sigma^2}{n} \mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = \lim \frac{\sigma^2}{n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \omega_t \quad (2)$$

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

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{V}, \quad (3)$$

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

$$\hat{\mathbf{h}} = \boldsymbol{\mathcal{X}}_n \sigma^2 \boldsymbol{\omega} + \mathbf{v} \quad (4)$$

⁵For our present purely exogenous nonstochastic \mathbf{X} , Eicker’s (1963) result suffices. White (1980) extends the result to include stochastic regressors as well.

⁶We use the notation “ $\mathbf{A} > \mathbf{0}$ ” to denote that the matrix \mathbf{A} is positive definite.

$$\begin{aligned}
\text{where } \hat{\mathbf{h}} &= \text{vech}(\hat{\mathbf{H}}) \sim K' \times 1, \\
\mathcal{X}_n &= [\text{vech}(\mathbf{x}_1\mathbf{x}'_1) \dots \text{vech}(\mathbf{x}_n\mathbf{x}'_n)] \sim K' \times n, \\
\boldsymbol{\omega} &= (\omega_1 \dots \omega_n)' \sim n \times 1, \text{ and } \boldsymbol{\omega} \subset \mathfrak{R}_+^n \\
\mathbf{v} &= \text{vech}(\mathbf{V}) \sim K' \times 1 \\
\text{and } K' &\equiv \frac{1}{2}K(K+1).
\end{aligned}$$

Some remarks to clarify the nature of the scedastic function regression model, as given in equation (4), are in order. Our first remark is on the error term, \mathbf{v} , and on the parameter vector $\boldsymbol{\omega}$. Distributional assumptions on \mathbf{v} are hard to come by, however. The only requirement on the nature of \mathbf{v} is as follows.

Remark 1. Elements of \mathbf{v} should make $\hat{\mathbf{H}} > 0$ through $\hat{\mathbf{h}}$ with the fact that $\boldsymbol{\omega} \subset \mathfrak{R}_+^n$.

We next turn our attention to the identifiability condition for the parameters in equation (4). It should be noted that the $\boldsymbol{\omega}$ parameters in equation (4) are identifiable only when the $\boldsymbol{\beta}$ vector in the original equation, *i.e.*, equation (1), is identifiable. Hence, the following should hold.

Remark 2. The $\boldsymbol{\omega}$ vector in equation (4) is identifiable if the inequalities $K' > n$ and $n > K$ simultaneously hold, where the regressor matrix in the equation, \mathcal{X}_n , has its dimension $K' \times n$ and $K' = K(K+1)/2$ as defined just after equation (4).

The above states that in a typical case like $n = 50$ and $K = 3$, $\boldsymbol{\omega}$ is unidentifiable since $K' = 6$. The chances are very slim for $\boldsymbol{\omega}$ to become identifiable, however. Take $n = 20$ and $K = 10$ then $K' = 55$ which makes $\boldsymbol{\omega}$ to be identifiable. We do admit that the conditions in the above remark, are not easily met in practice.

Remark 2 has clarified that $\boldsymbol{\omega}$ may become identifiable if the column dimension of \mathbf{X} increases. Then the “sample size” in equation (4), *i.e.*, K' , should also increase. We need to be careful employing this approach, however. What we need to do is to increase K' without altering \mathbf{e} , estimated regression residual vector from equation (1), that constitutes a part of dependent variable in equation (4). We may even call the situation “the curse of degrees of freedom.”

An approach that Amemiya (1983) proposes in a different context, may be adopted to free us from “the curse of degrees of freedom.” Amemiya proposed what he calls “partially generalized least squares” estimation method to improve upon efficiency over OLSE. He used a set of regressors that are orthogonal to the original regressors. For our model, the regressor matrix in the original regression equation now becomes

$$\mathbf{Z} = (\mathbf{X} \ \mathbf{W}) \sim n \times (K + K_W), \quad (5)$$

where we require that \mathbf{W} is orthogonal to all the variables in the original regression, *i.e.*, $\mathbf{W}'(\mathbf{y} \ \mathbf{X}) = \mathbf{0} \sim K_W \times (K+1)$ ⁷. It is easy to show that

⁷It is well known that the OLSE of $\boldsymbol{\beta}$ for the transformed regression, $\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} +$

\mathbf{b} and \mathbf{e} from equation (1) remain the same even after we regress \mathbf{y} on \mathbf{Z} instead of \mathbf{X} alone⁸. We are now ready to present our final remark of this section.

Remark 3. Use \mathbf{Z} defined in equation (5) in place of \mathbf{X} in equation (1). Then the identifiability condition of Remark 2, becomes

$$K'' > n \text{ and } n > K,$$

where $K'' \equiv (K + K_W)(K + K_W + 1)/2$.

We may increase K_W to the extent that K'' satisfies the above two conditions simultaneously. For instance, take the previous numerical case of $n = 50$ and $K = 3$. If we set $7 < K_W < 47$ then the conditions in the above remark are satisfied. 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 \mathbf{e} vector is unchanged.

We have now shown that using \mathbf{Z} instead of \mathbf{X} as the regressor matrix in equation (1), the $\boldsymbol{\omega}$ vector in equation (4) is always identified and estimable by the OLS provided that the parameters in the original regression equation, (1), are identified. The remaining question is whether the OLSE of $\boldsymbol{\omega}$ in (4) is efficient? Since the OLSE of $\boldsymbol{\omega}$ in equation (4) does not utilize any *a priori* condition on the parameter vector, *e.g.*, that $\omega_i > 0$ for $i = 1, \dots, n$, we doubt whether the nonparametric estimation method, *i.e.*, OLS of equation (4), is efficient? This motivates us to use a Bayesian procedure, which we describe in the next section.

3 Conditional Bayesian Inference

3.1 Bayesian setup

Notice that our whole argument is based upon the Eicker–White HCCM, equation (2). Hence, prior to proceeding to a Bayesian approach, we need to estimate equation (1) to obtain the regression residual vector \mathbf{e} . Note also that the scaling scalar parameter of the variance matrix $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{\Omega}$, *i.e.*, σ^2 , can be accurately estimated under the normalization assumption, $\text{tr}(\boldsymbol{\Omega}) = \sum_i \omega_i = n$.⁹ Therefore, we condition our Bayesian setup for the $\hat{\mathbf{h}}$ model, on estimated σ^2 and \mathbf{e} in the first stage. Hereafter, σ^2 is replaced by the estimate of it, $\hat{\sigma}^{2*} = \frac{1}{n} \mathbf{e}'\mathbf{e}$, from the first stage regression of \mathbf{y} on \mathbf{Z} .

Our Bayesian approach to estimate $\boldsymbol{\omega}$, thus, starts out by specifying the likelihood and prior pdf's (probability density functions) for equation (4), not (1). We shall call this approach “a conditional Bayesian” framework. In this approach the Bayesian setup is conditioned on the OLSE of $\boldsymbol{\beta}$ and

$\mathbf{Z}'\boldsymbol{\epsilon}$, is identical to the OLSE for the original equation. Incidentally, Amemiya (1983) showed that the GLS estimator for the transformed model is more efficient than the OLSE.

⁸In our *GAUSS* programming, we used the “NULL(·)” command to obtain $\mathbf{W} \sim n \times K_W$ that is orthogonal to \mathbf{X} . NULL command uses the QR decomposition to obtain the orthogonal complement of \mathbf{X} .

⁹See, for instance, Greene (2003, p.218) on this issue.

thereby \mathbf{e} in (1). Note that we need not begin with distributional specification of equation (1), since we assume a general set of conditions that guarantee a consistency for the HCCM.

We suppose a standard set of distributional assumptions for (4). First on \mathbf{v} :

$$\mathbf{v} \sim N(\mathbf{0}, \sigma_v^2 \mathbf{I}_{K''}), \quad (6)$$

where σ_v^2 is a scalar variance and $\mathbf{I}_{K''}$ denotes a K'' dimensional identity matrix. In Remark 1, we pointed out that there is no easily obtainable distributional assumption available to us, to form a likelihood function in the regression equation (4). Hence, the above normality assumption may seem at odds with Remark 1 at first sight. To elicit a suitable distributional assumption for \mathbf{v} , we have carried out some intensive experiments. As a result, we found that the above normality assumption on \mathbf{v} satisfies the requirement given in Remark 1.

Combining the normal based likelihood function of (6) with a joint informative prior pdf for σ_v and for the elements of $\boldsymbol{\omega}$,

$$P(\hat{\sigma}^{2*} \boldsymbol{\omega}, \sigma_v^2) \propto \sigma_v^{-1} \text{ for any element of } \boldsymbol{\omega} \in [\underline{\omega}, \bar{\omega}], \quad (7)$$

we obtain the posterior pdf¹⁰,

$$P(\hat{\sigma}^{2*} \boldsymbol{\omega}, \sigma_v^2 | \hat{\mathbf{h}}, \mathcal{X}_n) = P(\hat{\sigma}^{2*} \boldsymbol{\omega} | \sigma_v^2, \hat{\mathbf{h}}, \mathcal{X}_n) \times P(\sigma_v^2 | \hat{\mathbf{h}}, \mathcal{X}_n). \quad (8)$$

This shows that the joint posterior pdf is a product of the two conditional posterior distributions:

$$\hat{\sigma}^{2*} \boldsymbol{\omega} | \sigma_v^2, \hat{\mathbf{h}}, \mathcal{X}_n \sim N(\hat{\sigma}^{2*} \hat{\boldsymbol{\omega}}, \sigma_v^2 (\boldsymbol{\mathcal{X}}_n' \boldsymbol{\mathcal{X}}_n)^{-1}) \mathbf{1}_{\boldsymbol{\omega} \in [\underline{\omega}, \bar{\omega}]} \quad (9)$$

and

$$\sigma_v^2 | \hat{\mathbf{h}}, \mathcal{X}_n \sim IG[\hat{\mathbf{h}}' \mathbf{M} \boldsymbol{\mathcal{X}}_n \hat{\mathbf{h}}, K'' - n], \quad (10)$$

where $\mathbf{1}_{\boldsymbol{\omega} \in [\underline{\omega}, \bar{\omega}]}$ in (9) is an indicator function such that it takes on the value one if $\boldsymbol{\omega} \in [\underline{\omega}, \bar{\omega}]$, *i.e.*, the n elements of $\boldsymbol{\omega}$ all fall in the range $[\underline{\omega}, \bar{\omega}]$ and zero otherwise, and the lower and upper limits, $\underline{\omega}$ and $\bar{\omega}$, are to be specified later. Notice that the entire right hand side of (9) is an n dimensional truncated multivariate normal distribution. This motivates us to use an MCMC method to simulate the posterior $\boldsymbol{\omega}$ and σ_v^2 . $\hat{\boldsymbol{\omega}}$ in equation (9) is the OLSE of equation (4), where \mathbf{Z} is used instead of \mathbf{X} and $\frac{1}{n} \mathbf{e}' \mathbf{e}$ is inserted to $\hat{\sigma}^{2*}$. $IG[\cdot]$ denotes an inverted gamma distribution.

Given equation (9) and (10), what we need to do is to simulate $\hat{\sigma}^{2*} \boldsymbol{\omega}$ and σ_v^2 . If we divide $\hat{\sigma}^{2*} \boldsymbol{\omega}$ by $\hat{\sigma}^{2*} = \frac{1}{n} \mathbf{e}' \mathbf{e}$, then we obtain sets of $(\boldsymbol{\omega}' \sigma_v^2)' \subset \mathfrak{R}_+^{n+1}$. For each of $(n+1)$ elements, we may depict a posterior pdf, posterior mean and variance or standard deviation.

¹⁰See, *e.g.*, Koop (2003, section 4.3) for a discussion of inequality constraints in a regression framework.

3.2 Bayesian experiments and MCMC procedure

In this subsection, we shall generate a data set to make sure that our Bayesian procedure works. In doing so, we may illustrate the details of our MCMC (Markov Chain Monte Carlo) procedure. Using a prespecified set of parameter values on β , σ^2 and ω , we generate a data set on \mathbf{y} and \mathbf{X} . This is used to generate a \mathbf{W} matrix to form $\mathbf{Z} = (\mathbf{X} \mathbf{W})$. By regressing \mathbf{y} on \mathbf{Z} , we obtain one set of $\mathbf{e} \sim n \times 1$, and hence a $\hat{\mathbf{h}} \sim K'' \times 1$ vector. Our Bayesian inference starts out from here.

We set our sample size to be $n = 50$. It now seems reasonable to set the bounds for the prior of ω , *i.e.*, $[\underline{\omega}, \bar{\omega}]$, to be $[0.025, 12.5]$, and we shall explain the reasons for this in below. First, in view of our scaling assumption, $\sum_i \omega_i = n = 50$ which gives an average of $\sum_i \omega_i/n = 0.5$, the set of bounds $\underline{\omega} = 0.025$ and $\bar{\omega} = 12.5$ covers a large enough interval. Next, 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 $\sum_{i=5}^{50} \omega_i = 4$. It can be seen that

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

The above implies that even in this extreme case, our lower bound, $\underline{\omega}$, is smaller than the average of the latter forty six ω_i 's, while our upper bound, $\bar{\omega}$, exceeds the average of the first four ω_i 's. In addition, we emphasize that this data dependent prior on the bounds of ω_i for $i = 1, \dots, n$ are relatively easy to set, since the only information needed to form it is n . In this paragraph, we gave a rule of thumb for setting the bounds for ω_i 's. Our experiments confirm that draws are relatively insensitive to the bounds.

We used $\beta = (1, 3, -2)'$ values for β . The \mathbf{X} matrix is specified to be $\mathbf{X} = (\mathbf{1}_n, \mathbf{x}_2, \mathbf{x}_3) \sim n \times 3$, where $\mathbf{1}_n$ is an $n \times 1$ vector of all one's, $\mathbf{x}_2 \sim N(3, I_n)$, and $\mathbf{x}_3 \sim N(6, 2^2 I_n)$. There are 50 elements in ω and thus we do not care to present each one of them here. Instead, these values will be given in a chart that compares them to the Bayesian posterior mean values, later.

Given the OLS estimated regression residuals, \mathbf{e} and the \mathcal{X}_n matrix that uses \mathbf{Z} , we are now ready to estimate regression (4). Equation (4) could be estimated by the OLS, again as in the first stage. We speculate that the Bayesian posterior analysis should do better than the OLS since the former combines *a priori* assumption on ω , whereas the latter does not, *i.e.*, a nonparametric estimation method.

Let us now turn to discuss our MCMC procedure. σ_v^2 can be easily simulated from the Inverted Gamma density, $IG[\hat{\mathbf{h}}' \mathbf{M} \mathcal{X}_n \hat{\mathbf{h}}, K'' - n]$. For the 50 values of ω , we used a MH (Metropolis–Hastings) method, random walk chain¹¹,

$$\omega^{(s)} \sim N(\omega^{(s-1)}, \text{diag}(\hat{\omega} \odot \hat{\omega})) \mathbf{1}_{\omega \in [0.025, 12.5]}, \quad (11)$$

¹¹It would seem more natural to use a normal based proposal density that uses a sample-based variance matrix: $(\mathcal{X}_n' \mathcal{X}_n)^{-1}$. This avenue of approach does not work well here.

where $\omega^{(s)}$ is the s th simulated value of ω , $\hat{\omega}$ denotes an OLS estimated vector value, and “ \odot ” denotes an elementwise Hadamard product. 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 the Markov Chain procedure. After the burn-in’s, we simulated 20,000, which makes the total number of simulated set is 30,000.

We now have a set of $20,000/5 = 4,000$ simulated values for each of σ_v^2 and ω_i ’s ($i = 1, \dots, 50$) available. Since giving 51 posterior pdf’s would not contribute to the understanding of our procedure, we decided to present posterior means of the fifty $\hat{\sigma}^{2*}\omega_i$ ’s, and this is given in Figure 1. In the same chart, fifty assumed values are given using a solid line. We see that there are some discrepancies between the true (or given) value and the posterior mean value, however, mostly they are close. For instance, the second one, $\sigma^2\omega_2$ has both the given value and the posterior mean close to 2.5. We have selected this sample, *i.e.*, $\sigma^2\omega_2$, to depict the posterior density, and this is given in Figure 2. It seems to be a typical density for a variance that is unimodal in a positive region.

In our experiment, we now confirmed that the posterior mean values are fairly close to the true values. The next thing that we should examine is the posterior standard deviations. The 50 posterior standard deviations are given in Figure 3. They do fluctuate a lot. 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, 1.6, which is not small at all. But if we compare Figure 3 to the previous Figure 1, we notice a tendency that posterior means and standard errors, move in the same direction¹². We next compared the posterior mean values to the OLS estimated values, $\hat{\omega}$. We do not present this result here, since these two are very close. This is comforting to know since this is another evidence that our Bayesian method is more efficient, because of the appropriate prior information has been used.

3.3 One-factor model of Japanese pharmaceutical stocks

In this section, we use a panel data set to show that our procedure works well in practice. Panel data sets should be most suitable vehicle with which to experiment our procedure. We used data on the daily stock prices of fifty Japanese pharmaceutical/ biomedical venture capital companies. The data period is from May 6 2005 to May 1 2006. The sample size is 285.

Let us begin presenting our model with a multifactor multivariate return generating equation:

$$\mathbf{R} = (\iota_T \mathbf{F}) \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{B} \end{pmatrix} + \boldsymbol{\epsilon}, \quad (12)$$

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

¹²We should compute a sample correlation between these two series.

Figure 1: posterior means of the 50 ω_i 's

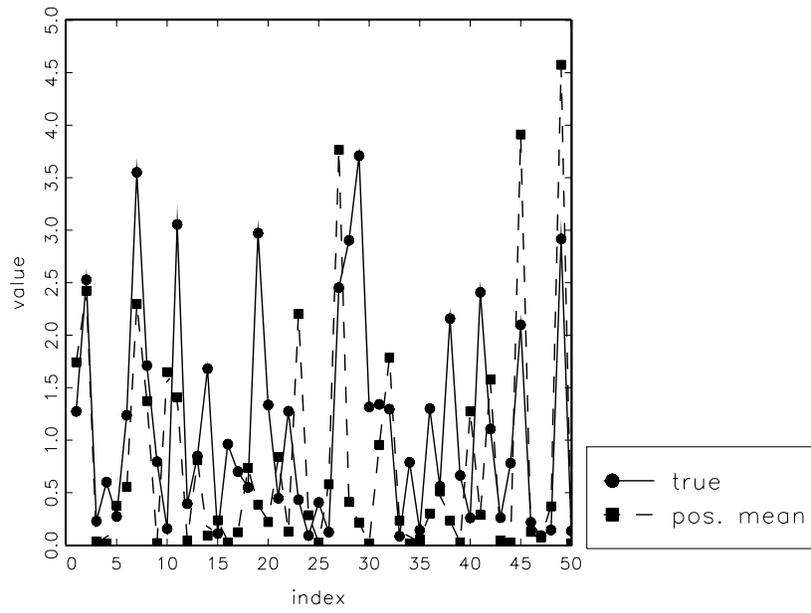


Figure 2: posterior density of $\sigma^2\omega_2$

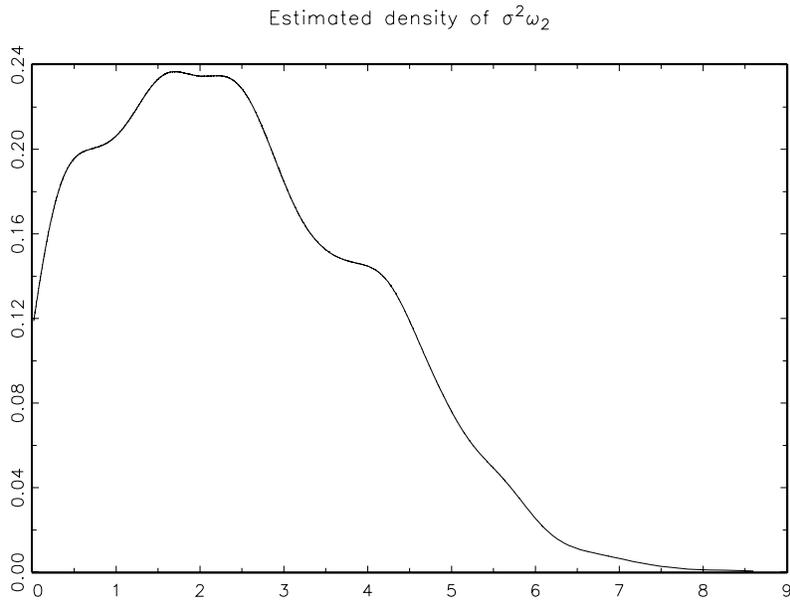
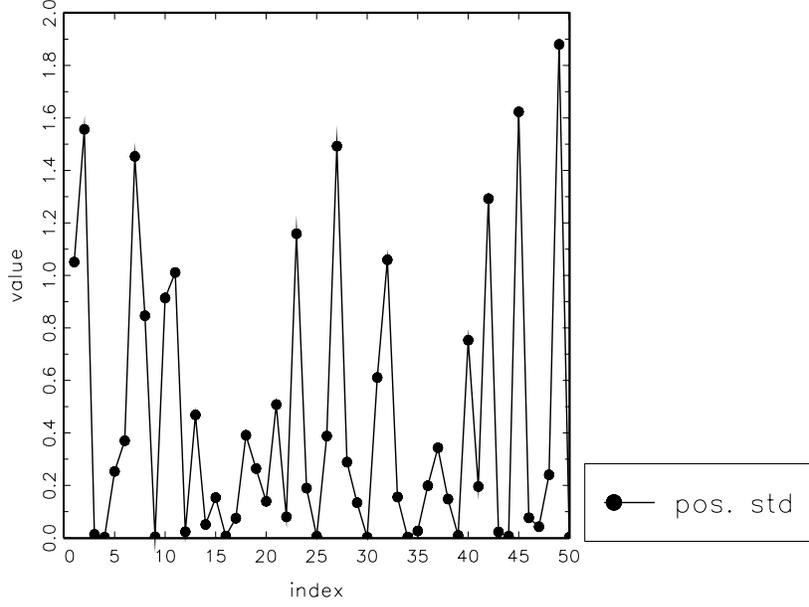


Figure 3: posterior standard deviations of the 50 ω_i 's



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:

$$\mathbf{R} = (\iota_T \mathbf{f}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \epsilon, \quad (13)$$

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

Equation (13) is just a set of N time series regressions. We obtain an OLSE of β' , $\hat{\beta}' \sim N \times 1$ from equation (13). Define sample mean of \mathbf{R} to be an N dimensional vector $\bar{\mathbf{R}}$, then we obtain

$$\bar{\mathbf{R}} = \hat{\beta}' \lambda + \mathbf{u}, \quad (14)$$

where $\bar{\mathbf{R}} = \frac{1}{T} \mathbf{R}' \iota_T = (\bar{R}_1 \cdots \bar{R}_N)' \sim N \times 1$ vector of average excess returns,
 $\lambda \sim \text{scalar}$ is a risk premium associated with one factor \mathbf{f} ,
 $\mathbf{u} \sim N \times 1$ is a vector of pricing errors.

This, *i.e.*, equation (14), is the one factor type CAPM model given in Cochrane (2001, p.235), among others. Since we used a set of notations that are common in empirical finance, we thought we need to clarify the notational correspondences between this and previous subsections. They are given in below.

$\bar{\mathbf{R}} \sim N \times 1$	(corresponds to \mathbf{y} in (1)),
$\hat{\beta}' \sim N \times 1$	(corresponds to \mathbf{X} in (1)),
$\mathbf{u} \sim N \times 1$	(corresponds to ϵ in (1)),
$\lambda \sim \text{scalar}$	(corresponds to β in (1)),
$F = 1$ is the number of factors	(corresponds to K in (1)),
N the number of stocks	(corresponds to n in (1)).

Our purpose in this subsection, is to investigate whether $\text{Var}(u_i) = \sigma^2\omega_i$ for $i = 1, \dots, n$ can be successfully estimated using our Bayesian method? To this effect, we have simulated posterior densities following the methodology outlined in the previous subsection. We used exactly the same priors and likelihood as in subsection 3.2. This includes the bounds for ω_i to be $[0.025, 12.5]$ too. The details of the MCMC, such as the number of replications and burn-in's, are also set exactly the same as in subsection 3.2.

The results are given in Figures 4 and 5. Posterior means are presented in Figure 4. Since we thought that the stocks traded in TSE-1 (the first section of the Tokyo Stock Exchange) and others (this includes venture capital type stocks) may have different $\text{Var}(u_i)$, *i.e.*, volatility, we decided to mark the simulated posterior mean by a certain positive value. Hence, in Figure 4, $\sigma^2\omega_3$ is rather small (around 0.001) and it is identified as “TSE-1” type return since “ts1” line sticks to the bottom. On the other hand $\sigma^2\omega_2$ is relatively large (around 0.017), which is identified as the non-TSE-1 type return by “ts1” line taking the value 0.0045. This value, *i.e.*, 0.0045, is the average of all return volatility.

Figure 4: posterior means of $\sigma^2\omega_i$ for the 50 pharmaceutical/biomedical firms

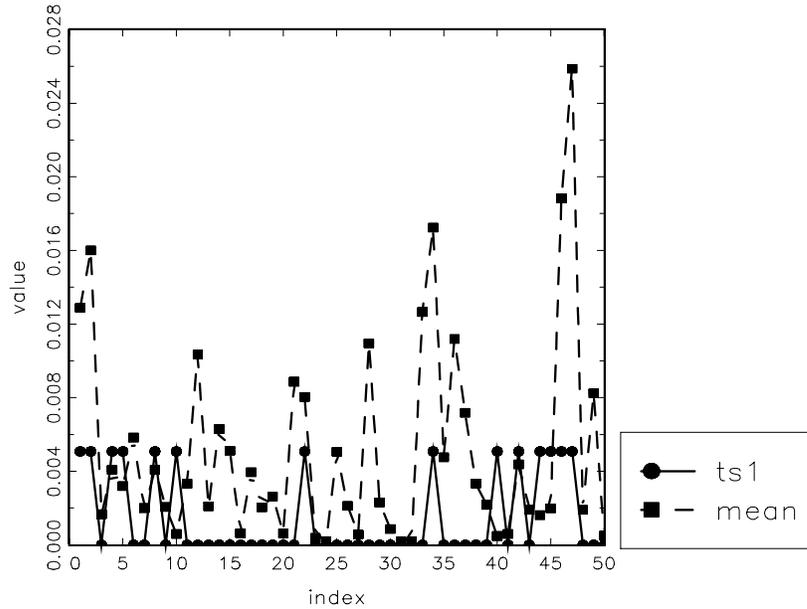
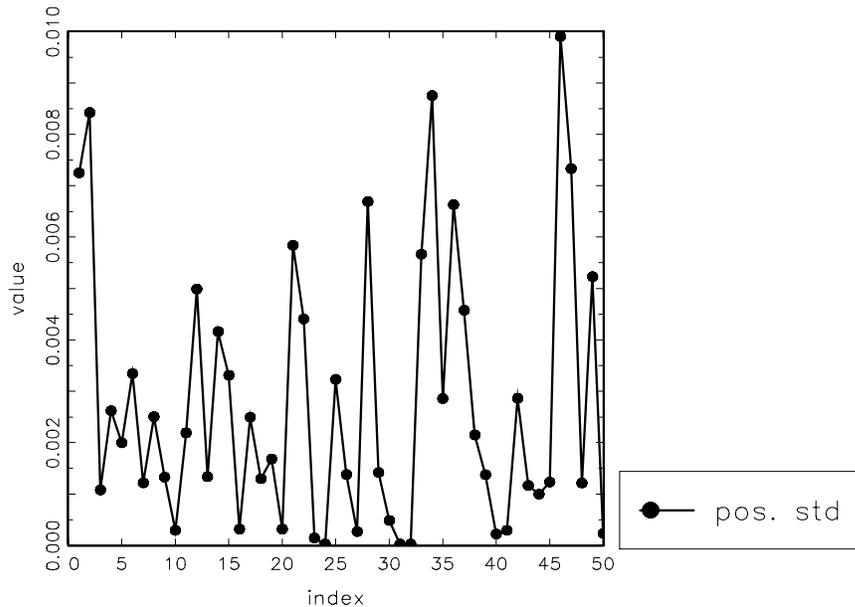


Figure 4 demonstrates that our Bayesian method works. On the average, non-TSE-1 type returns have higher volatility than the TSE-1 type returns. This fact is also demonstrated in the following descriptive statistics.

Figure 5: posterior standard deviations of $\sigma^2\omega_i$ for the 50 pharmaceutical/biomedical firms



Type of Stocks	average of posterior means	average of posterior standard deviations
TSE-1(36)	0.003596	0.002191
non TSE-1 (14)	0.008890	0.004463

n.b.: Inside (·) is the number of stocks.

From the table above, we see that the non-TSE-1 type stocks have, on the average, twice as high estimated volatility compared to the TSE-1 type stocks. This is quite reasonable in view of the well known anomaly in the Japanese stock market: smaller company's stocks tend to have higher volatility than the larger company's.¹³

4 Concluding Remarks

In this paper, we developed a conditional Bayesian approach to estimate diagonal elements of the regression error term variance-covariance matrix, ω . Our method may be divided into two parts: (i) a nonparametrical OLS estimation of the Eicker-White HCCM, and (ii) a conditional Bayesian estimation of ω . We have supplied a numerical example, and also an empirical

¹³See, *e.g.*, Chang and Dong (2006), among others.

example to show that our method works reasonably. Finally, as demonstrated in our empirical research, obtaining \mathbf{W} such that $\mathbf{W}'(\mathbf{y} \ \mathbf{X}) = \mathbf{0}$ is no problem since \mathbf{W} is not found as an empirical data but rather computer generated data.

Let us suggest further task of this research. It would be nice if we could go fully Bayesian from of the current conditional Bayesian. That is to start assuming priors and likelihood for the first stage regression. Theoretically, this is more complete compared to the present setup, however, this also implies that we abandon the HCCM. If we dispense with the HCCM, there is very little information that can be extracted for scedastic function from data alone, *i.e.*, (\mathbf{y}, \mathbf{X}) .

References

- [1] Amemiya, T., 1983, "Partially generalized least squares and two-stage least squares estimators," *Journal of Econometrics*, 23, 275–283.
- [2] 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.
- [3] 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.
- [4] Cochrane, J.H., 2001, *Asset Pricing*, Princeton University Press.
- [5] 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.
- [6] 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, 641–649, Clarendon Press.
- [7] Godfrey, L.G., 2006, "Tests for regression models with heteroskedasticity of unknown form," *Computational Statistics & Data Analysis*, 50, 2715–2733.
- [8] Greene, W.H., 2003, *Econometric Analysis*, 5th ed., Prentice-Hall.
- [9] Koop, G., 2003, *Bayesian Econometrics*, John Wiley & Sons.
- [10] Long, J.S. and L.H. Ervin, 2000, "Using heteroscedasticity consistent standard errors in the linear regression model," *American Statistician*, 54, 3, 217–224.
- [11] Lütkepohl, H., 1996, *Handbook of Matrices*, John Wiley & Sons.
- [12] Robinson, P.M., 1987, "Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form," *Econometrica*, 55, 4, 875–891.

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

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

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

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

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

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

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

proof : Omitted.

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

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

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

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

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