1990

Estimation of Systematic Risk Using Bayesian Analysis with Hierarchical and Non-Normal Priors

Anil K. Bera
José AF Machado

Citation of this paper:
RESEARCH REPORT 9008

ESTIMATION OF SYSTEMATIC RISK USING BAYESIAN ANALYSIS WITH HIERARCHICAL AND NON-NORMAL PRIORS

by

Anil K. Bera
and
José A.F. Machado

June 1990

Department of Economics
University of Western Ontario
London, Ontario, Canada
N6A 5C2
ESTIMATION OF SYSTEMATIC RISK USING BAYESIAN ANALYSIS WITH HIERARCHICAL AND NON-NORMAL PRIORS*

Anil K. Bera  
Department of Economics, University of Illinois  
and  
Department of Economics, University of Western Ontario  
and  
José A. F. Machado  
Department of Economics, University of Illinois  
and  
Faculdade de Economia, Universidade Nova de Lisboa

*We would like to thank Bill Griffiths for numerous suggestions which improved the paper. Thanks are also due to an anonymous referee, Paul Newbold, Sidhartha Chib and participants of a Conference in Bayesian Inference in Econometrics for helpful comments on an earlier version of this paper and Pin Ng for drawing the figures. Responsibility for any errors and omissions is naturally solely ours. Financial support from the Bureau of Economic and Business Research and the Research Board of the University of Illinois is gratefully acknowledged.
ABSTRACT

Estimation of systematic risk is one of the most important aspects of investment analysis, and has attracted the attention of many researchers. In spite of substantial contributions in the recent past, there still remains room for improvement in the methodologies currently available for forecasting systematic risk. This paper is concerned with some improved methods of estimating systematic risk for individual securities. We use Bayesian analysis with hierarchical and non-normal priors.
1. **Introduction**

The central model in most of the research pertaining to systematic risk has been the single index model

\[ R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it} \quad i = 1, 2, \ldots, N \]

\[ t = 1, 2, \ldots, T \quad (1) \]

where \( R_{it} \) and \( R_{mt} \) are, respectively, the random return on security \( i \) and the corresponding random market return in period \( t \). \( \alpha_i \) and \( \beta_i \) are the regression parameters appropriate to security \( i \) and \( \epsilon_{it} \) is the random disturbance term which is distributed as normal with mean zero and variance \( \sigma_i^2 \). The parameter \( \beta_i \), called beta, measures the systematic risk of the security \( i \) and is defined as \( \text{Cov}(R_i, R_m)/\text{Var}(R_m) \).

Estimation of this systematic risk is one of the most important aspects of investment analysis and has attracted the attention of many researchers. Betas are used by the investors to evaluate the relative risk of different portfolios. In the future market context, betas of different stock portfolios are needed to calculate the number of contracts to be bought or sold. In spite of substantial contribution in the recent past, there still remains room for improvement in the methodologies currently available to forecast betas.

Blume (1971) observed that over time betas appear to take less extreme values and exhibit a tendency towards the market risk. This would mean that the historical betas based on ordinary least squares (OLS) estimation would be poor estimators of the future betas. Therefore, it is necessary to adjust the OLS estimators of \( \beta_i \). Ohlson and Rosenberg (1982) tried to take account for the variation of beta over
time by treating it as a stochastic parameter, and provided confidence regions using a mixture of classical and Bayesian perspective. Vasicek (1973) suggested a Bayesian adjustment technique using a normal prior for $\beta_i$. As it will be clear from the subsequent discussion, Vasicek's procedure has some drawbacks. It utilizes the information from the other stocks only through the cross-sectional mean and variance. In the New York Stock Exchange more than 2,000 stocks are traded; improved estimates of one stock might be obtained by combining the data from the other stocks as far as possible. We propose to do this utilizing Lindley and Smith's (1972) hyperparameter model and the concept of exchangeable priors. Under this framework, parameters of our linear model (1), themselves will have a general linear structure in terms of other quantities which are called hyperparameters. And exchangeability means that the joint distribution of $\beta_i$'s is unaltered by any permutation of the suffixes. This assumption is weaker than the traditional independent and identically distributed (IID) set up.

Lindley and Smith's linear hierarchical model has been used in many econometric applications, see e.g., Trivedi (1980), Haitovsky (1986), Ilmakunnas (1986), and Kadiyala and Oberhelman (1986). In Section 2, we set up the model in a convenient form and carry out the Bayesian analysis using the hierarchical model.

Recently, Bera and Kannan (1986) studied extensively the empirical distribution of betas. They considered the time period from July 1948 through June 1983, and divided that period into seven non-overlapping estimation periods of 60 months each. They found that the empirical distributions of betas were highly positively skewed and often
platykurtic. However, with a square-root transformation the values of skewness and kurtosis changed in such a way that using the Jarque and Bera (1987) test statistic the normality hypothesis could be accepted in four out of seven periods. Also the values of the test statistic in the remaining three periods were not very high. Therefore, it appears that beta has a root-normal distribution, i.e., the square-root of the variable is normal. The finding casts some doubts on the validity of Vasicek's selection of normal priors. Therefore, our second aim is to do a Bayesian analysis assuming that $\sqrt{\beta_1}$ is normally distributed, or in other words, $\beta_1$ has a noncentral $\chi^2$ distribution with one degree of freedom. In Section 3, we generalize our results of Section 2 by considering a hierarchical model with non-normal priors; while in Section 4, we present an empirical Bayesian analysis, similar to Vasicek's and based on our non-central chi-square prior distribution. In the last section of the paper, some concluding remarks are offered.

After the publication of Vasicek's paper in 1973, to our knowledge, there is no work along Bayesian lines which attempts to improve upon it. Also in the statistics literature, most Bayesian regression analysis are based on normal priors primarily because of their simplicity. Since here we have some empirical evidence on the distribution of betas, it is appropriate that we utilize that information in the analysis. We hope this will lead to improved estimation of betas.

2. **Analysis with Hierarchical Priors**

Since we are interested only in the $\beta_1$ parameters, it is convenient to work with the model in deviation form
\[ y_{it} = \beta_i x_{it} + u_{it} \]  \hspace{1cm} (2)

where \( y_{it} = R_{it} - \bar{R}_i \), \( x_{it} = R_{mt} - \bar{R}_m \) and \( u_{it} \) is the new disturbance term.

This deviation form model can be obtained in the following way.

The sample information for the security \( i \) is given by

\[
f(R_i | \alpha_i, \beta_i, \sigma_i) = \left( 2\pi \right)^{-T/2} \left( \sigma_i^2 \right)^{-T/2} e^{-\frac{1}{2 \sigma_i^2} \sum_{t=1}^{T} (R_{it} - \alpha_i - \beta_i R_{mt})^2}
\]

where \( R_i = (R_{i1}, R_{i2}, \ldots, R_{iT})' \). Assuming non-informative prior for \( \alpha_i \) and integrating \( \alpha_i \) out we obtain

\[
f(y_i | \beta_i, \sigma_i) = \left( \frac{2\pi}{T} \right)^{1/2} \sigma_i^{-1} \left( \frac{1}{2\pi} \right)^{-T/2} e^{-\frac{1}{2 \sigma_i^2} \sum_{t=1}^{T} (y_{it} - \beta_i x_{it})^2}
\]

where \( y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})' \). Therefore, given the prior on \( \alpha_i \), the sample information can be written as

\[
y_i | \beta_i, \sigma_i \sim N(x_i^* \beta_i, (T/2\pi)I_n)
\]

where \( y_i^* = y_{it} (2\pi \sigma_i / T)^{1/2} \) and \( x_i^* = x_{it} (2\pi \sigma_i / T)^{1/2} \). Furthermore, since

\[
\frac{x_i y_i}{T/2\pi} = \frac{x_i y_i^*}{\sigma_i^2} \quad \text{and} \quad \frac{x_i^* y_i}{T/2\pi} = \frac{x_i y_i}{\sigma_i^2}
\]

for our results below (namely (7) and (8)), it is equivalent to work with \( x_i \) and \( y_i \) instead of \( x_i^* \) and \( y_i^* \).
Under the classical framework, the maximum likelihood or the OLS estimator of \( \beta_i \) in model (2) is given by

\[
\hat{\beta}_i = \frac{\sum_{t=1}^{T} x_{it} y_{it}}{\sum_{t=1}^{T} x_{it}^2}
\]

\( i = 1, 2, \ldots, N \)

with \( \text{Var}(\hat{\beta}_i) = \sigma_i^2 / \sum_{t=1}^{T} x_{it}^2 \equiv S_i^2 \). Vasicek (1973, pp. 1235–1236) suggested a Bayesian approach with normal prior for \( \beta_i \)

\[
\beta_i \sim N(\bar{\beta}, \phi^2)
\]

and showed that for large \( T \), the posterior distribution of \( \beta_i \) conditional on \( \bar{\beta} \) and \( \phi^2 \) is approximately normal with mean

\[
\frac{(\bar{\beta}/\phi^2) + (\hat{\beta}_i/S_i^2)}{(1/\phi^2) + (1/S_i^2)}
\]

and variance \((1/\phi^2) + (1/S_i^2)^{-1}\). Vasicek suggested to use the mean and variance of the cross-sectional betas in place of \( \bar{\beta} \) and \( \phi^2 \) respectively and

\[
\hat{\sigma}_i^2 = \frac{\sum_{t=1}^{T} (y_{it} - \hat{\beta}_i x_{it})^2}{(T-2)}
\]

As we should expect, had Vasicek assumed complete ignorance about \( \bar{\beta} \) instead of \( \bar{\beta} \) being known, that is \( \phi^2 = \infty \), the posterior mean would be the OLS estimator.

Empirical results in Bera and Kannan (1986, Tables VII and VIII) show that forecasts based on the Vasicek's adjusted betas (posterior mean) are superior to the OLS (unadjusted) betas. This indicates that we can improve prediction performance for a security by pooling information from other securities.
Let us now cast the model in Lindley and Smith's hierarchical framework. Assuming a prior \( \pi(\sigma) = 1/\sigma \), from our earlier expression of \( f(y_i|\beta_i, \sigma_i) \), we can write

\[
y_i, \sigma_i|\beta_i \sim N(x_i \beta_i, \sigma_i I_T) \quad i = 1, 2, \ldots, N.
\]  

(5)

Given this formulation for the sample information our posterior distributions will be conditional on \( \sigma_i \). Next we assume exchangeability among the \( \beta_i \), specifically

\[
\beta_i|\xi \sim N(\xi, \tau^2)
\]

(6)

with a second stage non-informative prior for \( \xi \). Due to the randomness of \( \xi \), this is a weaker assumption than the IID assumption in (3).

To see it clearly, note that the joint prior distribution of \( \beta_1, \beta_2, \ldots, \beta_N \) is given by

\[
\pi(\beta_1, \beta_2, \ldots, \beta_N) = \int (\Pi_{i=1}^N \pi(\beta_i|\xi)) f(\xi) d\xi
\]

where \( f(\xi) \) is the probability density function of \( \xi \). Therefore, \( \pi(\beta_1, \beta_2, \ldots, \beta_N) \) is a mixture of IID distributions conditional on \( \xi \), but unconditionally the joint distribution does not satisfy the IID assumption. The above specification is a simple special case of Lindley and Smith's (1972, p. 6) general hierarchical model

\[
y|\theta_1 \sim N(A_1 \theta_1, C_1)
\]

\[
\theta_1|\theta_2 \sim N(A_2 \theta_2, C_2)
\]

\[
\theta_2|\theta_3 \sim N(A_3 \theta_3, C_3)
\]
with $C_3^{-1} = 0$. For this model Bayesian inference can be drawn from the posterior for $\theta_1$ given $\{A_1\}, \{C_1\}$, and $y$ which is given by $N(Dd,D)$ where

$$D^{-1} = A_1'C_1^{-1}A_1 + C_2^{-1} - C_2^{-1}A_2(A_2'C_2^{-1}A_2)^{-1}A_2'C_2^{-1}$$

and

$$d = A_1'C_1^{-1}y.$$

Identifying the appropriate values of $A_1$, $C_1$ and $\theta_1$, for our specification (5) and (6), we obtain the following posterior distribution for $\theta_1 = (\beta_1, \beta_2, ..., \beta_n)'$ as $N(Dd,D)$ where

$$D^{-1} = \text{diag}\left(\frac{x_1'x_1}{\sigma_1^2} + \frac{1}{\tau^2}, ..., \frac{x_N'x_N}{\sigma_N^2} + \frac{1}{\tau^2}\right) - \frac{J_N}{N\tau^2}$$

(7)

where $J_N$ is an $N\times N$ matrix whose all elements are one, and

$$d' = \left(\frac{x_1'y_1}{\sigma_1^2}, ..., \frac{x_N'y_N}{\sigma_N^2}\right).$$

(8)

Using the above expression for the $i$-th security, the estimate of the systematic risk under a quadratic loss function can be expressed as

$$\beta_i^* = \left(\frac{x_i'x_i}{\sigma_i^2} + \frac{1}{\tau^2}\right)^{-1} \frac{x_i'x_i}{\sigma_i^2} \beta_i + \left(\frac{x_i'x_i}{\sigma_i^2} + \frac{1}{\tau^2}\right)^{-1} \frac{1}{\tau^2} \sum_{j=1}^{N} w_j \beta_j$$

where

$$w_j = \left[\sum_{i=1}^{N} \frac{x_i'x_i}{\tau^2 x_i'x_i + \sigma_i^2}\right]^{-1} \frac{x_j'x_j}{\tau^2 x_j'x_j + \sigma_j^2}.$$

(9)
A quick comparison of (4) and (9) reveals that both estimates are linear combinations of OLS estimator and the mean of cross-section beta, but for the hierarchical estimate a weighted average of the cross-section betas are used instead of simple average as in the Vasicek case. \( \beta^*_i \) can also be expressed as

\[
\beta^*_i = \frac{\sigma^2_{i_1}}{\sigma^2_{i_1}} \left( \frac{\hat{\beta}^*_i}{\sigma^2_{i_1}} + \frac{\bar{\beta}^*}{\sigma^2} \right)
\]

where \( \bar{\beta}^* = \frac{1}{N} \sum_{j=1}^{N} \beta^*_j \). This formula, which is directly comparable with (4), reveals how the information from other securities is used in estimating the systematic risk for \( i \)-th security. Unlike in (4), the information conveyed by other stocks which is reflected in \( \bar{\beta}^* \) is incorporated in a self-fulfilling way, since the cross-sectional average beta is consistent with the estimates for individual securities. This has a rational expectations interpretation in the sense that the market information embodied in \( \bar{\beta}^* \) is compatible with all individual \( \beta^*_i \).

To compare the above estimate with Vasicek's one, let us put his specification in Lindley and Smith's framework as

\[
y_i, \sigma_i | \beta_i \sim N(x_i \beta_i, \sigma^2_{i_1}),
\]

\[
\beta_i | \bar{\beta} \sim N(\bar{\beta}, \phi^2).
\]

(10)

Here the first stage prior is "completely specified" by the cross-section data. The counterparts (7) and (9) are respectively,
\[ D_V^{-1} = \text{diag}\left(\frac{x_1'x_1}{\sigma_1^2} + \frac{1}{\phi^2}, \ldots, \frac{x_N'x_N}{\sigma_N^2} + \frac{1}{\phi^2}\right) \] (11)

and

\[ \beta_{1V}^* = \left(\frac{x_1'x_1}{\sigma_1^2} + \frac{1}{\phi^2}\right)^{-1} \frac{x_1'\hat{\beta}_1}{\sigma_1^2} + \left(\frac{x_i'x_i}{\sigma_i^2} + \frac{1}{\phi^2}\right)^{-1} \frac{1}{\phi^2} \left(\sum_{j=1}^{N} \frac{\hat{\beta}_j}{N}\right). \] (12)

Comparing (9) and (12), we note that to find the average systematic risk in (12), a simple average is used whereas under the Lindley and Smith framework, we use a weighted average. The latter is more reasonable since the precision in estimating the systematic risks of different securities are different from each other.

It is also interesting to note that

\[ D_V^{-1} - D^{-1} = \text{diag}\left(\frac{1}{\phi^2} - \frac{1}{\tau^2}, \ldots, \frac{1}{\phi^2} - \frac{1}{\tau^2}\right) + \frac{J_N}{N\tau^2} \]

and as such we cannot say much about this matrix. However, when we put \( \phi^2 = \tau^2 \), i.e., the first stage prior variances are the same then \( D_V^{-1} - D^{-1} \) is positive semi-definite. In other words, the Vasicek's estimator has higher precision. This result is not at all surprising if we compare the prior distributions. Under the Vasicek prior at the second stage \( V(\beta) = 0 \) whereas for the hyperparameter model we assume a second stage non-informative prior, i.e., \( V(\xi)^{-1} = 0 \). Therefore, the basic difference between the two approaches are two alternative ways of handling the unknown prior mean \( \beta \). In Vasicek's approach, the Bayes' estimator is derived assuming \( \beta \) is known and at the end it is replaced by the simple average. On the other hand, the Lindley and
Smith method is more fully Bayesian in the sense that it uses non-informative prior for $\xi$ (or $\bar{\theta}$).

Finally, we should note that the estimator in (9) depends on $\sigma_i^2$ and $\tau^2$. In practice these parameters will have to be estimated. Assuming noninformative scale invariant priors, that is, $\pi(\sigma_i) \propto \sigma_i^{-1}$ and $\pi(\tau) \propto \tau^{-1}$, the modal values from the joint posterior distribution of $(\beta_1, \beta_2, \ldots, \beta_N, \sigma_1, \sigma_2, \ldots, \sigma_N, \tau)$ are

$$
\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} (y_{it} - x_{it} \hat{\beta}_i)^2
$$

and

$$
\hat{\tau}^2 = \frac{1}{N} \sum_{i=1}^{N} (\hat{\beta}_i - \bar{\beta})^2
$$

where $\beta_i^*$ and $\bar{\beta}^*$ are as defined earlier. These equations, together with (9), can be solved iteratively. Other priors for $\sigma_i^2$ and $\tau^2$, such as inverse $\chi^2$ could also be handled easily [see, Lindley and Smith (1972, p. 13)].

3. Analysis of Hierarchical Model with Non-normal Prior

Vasicek used a normal prior for the cross-sectional distribution of the beta coefficients. As mentioned earlier, the cross-sectional betas are not normally distributed, and recent work by Bera and Kannan (1986) indicates that their distribution tends to normal after a square-root transformation. It is therefore natural to explore the consequences of assuming a root-normal prior for beta, i.e.,

$$
\sqrt{\beta} \sim N(\xi, \tau^2).
$$
Then $\beta/\tau^2$ will be distributed as a noncentral $\chi^2$ with one degree of freedom and noncentrality parameter $(\xi/\tau)^2$, denoted by $\chi_1^2(\xi^2/\tau^2)$. The p.d.f. of $\beta$ can be written as

$$
\pi(\beta|\xi, \tau^2) = (2\pi)^{-1/2}(\beta \tau^2)^{-1/2} e^{-\frac{1}{2\tau^2}(\beta + \xi)^2} \cosh(\frac{\xi \sqrt{\beta}}{\tau}) I_{[0, \infty]}(\beta)
$$

where $\cosh(z) = (e^z + e^{-z})/2$ and $I_{[0, \infty]}$ is an indicator function.

We shall develop a hierarchical framework by assuming non-informative priors for the hyperparameters $\xi$ and $\tau$. The hierarchical model is,

$$
y_1^*|\beta_1, \sigma_1 \sim N(x_1^* \beta_1, \sigma_1^2 I_n)
$$

$$
\beta_1|\xi, \tau \sim \tau^2 \chi_1^2(\xi^2/\tau^2)
$$

with $\pi(\sigma_1) = \sigma_1^{-1}$

and $\pi(\xi, \tau) = \tau^{-1}$, $i = 1, 2, \ldots, N$.

The first and third distributions above can be pooled together implying that

$$
y_1, \sigma_1|\beta_1 \sim N(x_1^* \beta_1, \sigma_1^2 I_n).
$$

From the second and fourth distributions, we can write the joint prior distribution for $\beta_1, \beta_2, \ldots, \beta_N$, $\xi$ and $\tau$ as

$$
\pi(\beta_1, \beta_2, \ldots, \beta_N, \xi, \tau) = \tau^{-(N+1)}(\prod_{i=1}^{N} \beta_i^{-1/2}) e^{-\frac{1}{2\tau^2} \sum_{i=1}^{N} (\beta_i + \xi)^2} \\
\times \prod_{i=1}^{N} \cosh(\frac{\xi \sqrt{\beta_i}}{\tau})
$$

\[\]
Integrating $\xi$ out of the above expression and after some simplification, we obtain

$$\pi(\beta_1, \beta_2, \ldots, \beta_N, \tau) = \tau^{-N} \left( \prod_{i=1}^{N} \beta_i^{-1/2} \right) e^{-\frac{1}{2\tau} \sum_{i=1}^{n} \left( \beta_i^2 - (\bar{B} - A)^2 \right)}$$

where

$$\bar{B} = \frac{1}{N} \sum_{i=1}^{n} \beta_i$$

and

$$\overline{A} = \frac{1}{N} \sum_{i=1}^{N} \log\cosh(\sqrt{\beta_i}).$$

From (15), integrating with respect to $\tau$ gives us the joint prior for $(\beta_1, \beta_2, \ldots, \beta_N)$ as

$$\pi(\beta_1, \beta_2, \ldots, \beta_N) = \left( \prod_{i=1}^{N} \beta_i^{-1/2} \right) \left[ \sum_{i=1}^{N} \beta_i^2 - (\bar{B} - A)^2 \right]^{-\frac{N-1}{2}}.$$ 

Combining this with our sample information as given in (14), we have the posterior distribution

$$\pi(\beta_1, \beta_2, \ldots, \beta_N | y, x) = \left[ \sum_{i=1}^{N} \beta_i^2 - (\bar{B} - A)^2 \right]^{-\frac{N-1}{2}}$$

$$\times \prod_{i=1}^{n} \left[ \beta_i^{-1/2} \left( \frac{\beta_i - \hat{\beta}_i}{s_i} \right)^2 \right]^{-1/2}$$

$$= \bar{N} \times \prod_{i=1}^{N} H(\beta_i) \quad \text{(say).}$$

The second part of this posterior is the p.d.f. that would result had we assumed $\xi$ is a constant. In that case no pooling of time series and cross section information occurs. In Figure 1, we plot this part of the posterior density (though still it is called posterior p.d.f.)
for $\hat{\beta}_i = 1$, $s_{\hat{\beta}_i} = .5$ and $T = 21$. As we observe from the figure, this will shrink the least squares estimator of $\beta_i$. Specifically the factor $(\beta^{-1/2})$ moves the "mode" towards zero, as follows easily from the fact that

$$\frac{dH(\beta_i)}{d\beta_i} < 0.$$ 

Cross-sectional information is truly reflected in the first part of the posterior

$$\overline{H} = \frac{-N-1}{2} \left[ \sum_{i=1}^{N} (\beta_i^2 - (\overline{\beta} - \overline{A})^2) \right]$$

where $\overline{A}$, logarithm of the geometric mean of $\cosh(\sqrt{\beta_1})$, incorporates the information from the other securities. Bayes estimator implied by the posterior (16), is comparable to $\beta_i^*$ in (9) where the pooling takes place through a weighted average of the cross-section betas. Unfortunately, the expression (16) cannot be integrated further to obtain the posterior p.d.f. of individual security betas.

4. Non-hierarchical Approach with Non-normal Prior

Instead of a hierarchical model, we now adopt a single stage root-normal prior for $\beta$. As in Vasicek, the hyperparameters $\overline{\beta}$ and $\phi$ are assumed to be known, although in applications these will have to be estimated. The model can now be written as,
\[ y^* | \theta, \sigma_1, \sigma_2, \ldots, \sigma_N \sim N(\mathbf{x}^* \theta, (\tau/2\pi I_N) \Sigma^*) \]
\[ \beta_i \sim \phi^2 \cdot \chi^2_{(1)}(\beta^2/\phi^2) \]
\[ \pi(\sigma_1, \sigma_2, \ldots, \sigma_N) = \left( \prod_{i=1}^{N} \sigma_i \right)^{-1} \]

where \( \theta' = (\beta_1, \beta_2, \ldots, \beta_N) \) and now \( x^* \) and \( y^* \) contain data on all the securities. Combining first and third distributions above, we have,

\[ y, \sigma_1, \sigma_2, \ldots, \sigma_m | \theta \sim N(\mathbf{x}' \theta, \Sigma) \]

where \( \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2) \otimes I_T \).

The joint posterior distribution of \( (\theta, \sigma_1, \sigma_2, \ldots, \sigma_m) \) conditional on \( \bar{\beta} \) and \( \phi \) can be derived along the lines similar to the previous section. However, applied work with that posterior will be slightly messy. Simpler results can be obtained by using a central \( \chi^2 \) approximation to the priors for \( \beta_i \)'s. A central \( \chi^2 \) approximation to our prior p.d.f. is [see Johnson and Kotz (1970, p. 139)],

\[ \frac{\beta_i}{c\phi^2} \sim \chi^2_f \]

where \( c = (1+\delta)^{-1}(1+2\delta), f = 1+\delta^2(1+2\delta)^{-1} \) and \( \delta = (\bar{\beta}/\phi)^2 \). Therefore, the prior for \( \beta_i \) is a gamma density whose kernel is,

\[ \beta_i \frac{f}{2} e^{-1} \beta_i^2 e^{-\beta_i/h}, \quad \beta_i > 0. \]

where \( h = 2c\phi^2 \). It is worth noting that, even though \( 2c\phi^2 \) is, for fixed \( f \), a scale parameter, \( f \) is not a location parameter. It is...
therefore difficult to develop a hierarchical model with non-
informative second stage priors based upon this central $\chi^2$
approximation.

As the $\beta_i$'s are (conditional on $\tilde{\beta}$ and $\phi$) i.i.d., the joint prior
is the product of the marginal priors, and the joint posterior for
$(\theta, \sigma_1, \sigma_2, \ldots, \sigma_N)$ can be written as,

$$\pi(\theta, \sigma_1, \sigma_2, \ldots, \sigma_N | y, x, f, h) = \prod_{i=1}^{N} \pi(\beta_i, \sigma_i | y, x, f, h)$$

$$- \frac{1}{2\sigma_i^2} \sum_{t=1}^{T} (y_{it} - \beta_i x_{it})^2 \frac{f}{2} - 1 - \beta_i \frac{\beta_i}{h}$$

where $\pi(\beta_i, \sigma_i | y, x, f, h) = \sigma_i^{-T} e^{-\frac{1}{2\sigma_i^2} \beta_i x_{it}}$ is the joint posterior distribution of $(\beta_i, \sigma_i)$. Integrating $\sigma_i$ out
one gets the conditional posterior density of $\beta_i$,

$$\pi(\beta_i | y, x, f, h) = \beta_i^{\frac{f}{2} - 1} e^{-\frac{\beta_i}{h}} \left( (T-2) + \left( \frac{\beta_i - \hat{\beta}_i}{s_{\beta_i}} \right)^2 \right)^{-\frac{T-1}{2}}, \quad \beta_i > 0.$$ The plot in Figure 2 illustrates the shape of this density. The
values of $f$ and $h$ are those implied by $\tilde{\beta} = 1.03$ and $\phi = 0.22$. These
values were selected from the findings in Bera and Kannan (1986). As
in Figure 1, we set $\hat{\beta}_i = 1$, $s_{\beta_i} = 0.5$ and $T = 21.$

It can easily be shown that the posterior mean exists. Indeed
for $T \geq 3$,

$$\int_0^{\beta_i} \pi(\beta_i | y, x, f, h) d\beta_i \leq \frac{f}{2} - \frac{\beta_i}{h} \int_0^{\beta_i} e^{-\frac{\beta_i}{h}} d\beta_i = \frac{f}{2} + 1,$$

where the last equality follows from the gamma p.d.f.

It is apparent that, at least for the parameter values used, the
posterior p.d.f. is unimodal and almost symmetric though slightly
skewed to the right. The least squares estimator can be larger or smaller than the modal value. In fact, straightforward algebra yields

\[ \text{sign}\left(\frac{d\pi(\hat{\beta}_i | y)}{d\hat{\beta}_i}\right) = \text{sign}\{h(f - 1) - \hat{\beta}_i\}. \]

Therefore, if \( \hat{\beta}_i > h(f - 1) \) it follows from the shape of the p.d.f. that the Bayes estimator \( \hat{\beta}^*_i \) under a quadratic loss will satisfy \( \hat{\beta}^*_i > \hat{\beta}_i \). Under 0-1 loss function, the Bayes estimators can be obtained from the modal values of \( \pi(\beta_i | x, i, h) \), \( i = 1, 2, \ldots, N \). For example, with \( \hat{\beta} = 1.03, \phi = .22 \) (implying \( f = 11.72 \) and \( h = .19 \)) one has the Bayes' estimator \( \hat{\beta}^*_i = .957 \), when we set \( \hat{\beta}_i = 1, s_{\hat{\beta}_i} = .5 \) and \( T = 21 \). If we take Bayes estimates as an improved predictor for beta, then the above observation agrees with the findings of earlier researchers that relatively high and low OLS beta estimates tend to overpredict and underpredict, respectively, the corresponding betas for the subsequent time period [see, e.g., Blume (1971) and Klemkosky and Martin (1975)].

5. **Concluding Remarks**

We have presented only some theoretical results. It would be interesting to apply our procedures to real data, and see whether that leads to improved forecasts for systematic risk. On the theoretical side, some other prior distribution can be used instead of a non-central \( \chi^2 \) distribution. One possibility is to use a mixture of two (or a few) normal distributions. A second possibility is to take a
normal prior for $\beta_1$ with mean modelled in terms of a regression function of some firm specific variables. The prior variance could also be defined from a regression model. Lastly, a number of other approximations for non-central $\chi^2$ distribution are available. For example, $\chi^2_1(\delta)$ can be approximated by a central $\chi^2_1$ with $1+\nu$ degrees of freedom where $\nu$ is a Poisson random variable with mean $\delta/2$. 
References


D/140
Fig. 1. Posterior p.d.f of beta for the hierarchical model
Fig. 2. Posterior p.d.f of beta for the nonhierarchical model with central $x^2$ approximation