THE HECKMAN BINDERY, INC.
North Manchester, Indiana

KRL

H or V	JUST	FONT	SLOT	TITLE
H	CC	1W	22	BEBR
			21	FACULTY
			20	WORKING
			19	PAPER
H	CC	1W	8	1989
			7	NO.1555-1572
H	CC	1W		330
				B385<"CV">
				no.1555-1572
				cop.2
H	CC	7W		<IMPRINT>
				U. of ILL.
				LIBRARY
				URBANA

BINDING COPY

PERIODICAL ☐ CUSTOM ☐ STANDARD ☐ ECONOMY ☐ THESIS
BOOK ☐ CUSTOM ☐ MUSIC ☐ ECONOMY AUTH 1ST ☐
NO VOLS. THIS TITLE — LEAD ATTACH

ACCOUNT LIBRARY NEW: 66672 001 — RUB OR SAMPLE — TITLE I.D.: 6632 — COLOR: FOIL WHI, MATERIAL 488

ACCOUNT NAME: UNIV OF ILLINOIS — ISSN.

ACCOUNT INTERNAL I.D.: B01912400 — NOTES — BINDING FREQUENCY — WHEEL — SYS. I.D: 3 — 39256

I.D. #2: STX3

COLLATING: 35

ADDITIONAL INSTRUCTIONS
Dept=STX3 Lot-#20 Item=143 HNM=1XY#
ICR2ST3CR MARK BY # B4 911

SEP. SHEETS | PTS. BD. PAPER | TAPE STUBS | CLOTH EXT. | FILLER | GUM | STUB | LEAF ATTACH.

POCKETS: PAPER | BUCK | CLOTH — SPECIAL PREP.

INSERT MAT — ACCOUNT LOT NO.: #20 — JOB NO.: HV363

PRODUCT TYPE: 11 — ACCOUNT PIECE NO: 143 — PIECE NO.: 85

HEIGHT: 11.2 — GROUP CARD: 1 — VOL. THIS TITLE — COVER SIZE: X

001247549

BEBR
FACULTY WORKING PAPER NO. 89-1567

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

Anil K. Bera
Jose A. F. Machado

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 89-1567

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

May 1989

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

Anil K. Bera, Associate Professor
Department of Economics

Jose A. F. Machado
Universidade Nova de Lisboa

Some earlier results of this paper were presented at an NSF Conference in Bayesian Inference in Econometrics, Duke University. We are thankful to the participants of that conference for constructive criticism. Thanks are also due to Paul Newbold for some pertinent comments on an earlier version of this paper and Pin Ng for drawing the figures. Responsibility for any errors and omissions in 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

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \qquad \begin{array}{l} i = 1,2,\ldots,N \\ t = 1,2,\ldots,T \end{array} \tag{1}$$

where R_{it} and R_{mt} are, respectively, the random return on security i and the corresponding random market return in period t, α_i and β_i are the regression parameters appropriate to security i and ε_{it} is the random disturbance term with mean zero and variance σ_i^2. The parameter β_i, called beta, measures the systematic risk of the security i and is defined as $Cov(R_i, R_m)/Var(R_m)$.

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 β_i. Vasicek (1973) suggested a Bayesian adjustment technique using a normal prior for β_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 β_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 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 $\sqrt{\beta_i}$ is normally distributed, or in other words, β_i has a noncentral χ^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 go back to Vasicek's set up and combine it with our non-central chi-square prior distribution and its various approximations. 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 which attempts to improve upon it. Also in the statistics literature, most of the Bayesian regression analysis are based on normal priors primarily because of its 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 β_i parameters, it is convenient to work with the model in deviation form

$$y_{it} = \beta_i x_{it} + u_{it} \tag{2}$$

where $y_{it} = R_{it} - \bar{R}_i$, $x_{it} = R_{mt} - \bar{R}_m$ and $u_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$. So we will have $u_{it} \sim N(0, \rho_i^2 I_T)$ where $\rho_i^2 = (1 - \frac{1}{T})\sigma_i^2$. Under the classical framework, the maximum likelihood or the OLS estimator of β_i is given by

$$\hat{\beta}_i = \frac{\sum_{t=1}^{T} x_{it} y_{it}}{\sum_{t=1}^{T} x_{it}^2} \qquad i = 1,2,\ldots,N$$

with $V(\hat{\beta}_i) = \rho_i^2 / \sum_{t=1}^{T} x_{it}^2 \equiv S_i^2$. Vasicek (1973) suggested a Bayesian approach with normal prior for β_i

$$\beta_i \sim N(\bar{\beta}, \phi^2) \tag{3}$$

and a non-informative prior density for ρ

$$\pi(\rho_i) \alpha \frac{1}{\rho} .$$

The standard Bayes estimator is the posterior mean

$$\frac{(\bar{\beta}_i/\phi^2) + (\hat{\beta}_i/S_i^2)}{(1/\phi^2) + (1/S_i^2)} \tag{4}$$

and the posterior variance is given by $((1/\phi^2) + (1/S_i^2))^{-1}$. Vasicek suggested to use the mean and variance of the cross-sectional betas in place of $\bar{\beta}$ and ϕ^2 respectively. Empirical results in Bera and Kannan (1986, Tables VII and VIII) show that forecasts based on the Vasicek's adjusted betas 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.

From (2) we can write

$$y_i|\beta_i \sim N(x_i\beta_i, \rho_i^2 I_T) \qquad i = 1,2,\ldots,N \tag{5}$$

where $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ and $x_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$.

Next we assume exchangeability among the β_i, specifically

$$\beta_i|\xi \sim N(\xi, \tau^2) \tag{6}$$

with a second stage non-informative prior for ξ. Due to the randomness of ξ, 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

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

where $f(\xi)$ is the probability density function of ξ. Therefore, $\pi(\beta_1, \beta_2, \ldots, \beta_N)$ is a mixture of IID distributions conditional on ξ, 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

$$y|\theta_1 \sim N(A_1\theta_1, C_1)$$

$$\theta_1|\theta_2 \sim N(A_2\theta_2, C_2)$$

$$\theta_1|\theta_3 \sim N(A_3\theta_3, C_3)$$

with $C_3^{-1} = 0$. For this model Bayesian inference can be drawn from the posterior for θ_1 given $\{A_i\}$, $\{C_i\}$, 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_i, C_i and θ_i, for our specification (5) and (6), we obtain the following posterior distribution for $\theta_1 = (\beta_1, \beta_2, \ldots, \beta_N)'$ as N(Dd,D) where

$$D^{-1} = \mathrm{diag}\left(\frac{x_1'x_1}{\rho_1^2} + \frac{1}{\tau^2}, \ldots, \frac{x_N'x_N}{\rho_N^2} + \frac{1}{\tau^2}\right) - \frac{J_N}{N\tau^2} \tag{7}$$

where J_N is an N×N matrix whose all elements are one, and

$$d = \left(\frac{x_1'y_1}{\rho_1^2}, \ldots, \frac{x_N'y_N}{\rho_N^2}\right). \tag{8}$$

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

$$\beta_i^* = \left(\frac{x_i'x_i}{\rho_i^2} + \frac{1}{\tau^2}\right)^{-1} \frac{x_i'x_i}{\rho_i^2} \hat{\beta}_i + \left(\frac{x_i'x_i}{\rho_i^2} + \frac{1}{\tau^2}\right)^{-1} \frac{1}{\tau^2}\left(\sum_{j=1}^{N} w_j \hat{\beta}_j\right)$$

where $$w_j = \left[\sum_{i=1}^{N} \frac{x_i'x_i}{\tau^2 x_i'x_i + \rho_i^2}\right]^{-1} \frac{x_j'x_j}{\tau^2 x_j'x_j + \rho_j^2}. \tag{9}$$

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. β_i^* can also be expressed as

$$\beta_i^* = \frac{\rho_i^2 \tau^2}{\tau^2(x_i' x_i) + \rho_i^2} \left(\frac{\hat{\beta}_i}{\rho_i^2 / x_i' x_i} + \frac{\overline{\beta}^*}{\tau^2} \right)$$

where $\overline{\beta}^* = \sum_{j=1}^{N} \beta_j^*/N$. 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 $\overline{\beta}^*$ is incorporated in a self-fulfilling way, in the sense that the cross-sectional average beta is consistent with the estimates for individual securities.

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

$$y_i | \beta_i \sim N(x_i \beta_i, \rho_i^2 I_N)$$

$$\beta_i | \overline{\beta} \sim N(\overline{\beta}, \phi^2). \qquad (10)$$

Here the first stage prior is "completely specified" by the cross-section data. The counterparts (7) and (9) are respectively,

$$D_V^{-1} = \mathrm{diag}\left(\frac{x_1'x_1}{\rho_1^2} + \frac{1}{\phi^2}, \ldots, \frac{x_N'x_N}{\rho_N^2} + \frac{1}{\phi^2}\right) \tag{11}$$

and

$$\beta_{iV}^{*} = \left(\frac{x_i'x_i}{\rho_i^2} + \frac{1}{\phi^2}\right)^{-1} \frac{x_i'x_i}{\rho_i^2}\hat{\beta}_i + \left(\frac{x_i'x_i}{\rho_i^2} + \frac{1}{\phi^2}\right)^{-1} \frac{1}{\phi^2}\left(\sum_{j=1}^{N} \frac{\hat{\beta}_j}{N}\right). \tag{12}$$

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

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

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 $V(\bar{\beta}) = 0$ whereas for the hyperparameter model we assume second stage non-informative prior, i.e., $V(\xi)^{-1} = 0$.

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.,

$$\sqrt{\beta} \sim N(\tilde{\beta}, \phi^2).$$

Then β/ϕ^2 will be distributed as a noncentral χ^2 with one degree of freedom and noncentrality parameter $(\tilde{\beta}/\phi)^2$ denoted by $\chi_1^2(\tilde{\beta}^2/\phi^2)$. The p.d.f. of β can be written as

$$\pi(\beta|\tilde{\beta},\phi^2) = (2\pi)^{-1/2}(\beta\phi^2)^{-1/2} e^{-\frac{1}{2\phi^2}(\beta+\tilde{\beta})^2} \cosh\left(\frac{\tilde{\beta}\sqrt{\beta}}{\phi^2}\right) I_{(0,\infty)}(\beta) \quad (13)$$

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

To avoid cumbersome results and to make the comparison with Vasicek's analysis clearer, we shall focus on the case of a single security. Therefore, without the loss of generality we can suppress the index i.

We shall develop a hierarchical framework by assuming non-informative priors for the hyperparameters $\tilde{\beta}$ and ϕ. The hierarchical model is,

$$y|\beta,\rho \sim N(x\beta, \rho^2 I_T)$$

$$\beta|\tilde{\beta},\phi \sim \phi^2 \chi^2_{(1)}(\tilde{\beta}^2/\phi^2)$$

with $\quad \pi(\rho) \propto \rho^{-1}$

and $\quad \pi(\tilde{\beta},\phi) \propto \phi^{-1}.$

The major difference between this set-up and Lindley and Smith's (1972) is that now the prior for β is not linear in the hyperparameter $\tilde{\beta}$. Because of this feature one needs the explicit consideration of the non-informative priors for ρ and ϕ.

Direct application of the Bayes formula using the above setup and the prior density in (13) yields the joint posterior p.d.f. of the unknown parameters,

$$\pi(\beta,\tilde{\beta},\rho,\phi\,|y,x) \propto \rho^{-(T+1)} \exp\{-\frac{1}{2\rho^2}(y-x\beta)'(y-x\beta)\}$$

$$\times\ \phi^{-2}\beta^{-1/2}[\exp\{-\frac{1}{2\phi^2}(\sqrt{\beta}+\tilde{\beta})^2\} + \exp\{-\frac{1}{2\phi^2}(\sqrt{\beta}-\tilde{\beta})^2\}] \qquad (14)$$

for β, $\tilde{\beta}$, ρ and ϕ in $(0,\infty)$.

Noticing that (by the "normal integral")

$$\int_{-\infty}^{+\infty} \exp(-\frac{1}{2\phi^2}(\sqrt{\beta}\pm\tilde{\beta})^2)d\tilde{\beta}$$

is proportional to ϕ, integration of the posterior density in (14) with respect to $\tilde{\beta}$ gives

$$\pi(\beta,\rho,\phi\,|y,x) \propto \rho^{-(T+1)} \exp\{-\frac{1}{2\rho^2}(y-x\beta)'(y-x\beta)\}\beta^{-1/2}\phi^{-1}$$

$$\propto \rho^{-(T+1)}\exp\{-\frac{1}{2\rho^2}[(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2]\}\beta^{-1/2}\phi^{-1}$$

where $\hat{\beta} = (x'x)^{-1}x'y$ and $s^2_{\hat{\beta}} = (y-x\hat{\beta})'(y-x\hat{\beta})(x'x)^{-1}/(T-1)$. This expression clearly implies that

$$\pi(\beta,\rho\,|\,y,x) \propto \rho^{-(T+1)}\exp\{-\frac{1}{2\rho^2}[(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2]\}\beta^{-1/2}.$$

Integrating this joint density with respect to ρ (performing the change of variable, $z \equiv a/2\rho^2$, $a = (T-1) + [(\beta-\hat{\beta})/s_{\hat{\beta}}]^2$) one gets the marginal posterior p.d.f. of β as

$$\pi(\beta\,|\,y,x) \propto \beta^{-1/2}\{(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2\}^{-T/2}, \quad \beta > 0. \tag{15}$$

The main features of $\pi(\beta\,|\,y,x)$ can be seen in Figure 1 (plot for $\hat{\beta} = 1$, $s_{\hat{\beta}} = .5$, $T = 20$). It is apparent that the posterior density is slightly skewed to the left, and therefore, the posterior mean will be smaller than the least squares estimator of β.

When a non-informative or normal prior for β is assumed, the marginal posterior of β belongs to the t-family and the posterior mean is equal to the least squares estimator. Now the posterior p.d.f. is like a t-density with an additional factor ($\beta^{-1/2}$) and range $(0,\infty)$. Without the additional term the posterior mode (as well as the mean) would be $\hat{\beta}$, the least squares estimator. The term $\beta^{-1/2}$ moves the posterior mode towards zero, as follows easily from the fact that $d\pi(\hat{\beta}\,|\,y,x)/d\beta < 0$.

It can be shown easily that the posterior mean of β exists. Therefore, there exists the Bayes estimator under a quadratic loss function. Indeed, noticing that for $T \geq 3$ the term

$$\{(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2\}^{-T/2} \tag{16}$$

in (15) is smaller than one and that $\beta > \beta^{1/2}$ for $\beta > 1$,

$$\int_0^\infty \beta\pi(\beta|y,x)d\beta \leq \int_0^1 \beta^{1/2}d\beta + \int_0^\infty \beta\{(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2\}^{-T/2}d\beta.$$

Direct integration and the use of the t-family density [see, e.g., Rao (1973, p. 171)] reveal that the right-hand side is equal to

$$\frac{2}{3} + \frac{s_\beta^{*2}}{T-2}\,[(T-1) + \frac{\hat{\beta}^2}{s_\beta^2}]^{-\frac{T}{2}+1} + \frac{\hat{\beta}}{2}\,\frac{(T-1)^{\frac{T-1}{2}}}{B(\frac{1}{2},\frac{T-1}{2})}$$

where $B(\cdot,\cdot)$ is the beta function.

The posterior mode can be computed explicitly if we approximate the t-kernel by the normal, i.e., if one writes the posterior density in (15) as,

$$\beta^{-1/2}\exp\{-\frac{(\beta-\hat{\beta})^2}{2s_\beta^{*2}}\}.$$

For large data sets no crucial loss is incurred by using this approximation. It is easy to show that if the posterior mode exists, it is equal to

$$\frac{\hat{\beta}}{2} + \sqrt{\frac{\hat{\beta}^2}{4} - \frac{s_\beta^{*2}}{2}}\;.$$

As the first derivative of the approximate posterior density is quadratic the mode exists if $\hat{\beta}^2/4 \geq s_\beta^{*2}/2$. Otherwise, the posterior density will be monotonically decreasing, resembling a χ^2 distribution with one degree of freedom. The mode of the above approximation to the posterior p.d.f. provides an easy way to compute the Bayes estimator (under a 0-1 loss function) and shows that this estimator is obtained by shrinking the OLS estimator $\hat{\beta}$ towards zero.

In the following section we shall develop a non-hierarchical approach which may be easier to compete and is more in the line of Vasicek's work. Three cases will be considered: one based on the root-normal prior for beta and the other two based on central χ^2 and normal approximations to the prior density of beta.

4. Non-hierarchical Approach with Non-normal Prior

Instead of a hierarchical model, we now adopt a single stage root-normal prior for β. As in Vasicek, the hyperparameters $\tilde{\beta}$ and ϕ are assumed to be known, although in applications these will have to be estimated.

The model can now be written as,

$$y|\beta \sim N(x\beta, \rho^2 I_T)$$

$$\beta \sim \phi^2 \cdot \chi^2_{(1)}(\tilde{\beta}^2/\phi^2)$$

$$\pi(\rho) \propto \rho^{-1}$$

The conditional (on $\tilde{\beta}$ and ϕ) joint posterior p.d.f. of (β, ρ) is,

$$\pi(\beta,\rho|y,x,\tilde{\beta},\phi) \propto \rho^{-(T+1)} \exp\{-\frac{1}{2\rho^2}(y-x\beta)'(y-x\beta)\}$$
$$\times \beta^{-1/2}[\exp\{-\frac{1}{2\phi^2}(\sqrt{\beta}+\tilde{\beta})^2\} + \exp\{-\frac{1}{2\phi^2}(\sqrt{\beta}-\tilde{\beta})^2\}].$$

Integration with respect to ρ yields,

$$\pi(\beta|y,x,\tilde{\beta},\phi) \propto \beta^{-1/2}[\exp\{-\frac{1}{2\phi^2}(\sqrt{\beta}+\tilde{\beta})^2\} + \exp\{-\frac{1}{2\phi^2}(\sqrt{\beta}-\tilde{\beta})^2\}]$$
$$\times \{(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2\}^{-T/2} \qquad (17)$$

for $\beta > 0$.

It is apparent from this expression, and confirmed by Figure 2 that conditional marginal p.d.f. of β has a left tail similar to a χ^2 density. Figure 2 plots the posterior p.d.f. for $\tilde{\beta} = 1.03$, $\phi = .22$ ($\hat{\beta} = 1$, $s_{\hat{\beta}} = .5$, $T = 20$). The values of $\tilde{\beta}$ and ϕ were selected from the findings in Bera and Kannan (1986). For $T \geq 3$, as we noted in (16), the last term in (17) is less than one, and therefore,

$$\int_0^\infty \beta\pi(\beta|y,x,\tilde{\beta},\phi)d\beta \leq \int_0^\infty \beta k(\chi_1^2(\tilde{\beta}^2/\phi^2))d\beta = (1 + \frac{\tilde{\beta}^2}{\phi^2})(2\pi\phi^2)^{1/2},$$

where $k(\chi_1^2(\cdot))$ represents the kernel of a non-central χ^2 density, and the last equality follows from the fact that $E(\chi_\nu^2(\delta)) = \nu + \delta$. Hence the posterior mean of β conditional on $\tilde{\beta}$ and ϕ exists.

The conditional p.d.f. in (17) is directly comparable with expression (14) in Vasicek (1973). The major difference is that here the non-central χ^2 kernel replaces the normal kernel found in Vasicek's result. The "t-like" term is common to either expression. It is also worth recalling that in the hierarchical approach the two exponential terms were integrated out of the p.d.f.

The comparison with Vasicek's results is perhaps clearer if one uses a normal approximation to the non-central χ^2 prior for β. The approximated prior is [see Johnson and Kotz (1970, p. 139)]

$$\pi(\beta) = \frac{1}{\sqrt{2\pi}}\,\frac{A}{2}\,\frac{1}{\phi}\,\beta^{-1/2}\,\exp\{-\frac{1}{2}(\frac{A}{\phi}\sqrt{\beta} - B)^2\}$$

where $A = [(1+\delta)/(1+2\delta)]^{1/2}$, $B = [(1+2\delta^2+2\delta)/(1+2\delta)]^{1/2}$ and $\delta = (\tilde{\beta}/\phi)^2$ is the non-centrality parameter.

The same arguments as before yield the conditional posterior p.d.f. of β,

$$\pi(\beta \mid y, x, \tilde{\beta}, \phi) \alpha \beta^{-1/2} \exp\{-\frac{1}{2}(\frac{A}{\phi}\sqrt{\beta} - B)^2\}$$

$$\times [(T-1) + (\frac{\beta - \hat{\beta}}{s_{\hat{\beta}}})^2]^{-T/2}.$$

This expression highlights what was already remarked when commenting on the marginal (unconditional) p.d.f. of β: from a practical standpoint the term $\beta^{-1/2}$ appears to be the crucial modifier as it implies a thicker left tail than the one obtained using a normal prior. Therefore the posterior mean will be closer to zero.

Applied work with the above posteriors will be slightly messy. Simpler results can be obtained by using a central χ^2 approximation to the prior for β. A central χ^2 approximation to our prior p.d.f. is [see Johnson and Kotz (1970, p. 139)],

$$\frac{\beta}{c\phi^2} \sim \chi_f^2$$

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

$$\beta^{\frac{f}{2}-1} e^{-\beta/h}, \qquad \beta > 0.$$

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 χ^2 approximation.

Using the same model as before and integrating ρ out one gets the conditional posterior density of β,

$$\pi(\beta\,|y,x,f,h) \propto \beta^{\frac{f}{2}-1} e^{-\frac{\beta}{h}} \{(T-1) + (\frac{\beta-\hat{\beta}}{s_{\hat{\beta}}})^2\}^{-\frac{T}{2}}, \quad \beta > 0.$$

The plot in Figure 3 illustrates the shape of this density. The values of f and h are those implied by $\tilde{\beta} = 1.03$ and $\phi = .22$ (c.f. Figure 2), and again $\hat{\beta} = 1$, $s_{\hat{\beta}} = .5$ and $T = 20$.

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

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

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

$$\text{sign}\{\frac{d\pi(\hat{\beta}\,|y)}{d\beta}\} \equiv \text{sign}\{h(\frac{f}{2} - 1) - \hat{\beta}\}.$$

Therefore, if $\hat{\beta} \lessgtr h(\frac{f}{2} - 1)$ it follows from the shape of the p.d.f. that the Bayes estimator β^* under a quadratic loss will satisfy $\beta^* \gtrless \hat{\beta}$. 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 χ^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 β_i 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 χ^2 distribution are available. For example, $\chi_1^2(\delta)$ can be approximated by a central χ_1^2 with $1+\nu$ degrees of freedom where ν is a Poisson random variable with mean $\delta/2$.

References

Bera, A. K. and S. Kannan (1986), "An Adjustment Procedure for Predicting Systematic Risk," Journal of Applied Econometrics, 1, 317-332.

Blume, M. E. (1971), "On the Assessment of Risk," Journal of Finance, 26, 1-10.

Haitovsky, Y. (1986), "The Linear Hierarchical Model and Its Applications in Econometric Analysis," in Bayesian Inference and Decision Techniques, eds., P. Goel and A. Zellner (Amsterdam: Elsevier Science Publishers), pp. 119-138.

Ilmakunnas, P. (1986), "Stochastic Constraints on Cost Function Parameters: Mixed and Hierarchical Approaches," Empirical Economics, 11, 69-80.

Jarque, C. M. and A. K. Bera (1987), "Test for Normality of Observations and Regression Residuals," International Statistical Review, 55, 163-172.

Johnson, N. L. and S. Kotz (1970), Continuous Univariate Distributions, Vol. 2, New York: Wiley.

Kadiyala, K. R. and D. Oberhelman (1986), "Estimation of Actual Realizations in Stochastic Parameter Models," in Bayesian Inference and Decision Techniques, eds., P. Goel and A. Zellner (Amsterdam: Elsevier Science Publishers), pp. 363-373.

Klemkosky, R. C. and J. D. Martin (1975), "The Adjustment of Beta Forecasts," Journal of Finance, 30, 1123-1128.

Lindley, D. V. and A. F. M. Smith (1972), "Bayes Estimates for the Linear Model," Journal of the Royal Statistical Society, Series B, 34, 1-41.

Trivedi, P. K. (1980), "Small Samples and Collateral Information: An Application of the Hyperparameter Model," Journal of Econometrics, 12, 301-318.

Vasicek, O. A. (1973), "A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas," Journal of Finance, 28, 1233-1239.

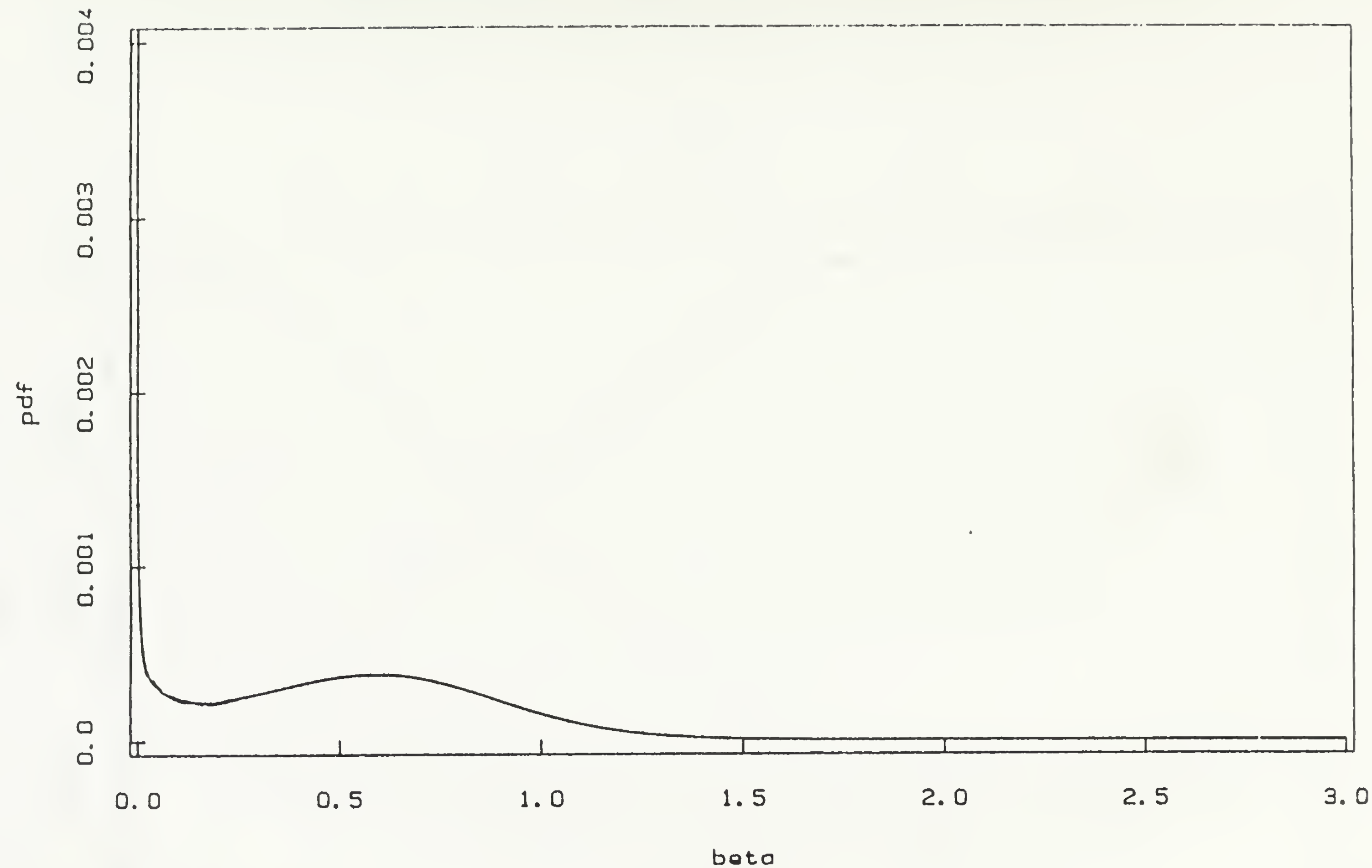

Figure 1 : Posterior p.d.f. of beta for the hierarchical model

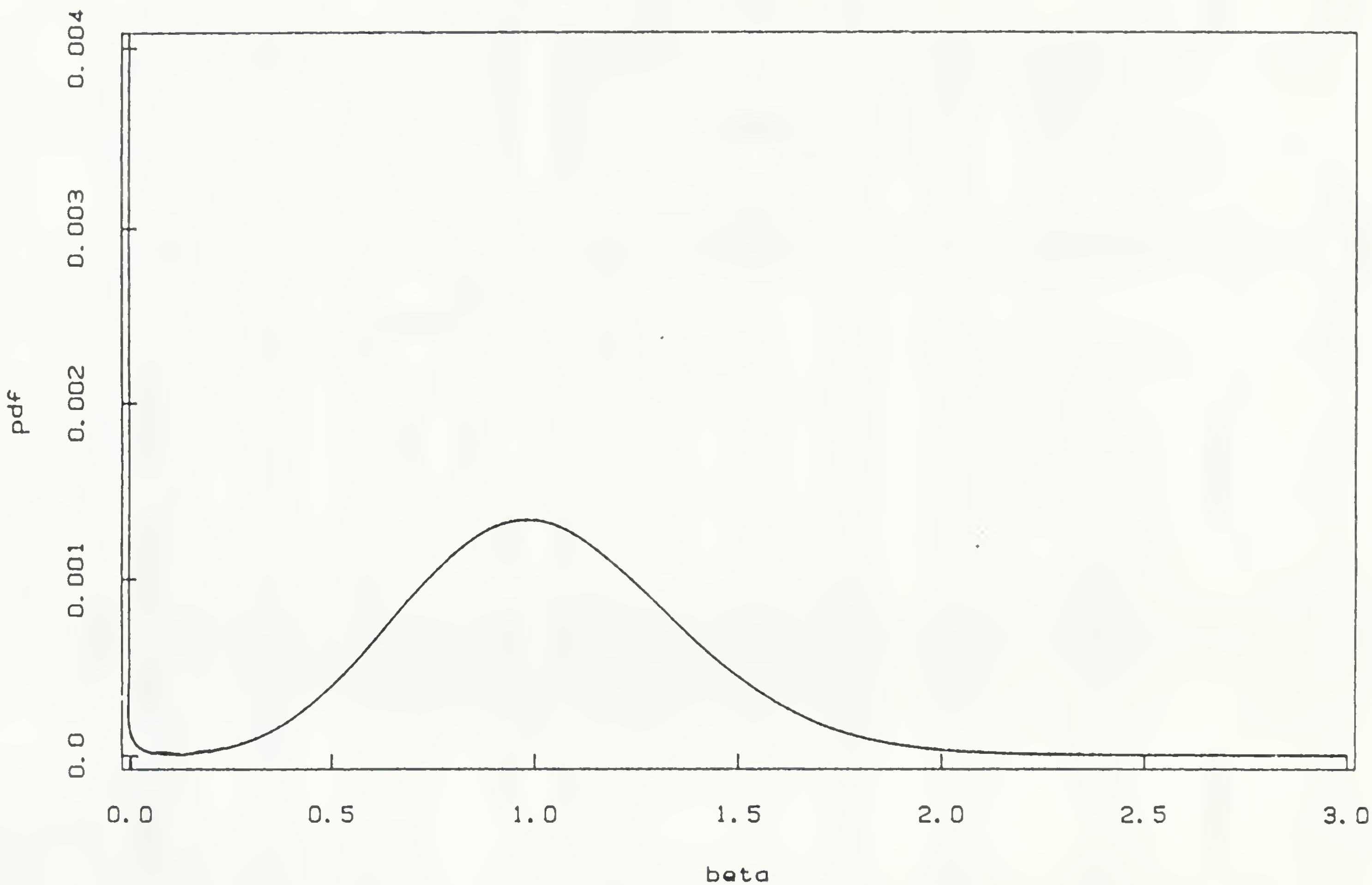

Figure 2 : Posterior p.d.f. of beta for the nonhierarchical model

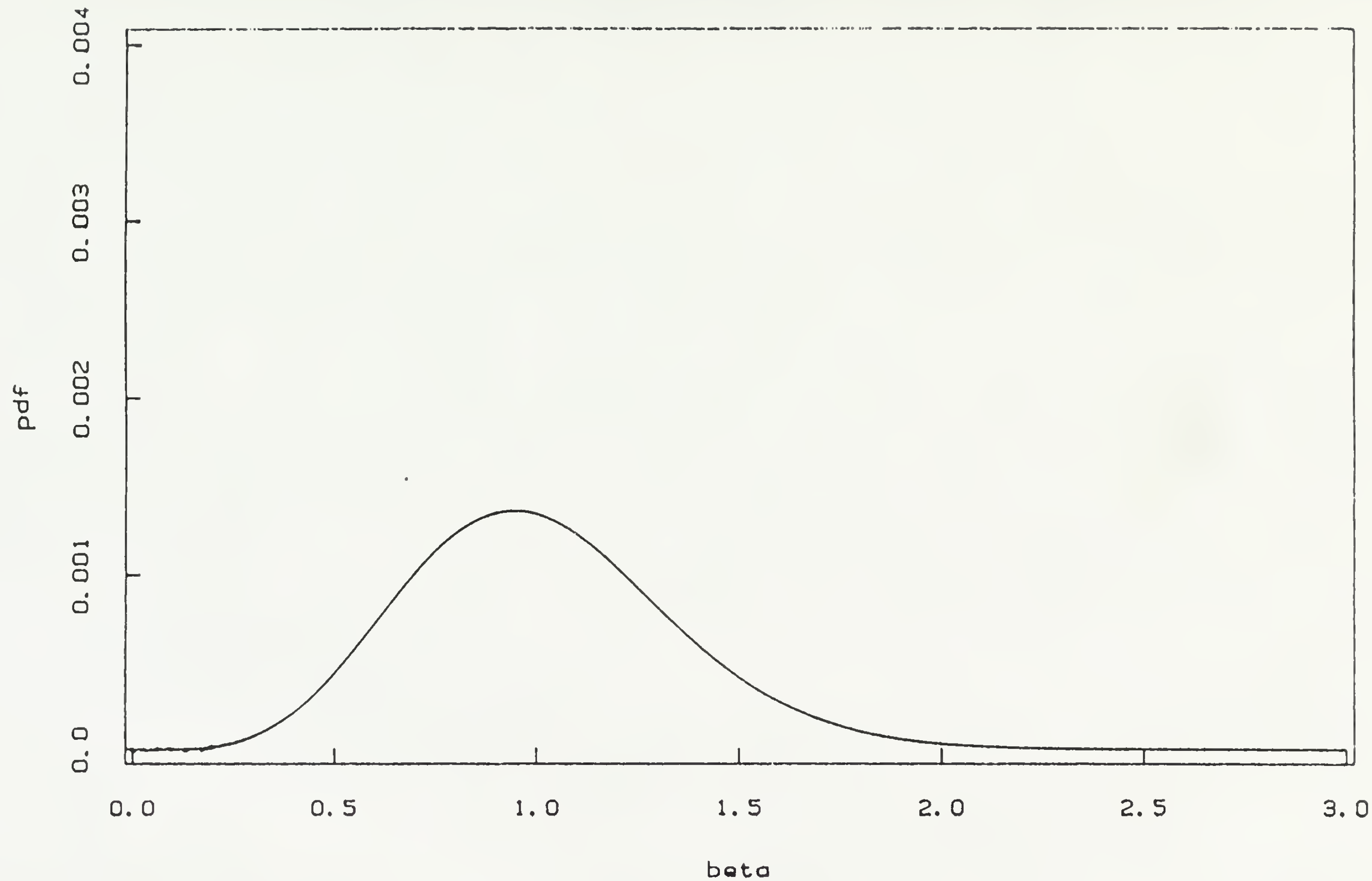

Figure 3 : Posterior p.d.f. of beta for the nonhierarchical model with central χ^2 approximation