

ROBUST ESTIMATION IN CAPITAL ASSET PRICING MODEL

WING-KEUNG WONG

ecswwk@nus.edu.sg

*Department of Economics, National University of Singapore
10 Kent Ridge Crescent, Singapore 119260*

GUORUI BIAN

stabiang@nus.edu.sg

*Department of Statistics and Applied Probability
National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260*

Abstract. Bian and Dickey (1996) developed a robust Bayesian estimator for the vector of regression coefficients using a Cauchy-type g -prior. This estimator is an adaptive weighted average of the least squares estimator and the prior location, and is of great robustness with respect to flat-tailed sample distribution. In this paper, we introduce the robust Bayesian estimator to the estimation of the Capital Asset Pricing Model (CAPM) in which the distribution of the error component is well-known to be flat-tailed. To support our proposal, we apply both the robust Bayesian estimator and the least squares estimator in the simulation of the CAPM and in the analysis of the CAPM for US annual and monthly stock returns. Our simulation results show that the Bayesian estimator is robust and superior to the least squares estimator when the CAPM is contaminated by large normal and/or non-normal disturbances, especially by Cauchy disturbances. In our empirical study, we find that the robust Bayesian estimate is uniformly more efficient than the least squares estimate in terms of the relative efficiency of one-step ahead forecast mean square error, especially for small samples.

Keywords: Robustness, Bayesian Estimate, Least Squares Estimate, Cauchy-type g -prior, Flat-tailed Distribution, Capital Asset Pricing Model.

1. Introduction

Both financial economists and statisticians have been concerned with the distributions of stock market returns. Fama (1963, 1965a, 1965b) and many others analyzed the empirical data and concluded that the normality assumption in the distribution of a security or portfolio return is violated such that the distribution is ‘flat-tailed’. They suggested the family of stable Paretian distributions between normal and Cauchy distributions for the stock returns.

On the other hand, Blattberg and Gonedes (1974) examined the return to security and suggested student- t as an alternative ‘flat-tail’ distribution for the return. Clark (1973), Christie (1983), Kon (1984) and Tse (1991) suggested a mixture of normal distributions for the stock return while Fielitz and Rozelle (1983) suggested that a mixture of non-normal stable distributions would be a better representation of the distribution of the return.

The distributional structure of the return may carry over into the structure of the disturbance in the Capital Asset Pricing Model (CAPM). In this situation, the dis-

tribution of the disturbance is ‘flat-tailed’ and the mixture of normal distributions or mixture of normal and Cauchy distributions may give a better description of the distribution of the disturbance in the CAPM. Harvey and Zhou (1993) supported this idea and pointed out that the non-normality in the return may carry over into non-normality of the disturbance in the CAPM. They examined the residuals of the world market portfolios in the CAPM and found that the distributions of the disturbances departed from normality in many cases. They then tested the sensitivity of the benchmark in the CAPM by specifying error structures that follow t -distributions or mixtures of normal distributions.

Bian and Dickey (1996) developed the robust Bayesian estimator for the vector of regression coefficients using a Cauchy-type g -prior. They showed that this robust Bayesian estimator is adaptive and markedly robust with respect to a flat-tailed sample distribution as compared to both the least squares estimator (LSE) and the usual Bayesian estimator.

Based on the ‘flat-tail’ characteristic on the distributions of the security or portfolio returns and their corresponding disturbances in the CAPM, we recommend the robust Bayesian estimator for the estimation of the parameters of the CAPM for the stock returns. The findings by Bian and Dickey (1996) lead us to hypothesize that the robust Bayesian estimator is more appropriate in the estimation of the CAPM in the sense that it is more efficient than the LSE.

To illustrate the superiority of the proposed Bayesian estimator, we simulate the LSE and the proposed estimator for a CAPM model. Based on the simulation results, we find that the proposed Bayesian estimator is superior to LSE when the CAPM model is contaminated by large normal and/or non-normal disturbances, especially by Cauchy disturbances.

To test our hypothesis, we also apply the one-step ahead forecasting technique to compare the robust Bayesian estimator with the traditional least squares estimator, LSE, in the estimation of the parameters in the CAPM for the US annual and monthly stock returns. The one-step ahead forecasting technique is commonly used to compare the performance of different models, see Clements and Hendry (1997). In our empirical study, we find that the robust Bayesian estimate is uniformly more efficient than the LSE in terms of relative efficiency of one-step ahead forecast mean square error, especially for small samples. Hence we recommend the robust Bayesian estimator for the estimation of the CAPM.

Many applications in finance involve prior beliefs about the behavior of the data. However, almost all empirical analyses have been carried out in the classical framework. There have been relatively few studies done, which applied the Bayesian approach in finance. Among them are Shanken (1987), Gibbons, Ross, and Shanken (1989), McCulloch and Rossi (1991) and Harvey and Zhou (1990). Two practical difficulties in implementing the approach have resulted in the slow adoption of Bayesian econometrics. The first is how to choose a prior and how to specify prior parameters. The other difficulty lies in evaluating the posterior distribution.

To overcome these difficulties, Harvey and Zhou (1990) imposed a prior on all the parameters of the multivariate regression model and used Monte Carlo numerical

integration to accurately evaluate 90-dimensional integrals to estimate the parameters in the posterior distribution. They developed a Bayesian framework to test the mean-variance efficiency of a given portfolio. The test is more direct than Shanken's (1987).

In recent studies, MacKinlay and Richardson (1991) developed the tests of unconditional mean-variance efficiency under weak distributional assumptions using a Generalized Method of Moments framework and concluded that the efficiency indexes can be sensitive to the test considered. Kandel, McCulloch and Stambaugh (1995) used a Bayesian approach to investigate a sample's information about a portfolio's degree of inefficiency and found that the NYSE-AMEX market portfolio is rather inefficient in the presence of a riskless asset.

There are two main issues on CAPM. One is testing the efficiency hypothesis, another is the estimation of the CAPM model, refer to Chapter 5 in Campbell, Lo and MacKinlay, 1997. In our paper, we address the latter issue by proposing an efficient method to overcome the difficulties in both getting the prior information and evaluating the posterior distribution. The proposed prior is an independent Cauchy and improper g -prior which is a robust prior. As such, the resulting estimator is adaptive and robust. We may acquire the information from the previous corresponding sample to specify the values of the prior parameters in practice. This approach makes the computation of the Bayesian estimate as easy as that of the LSE. It overcomes the need for computing integrals of any dimension for the estimation.

In Section 2, we review the least squares estimator, LSE, the usual Bayesian estimator and the robust Bayesian estimators. Section 3 reviews the theory of the standard CAPM, the non-stationarity of Beta parameter, the 'flat-tail' distribution of the security return and discusses applying the robust Bayesian estimator for the estimation of the CAPM. Section 4 reveals the results of the simulation for the Bayesian estimator and the least squares estimator when the CAPM is contaminated by normal and/or non-normal disturbances. We apply both the robust Bayesian estimator and the least squares estimator, LSE, in the estimation of the CAPM for the US annual and monthly stock returns and compare their efficiency in Section 5. The conclusion is in the last section.

2. A review of the least squares estimator and the Bayesian estimators of regression coefficients

The model considered is the normal linear multiple regression model (NLR) with the standard form:

$$y = X\beta + e \tag{1}$$

where y is an $n \times 1$ vector of observations on the dependent variable, X is an $n \times p$ design matrix with rank p , β is a $p \times 1$ vector of regression parameters with unknown value, and e is the $n \times 1$ vector of disturbances. It is assumed that the elements

of e are independently drawn from a normal distribution with mean 0 and finite variance σ^2 . The likelihood function for the NLR is

$$l(\beta, \sigma|y, X) \propto \sigma^{-n} \exp[-(y - X\beta)'(y - X\beta)/2\sigma^2]$$

In this model, the traditional estimator is the LSE which is also the maximum likelihood estimator for β . It is given by

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2)$$

So far in the literature of Bayesian estimation of regression coefficients, only the conjugate prior and the non-informative prior have been employed extensively in statistical estimation. The conjugate prior for the regression model (1) is a normal-reciprocal gamma distribution given by:

$$p_N(\beta, \sigma^{-2}) = f_N(\beta|\sigma) h(\sigma^{-2})$$

$$f_N(\beta|\sigma) \propto \sigma^{-p} \exp[-(\beta - \beta_0)'A(\beta - \beta_0)/2\sigma^2] \quad (3)$$

and

$$h(\sigma^{-2}) \propto (\sigma^{-2})^{\nu_0/2-1} \exp[-(\nu_0 s_0^2)/2\sigma^2]$$

The usual non-informative prior is:

$$p(\beta, \sigma) \propto \frac{1}{\sigma}$$

The posterior density of β and σ^{-2} associated with the conjugate prior (3) is:

$$f(\beta, \sigma^{-2}|y) \propto (\sigma^{-2})^{(n+\nu_0+p)/2-1} \times \exp\left\{-\frac{\sigma^{-2}}{2}[\nu_0 s_0^2 + \|y - X\hat{\beta}\|^2 + \|X(\beta - \hat{\beta})\|^2 + (\beta - \beta_0)'A(\beta - \beta_0)]\right\}$$

The Bayesian estimator of β , under quadratic loss, is the posterior mean of β . It is given by

$$E_N(\beta|y) = (X'X + A)^{-1}(X'X\hat{\beta} + A\beta_0) \quad (4)$$

where $\hat{\beta}$ is the LSE of β specified by (2). Zellner (1986) modified the above approach by considering the normal g -prior specified by the following forms:

$$p(\beta, \sigma^{-2}) \propto f(\beta|\sigma, g)h(\sigma^{-2})$$

$$f(\beta|\sigma, g) \propto \sigma^{-p} \exp[-g(\beta - \beta_0)'X'X(\beta - \beta_0)/2\sigma^2]$$

and

$$h(\sigma^{-2}) \propto (\sigma^{-2})^{\nu_0/2-1} \exp[-(\nu_0 s_0^2)/2\sigma^2]$$

This is a special case of the conjugate prior with the covariance matrix A^{-1} proportional to $(X'X)^{-1}$, the covariance matrix of the LSE. From (4), the Bayesian estimator, $\hat{\beta}_N$, of β becomes

$$\hat{\beta}_N = E_N(\beta|y, g) = \frac{\hat{\beta} + g\beta_0}{1 + g} \quad (5)$$

The prior that has been extensively employed is the conjugate prior. Mathematical simplicity of making analytical evaluation for inference is the most important advantage of the conjugate prior. Unfortunately, the resulting estimator is not robust. When the robustness is of concern, an attractive way to develop robust Bayesian inference is to use robust priors which possess flat but not too flat tails to form Bayesian estimators, see Bian (1995), Bian and Tiku (1997), Dickey (1974), Ramsay and Novick (1980), Berger (1980, 1984) and Press (1989). However, it is difficult to make analytical evaluation for inference because of the ugly forms of the resulting posterior densities. Bian and Dickey (1996) overcame this difficulty by introducing a prior in which the prior knowledge regarding β and σ^2 is assumed to be independently distributed as a Cauchy g -prior and a reciprocal gamma distribution such that

$$p(\beta, \sigma^{-2}) \propto f(\beta|g)h(\sigma^{-2})$$

where

$$f(\beta|g) \propto [1 + g(\beta - \beta_0)'X'X(\beta - \beta_0)]^{-(p+1)/2}$$

and

$$h(\sigma^{-2}) \propto (\sigma^{-2})^{\nu_0/2-1} \exp[-(\nu_0 s_0^2)/2\sigma^2]$$

This prior distribution has the same marginal density as the conjugate prior specified in (3) when $\nu_0 = 1$. Combining this prior with the likelihood function, the posterior density of β and σ^{-2} is:

$$f(\beta, \sigma^{-2}|y) \propto (\sigma^{-2})^{(n+\nu_0)/2-1} [1 + g\|(X'X)^{1/2}(\beta - \beta_0)\|^2]^{-(p+1)/2} \\ \times \exp\{-[\|y - X\hat{\beta}\|^2 + \|(X'X)^{1/2}(\beta - \beta_0)\|^2 + \nu_0 s_0^2]/2\sigma^2\}$$

Integrating out σ^{-2} yields the posterior density of β ,

$$f(\beta|y; \beta_0, g_2) \propto [1 + g_2\|(X'X)^{1/2}(\beta - \beta_0)\|^2]^{-(p+1)/2} \\ \times [\|y - X\hat{\beta}\|^2 + \|(X'X)^{1/2}(\beta - \hat{\beta})\|^2 + \nu_0 s_0^2]^{-(n+\nu_0)/2}. \quad (6)$$

When $\nu_0 = p + 1 - n$, and $s_0 = 0$, the marginal posterior density of β is a poly-Cauchy density

$$f(\beta|y) \propto [1 + g\|(X'X)^{1/2}(\beta - \beta_0)\|^2]^{-(p+1)/2} \\ \times [\|y - X\hat{\beta}\|^2 + \|(X'X)^{1/2}(\beta - \beta_0)\|^2]^{-(p+1)/2}.$$

The robust Bayesian estimator, $\hat{\beta}_C$ of β , under quadratic loss, is the posterior mean:

$$\hat{\beta}_C = E(\beta|y, g) = w\hat{\beta} + (1 - w)\beta_0 \quad (7)$$

with mean

$$E(\hat{\beta}_C) = E(w)\beta + [1 - E(w)]\beta_0$$

and the variance-covariance matrix

$$Var(\hat{\beta}_C) = E(w^2)Var(\hat{\beta}) + Var(w)(\beta - \beta_0)(\beta - \beta_0)'$$

where $w = (1 + g^{1/2}\|y - X\hat{\beta}\|)^{-1}$.

The new estimator $\hat{\beta}_C$ is a non-linear function of $\hat{\beta}$. The weight w in (7) is a decreasing function of the prior parameter g and the residual $\|y - X\hat{\beta}\|$. When g goes to zero, the prior of β diffuses to the non-informative prior and the weight w increases to 1. In this situation, $\hat{\beta}_C$ approaches to the LSE $\hat{\beta}$ which is the Bayes estimator arising from the usual non-informative prior.

The main attraction of $\hat{\beta}_C$ is its weight w depending reasonably on the residual $\|y - X\hat{\beta}\|$. When wild or extreme observations occur, the value of residual $\|y - X\hat{\beta}\|$ rises. Hence, there is higher uncertainty for the LSE $\hat{\beta}$ and consequently the weight w becomes smaller. In this sense, the estimator $\hat{\beta}_C$ is an adaptive weighted average and tends to be considerably more robust.

To compare the robust Bayesian Estimator $\hat{\beta}_C$ in (7) with the least squares estimator $\hat{\beta}$ in (2) and the usual Bayesian Estimator $\hat{\beta}_N$ in (5), Bian and Dickey (1996) simulated the simple regression model

$$y = a + bx + e$$

in which the random term e is distributed as the ε -contaminated normal distribution such that

$$e \sim (1 - \varepsilon)N(0, 1) + \varepsilon N(0, k^2), \quad k = 5 \quad (8)$$

or the ε -contaminated Cauchy distribution such that

$$e \sim (1 - \varepsilon)N(0, 1) + \varepsilon \text{ Standard Cauchy} \quad (9)$$

for small samples.

In their simulation results, they found that the efficiency of $\hat{\beta}_C$ relative to both $\hat{\beta}_N$ and $\hat{\beta}$ grows rapidly as ε grows in value. The higher the proportion of the observations contaminated by large fluctuations, the more efficient is $\hat{\beta}_C$ relative to both $\hat{\beta}_N$ and $\hat{\beta}$ in the simulation with the random terms in (8) and (9).

When the error terms are distributed as the ε -contaminated Cauchy distribution, the means and variances of $\hat{\beta}$ and $\hat{\beta}_N$ do not exist theoretically while the means and variances of $\hat{\beta}_C$ do exist and $\hat{\beta}_C$ is unbiased if the prior center β_0 hits the true value of β perfectly. In the simulation of this situation, they found that the efficiency of $\hat{\beta}_C$ relative to both $\hat{\beta}_N$ and $\hat{\beta}$ is extremely large even if the value of ε is as small as 0.01. This shows that $\hat{\beta}_C$ is considerably robust relative to both $\hat{\beta}_N$ and $\hat{\beta}$.

3. The application of the robust Bayesian estimator in CAPM

The Capital Asset Pricing Model is a parsimonious general equilibrium model developed by Sharpe (1963, 1964), Treynor (1961) and Lintner (1965). They suggested that the excess return R on a security is formulated by:

$$R = a + bR_m + e \quad (10)$$

where R_m is the excess return on market portfolio, and e is the random error. From Equations (1) and (10), we have $\beta = (a, b)'$. In this paper, we do not consider the Black version of CAPM which treats the zero-beta portfolio return as an unobserved quantity, making the analysis more complicated than that of the Sharpe-Lintner version.

Blume (1975), Brenner (1974), Pettit and Westerfield (1974), Leavy (1971), Hamada (1972) and many others found that the measure of security risk is empirically non-stationary over time. To handle the non-stationarity of β , Bodurtha and Nelson (1991) applied the conditional heteroskedastic error using autoregressive conditional heteroskedastic model (ARCH) for the estimation of the CAPM.

In order to capture the stationary Beta parameter, one may estimate the model from a reasonably short subperiod. In this situation, the Bayesian approach is a good choice. Vasicek (1973) is one of the earliest papers that discusses the application of Bayesian estimation to the CAPM. However, Vasicek's approach is not robust.

Many papers such as Fama (1963, 1965a, 1965b) analyzed the empirical data and concluded that the distribution of the security or portfolio return is 'flat-tail' and the normality assumption is violated. They suggested the family of stable Paretian distributions between normal and Cauchy distributions for the stock returns.

Blattberg and Gonedes (1974) examined the security returns and suggested student- t as an alternative 'flat-tail' distribution. Clark (1973), Christie (1983), Kon (1984), and Tse (1991) suggested a mixture of normal distributions for the stock return while Fielitz and Rozelle (1983) suggested that a mixture of non-normal stable distributions would be a better representation of the distribution of security and portfolio return.

The structure of the distribution for the return may carry over into the structure of the disturbance. As such, the disturbance's distribution is 'flat-tailed' and the mixture of normal distributions or the mixture of normal and Cauchy distributions may give a better description of the distribution of the disturbance in the CAPM. This is supported by Harvey and Zhou (1993) who pointed out that the non-normality in the return may carry over into non-normality of the disturbance in the CAPM. Harvey and Zhou examined the residuals of the world market portfolios in the CAPM and found that in many cases the distributions of the disturbances departed from normality. They then tested the sensitivity of the benchmark in the CAPM by specifying error structures that were t -distributed or follow the mixtures of normal distributions.

Based on the findings by Bian and Dickey (1996), we hypothesize that the robust Bayesian estimator is more appropriate in the estimation of the parameters of the CAPM in the sense that it is more efficient than the LSE in the estimation of the parameters.

To support our hypothesis, the first consideration pertains to the robustness of the estimator. Mandelbrot (1963) and others show empirically that the distribution of the return is intermediate between normal and Cauchy distributions and therefore the tails of the distribution are flatter than normal but thinner than Cauchy. Since the Cauchy prior distribution is a robust prior with tails very much flatter than the normal distribution, the robust Bayesian estimator $\hat{\beta}_C$ arising from a Cauchy type g -prior with normal-distributed sample distribution performs well in yielding an estimator which is robust with respect to wild fluctuations and extreme observations of the stock return in the CAPM. The estimator is highly representative in the situation in which the distribution is between normal and Cauchy.

The next consideration concerns the sampling distribution following the mixture of normal distributions, as found by Brenner (1974), Boness, Chen and Jatusipitak (1974), Kon (1984) and Tse (1990). The simulation results in Bian and Dickey (1996) showed that $\hat{\beta}_C$ is more efficient than both $\hat{\beta}_N$ and $\hat{\beta}$ under the mixture of normal distributions for the error term. This suggests that our approach should provide a better estimation for the CAPM with respect to the issue of the mixture of normal distributions.

The last consideration refers to the mixture of normal and Cauchy distributions. Fielitz and Rozelle (1983) found that the distribution of some security returns fitted the mixture of normal and non-normal stable distributions with different characteristic exponents. Bian and Dickey (1996) have already demonstrated that $\hat{\beta}_C$ is more efficient in the case of the mixture of normal and Cauchy distributions. This suggests our hypothesis is justified for the issue of the mixture of normal and Cauchy distributions.

4. Simulation results

From the practical point of view, it is very important to examine the sensitivity of a statistical procedure to deviations from an assumed model. We thus evaluate, in the traditional sense, the performance under departures from the assumed model of the $\hat{\beta}_C$ relative to the $\hat{\beta}$ based on the CAPM model (10) with $\beta_0 = (a, b) = (0, 1)$ and e_i 's being distributed as ε -contaminated distributions, as displayed in Table 1.

We then compare $\hat{\beta}_C$ with $g = 0.01$ to $\hat{\beta}$. The values of the mean error (bias) and MSE of these three estimators for different error distributions and different locations of the prior center $\beta_0 = (a_0, b_0)$ are evaluated based on 10,000 runs. The results are tabulated in Table 1. For convenience, we define the bias and MSE as follows:

$$\text{bias}(\hat{\beta}, \beta) = \|E(\hat{\beta}) - \beta\| \quad \text{and} \quad \text{MSE}(\hat{\beta}) = E(\|\hat{\beta} - \beta\|^2). \quad (11)$$

We note that, usually, the bias is defined as $E(\hat{\beta}) - \beta$ when the dimension of β is equal to one. However, we define the bias as in (11) because in our model the dimension of β is greater than 1.

Based on the tabulated values, we obtain the following findings:

1. The $\hat{\beta}_C$ has no bias when the prior center hits the true values of β perfectly, and has negligible biases when the prior center deviates moderately.
2. $\hat{\beta}_C$ is uniformly superior to $\hat{\beta}$ when the model is contaminated by non-normal disturbances or when the prior center hits the true location of β perfectly.
3. $\hat{\beta}_C$ is remarkably superior to $\hat{\beta}$ when the model is contaminated by Cauchy disturbances. Relative efficiency is from 8330.88 to as large as 18433.48 in our simulation.

Cauchy disturbances cause damage to LSE $\hat{\beta}$. When Cauchy errors occur in observations, the sampling means and sampling variances of $\hat{\beta}$ do not exist. Hence the values of LSE fluctuate violently and therefore the values of the MSE for $\hat{\beta}$ shown in Table 1 are very large. Thus one may conclude that LSE is inappropriate when some or all of the errors follow a Cauchy distribution. On the contrary, the values of both bias and MSE for $\hat{\beta}_C$ are quite small. In addition, the sampling mean and sampling variance of $\hat{\beta}_C$ do exist. Hence, $\hat{\beta}_C$ is highly robust relative to $\hat{\beta}$. At least, it can be viewed as a promising alternative method in a number of CAPM model where the error terms are distributed as mixture distributions.

We note that in Table 1 the MSE for the bayesian estimate with $\beta_0 = (0, 1)$ is less than the MSE for the LSE in the $N(0, 1)$ case. This is because the prior $\beta_0 = (0, 1)$ hits the exact value of the parameters in the model. When $\beta_0 = (-2, 3)$, reasonably far away from the parameters, the MSE for the proposed bayesian estimator is greater than the MSE for the LSE. When $\beta_0 = (-1, 2)$, it is close to the true value and hence the MSE is smaller than the MSE for the situation with $\beta_0 = (-2, 3)$.

We also note that in Table 1 the MSE for the LSE in $.75N(0, 1) + .25C(0, 1)$ is less than the LSE in $.90N(0, 1) + .10C(0, 1)$. This is possible because the variance of $C(0, 1)$ does not exist and hence the MSE has huge variability and depends on the samples chosen.

5. Empirical Study

In this section, we demonstrate that the robust Bayesian estimator is a more appropriate estimator of the parameters in the CAPM by examining the US annual and monthly stock returns.

Twelve industrial portfolios of U.S. data are employed in the study. The industry classifications conform to Sharpe (1982), Breeden, Gibbons and Litzenberger (1989) and Gibbons, Ross, and Shanken (1989). The portfolios are value-weighted. The monthly market return is the value weighted NYSE return. The portfolio returns are available from the Center for Research in Security Prices (CRSP) at

the University of Chicago. These monthly returns from the period 1926-1987 are in excess of 30-day Treasury-bill rate available from Ibbotson Associates. Harvey and Zhou (1990) introduced a Bayesian test and calculated posterior odds ratios for the industry portfolios of these returns to test the mean-variance efficiency. We use the same data set to demonstrate that the robust Bayesian estimator is a more appropriate approach in the CAPM estimation.

Table 1. Values of biases and MSE's for the estimators with $\beta = (0, 1)$

Sample of Error Terms	LSE		Proposed Bayesian Estimator					
	Bias	MSE	$\beta_0 = (0, 1)$		$\beta_0 = (-1, 2)$		$\beta_0 = (-2, 3)$	
			Bias	MSE	Bias	MSE	Bias	MSE
$N(0, 1)$	0.000	.479	.000	.298	.300	.392	.600	.673
$.95N(0, 1) + .05N(0, 9)$	0.011	.669	.007	.369	.326	.481	.657	.827
$.90N(0, 1) + .10N(0, 9)$	0.007	.837	.006	.431	.361	.570	.717	.983
$.75N(0, 1) + .25N(0, 9)$	0.012	1.426	.008	.627	.435	.830	.864	1.429
$.95N(0, 1) + .05(3T_4)^*$	0.002	.969	.001	.428	.344	.559	.688	.950
$.90N(0, 1) + .10(3T_4)$	0.002	1.291	.002	.502	.375	.663	.752	1.142
$.75N(0, 1) + .25(3T_4)$	0.030	2.574	.020	.807	.457	1.040	.929	1.770
$3T_4$	0.023	8.556	.018	1.809	.714	2.334	1.430	3.925
$.95N(0, 1) + .05C(0, 1)$	0.065	108.2	.007	.428	.342	.567	.679	.975
$.90N(0, 1) + .10C(0, 1)$	0.414	812.5	.008	.550	.374	.726	.741	1.238
$.75N(0, 1) + .25C(0, 1)$	0.026	445.6	.018	.883	.442	1.146	.898	1.951
$C(0, 1)$	0.472	40738	.003	2.208	.764	2.880	1.530	4.888

$N(0, 1)$ denotes a normal distribution with mean 0 and variance 1, $3T_4$ denotes a scaled Student t distribution with 4 degree of freedom with a scale of 3, and $C(0, 1)$ denotes a Cauchy distribution with mean 0 with a scale of 1.

We specify the CAPM for the excess return R_i for the i th industrial classification portfolio such that:

$$R_i = a_i + b_i R_m + e_i \quad \text{for } i = 1, 2, \dots, 12$$

where R_m is the market excess return, and e_i is the error term of the i th industrial portfolio.

We first apply the normality test concerning the measures of skewness and kurtosis for the returns and the corresponding residuals in the CAPM to test the hypothesis that the returns R_i are normally distributed and to test the hypothesis that the disturbances e_i come from a normal distribution. The results are shown in the following tables:

The results in Table 2 lead us to reject the hypothesis that the monthly returns R_i as well as their corresponding disturbances come from a normal distribution at 0.01 level of significance. The above finding supports the hypothesis that the

Table 2. Tests for departure from normality for monthly excess portfolio returns and the corresponding residuals in CAPM by industrial classifications.

Portfolio	Returns		Residuals	
	Skewness	kurtosis	Skewness	kurtosis
NYSE value-weighted	0.3059**	10.6030**	—	—
Petroleum	0.3103**	7.4277**	0.2477**	4.1315**
Finance & Real Estate	0.2257**	10.6255**	0.0060	4.7600**
Consumer Durables	1.0134**	15.3646**	0.6193**	10.7926**
Basic Industries	0.8691**	13.6209**	0.6333**	9.6177**
Food & Tobacco	0.0178	10.1611**	-0.1866*	4.9496**
Construction	0.8995**	11.5376**	0.5306**	6.6211**
Capital Goods	0.2375**	9.0959**	0.1785*	4.7571**
Transportation	1.1614**	15.2275**	1.1199**	8.7320**
Utilities	0.1446	10.7665**	-0.0405	5.0824**
Textile & Trade	0.1218	8.6145**	-0.0940	4.8637**
Services	0.0349	7.0560**	0.3336**	11.8533**
Recreation	0.2925**	9.1474**	-0.4153**	5.3689**

* $p < .05$

** $p < .01$

non-normality in the returns will carry over into the non-normality of the disturbances in the CAPM, as mentioned in Harvey and Zhou (1993). However, Table 3 leads us to accept the normality hypothesis for the annual returns of all portfolios except Construction and Basic Industries at 0.01 level of significance but reject the normality hypothesis for their corresponding disturbances in some cases. This suggests that the normality in the return may not carry over into the normality of the disturbance. In the situation where the disturbance is normally distributed or non-normally distributed for the U.S. portfolio return, we apply both $\hat{\beta}_C$ and $\hat{\beta}$ to study the efficiency of estimation in the CAPM. We note that the return may process possess ARCH effects which may cause the return to depart from normal. However, temporal aggregation will reduce this ARCH effects; for examples, see Drost and Nijman (1995). Hence, the annual excess portfolio returns are closer to normal as compared to the monthly excess portfolio returns, see Tables 2 and 3 respectively.

Since the Bayesian estimation involves subjective judgement, we have to specify the values of the prior parameters. Ideally, the specification of the hyper-parameters should be obtained from experts with thorough knowledge in the market. The expert opinion may come from the detailed information of the fundamentals such as corporate profitability, capital structure and leverage, and from confidential and restricted information such as the latest preliminary corporate accounts and investment plans.

Table 3. Tests for departure from normality for annual excess portfolio returns and the corresponding residuals in CAPM by industrial classifications.

Portfolio	Returns		Residuals	
	Skewness	kurtosis	Skewness	kurtosis
NYSE value-weighted	0.1361	3.0145	—	—
Petroleum	-0.1036	3.6989	-0.2063	3.0292
Finance & Real Estate	0.3696	3.0602	-0.2794	3.7495
Consumer Durables	-0.4871	3.8769	-0.5831*	4.4674*
Basic Industries	-0.0819	5.4423**	-1.4005**	8.9522**
Food & Tobacco	-0.0129	2.8486	-0.2207	2.1284*
Construction	-1.3041**	8.1547**	-2.1054**	15.4045**
Capital Goods	-0.1921	3.5966	-0.7956**	3.6278
Transportation	0.0455	3.3221	-0.1741	2.6903
Utilities	-0.5137*	4.5002*	0.2611	6.8466**
Textile & Trade	0.0230	2.4575	-0.4527	3.9870*
Services	0.0755	3.0064	0.4039	4.7447**
Recreation	-0.0037	3.0391	-0.3682	3.5089

* $p < .05$

** $p < .01$

However, sometimes there is not enough information for statisticians or financial analysts to specify the values of the prior parameters. In this situation, we may acquire the information of the previous corresponding month to specify the information for the prior.

There are two prior parameters in the robust Bayesian estimator $\hat{\beta}_C$: β_0 and g . The parameter β_0 is the prior centre of β while g is prior precision of β . In this study, we use the estimate of β from the previous sample with the same sample size as the value of β_0 in the updated estimation. This is essentially an empirical Bayes approach (see Maritz and Lwin 1989).

We adopt the one-step ahead forecast MSE, see Clements and Hendry (1997) for more detail, as a basis for comparison between $\hat{\beta}_C$ and $\hat{\beta}$ for the U.S. monthly and annual data. In the computation, the sample size n is chosen from 6 to 36 for monthly data and from 5 to 20 for annual data. The value of g is chosen from 0.1 to 20. We note that the first n data ($t = 1, \dots, n$) are being used only to compute the prior information for $\hat{\beta}_C$ in the first sample ($t = n + 1, \dots, 2n$). The second n data ($t = 2, \dots, n + 1$) are being used only to compute the prior information for $\hat{\beta}_C$ in the second sample ($t = n + 2, \dots, 2n + 1$), and so on.

For each sample size n and for each g value, the estimates of both $\hat{\beta}_C$ and $\hat{\beta}$ are first computed for each industrial portfolio for $t = n + 1, \dots, T - 1$ where T is December 1987 for monthly data and 1987 for annual data. We then compute their one-step ahead forecasts, \hat{R}_{it} , by applying $\hat{\beta}_C$ and $\hat{\beta}$ respectively for each portfolio

and for $t = 2n + 1, \dots, T$ and subsequently the one-step ahead forecast MSE,

$$\frac{\sqrt{\sum_{t=2n+1}^T (\hat{R}_{it} - R_{it})^2}}{T - 2n}$$

for the i^{th} Portfolio with respect to $\hat{\beta}_C$ and $\hat{\beta}$ for each n and each g . The average one-step ahead forecast MSE

$$\frac{\sum_{i=1}^{12} \sqrt{\sum_{t=2n+1}^T (\hat{R}_{it} - R_{it})^2}}{12(T - 2n)}$$

with respect to $\hat{\beta}_C$ and the average with respect to $\hat{\beta}$ are then computed for each g and each n . Their relative efficiency

$$\frac{\text{average one-step ahead of forecast MSE of } \hat{\beta}}{\text{average one-step ahead of forecast MSE of } \hat{\beta}_C}$$

is then computed for each g value and sample size n .

In our empirical study, we find that $\hat{\beta}_C$ is uniformly more efficient than $\hat{\beta}$ in the sense of the relative efficiency of one-step ahead forecast mean square error for any sample size and for any g value. For simplicity, we only present the average relative efficiency for $g = 0, 0.1, 0.5, 1, 2, 5, 10, 15$ and 20 and sample size from 6 to 36 with an increment of 6 for monthly data and from 5 to 20 with an increment of 5 for annual data. The results of the average one-step ahead forecast MSE obtained by applying $\hat{\beta}_C$ in the CAPM for monthly and annual US returns are in Table 4 and Table 6 respectively. We note that the values in the tables are 1000 times the original values and the average one-step ahead forecast MSE with respect to $\hat{\beta}_C$ is equal to that of $\hat{\beta}$ when $g = 0$. The results of the efficiency of $\hat{\beta}_C$ relative to $\hat{\beta}$ for monthly and annual US stock returns are in Table 5 and Table 7 respectively.

Table 4. Average one-step ahead forecast MSE obtained by applying $\hat{\beta}_C$ for monthly US stock returns ($\times 1000$)

Sample Size	g value								
	0	0.1	0.5	1	2	5	10	15	20
6	1.298	1.279	1.258	1.245	1.229	1.203	1.181	1.168	1.159
12	1.032	1.019	1.007	0.999	0.992	0.981	0.974	0.971	0.969
18	0.963	0.952	0.942	0.937	0.932	0.925	0.923	0.922	0.923
24	0.927	0.918	0.911	0.907	0.904	0.901	0.901	0.902	0.903
30	0.899	0.892	0.888	0.886	0.885	0.885	0.888	0.890	0.891
36	0.871	0.866	0.863	0.862	0.861	0.863	0.866	0.868	0.870

From the results in these tables, we find that the estimate of $\hat{\beta}_C$ is more efficient than that of $\hat{\beta}$ for any g value and for any sample size n in our study especially

Table 5. Percentage average relative efficiency of $\hat{\beta}_C$ to $\hat{\beta}$ for the monthly US stock returns.

Sample Size	g value							
	0.1	0.5	1	2	5	10	15	20
6	101.53	103.18	104.27	105.65	107.91	109.91	111.14	112.02
12	101.25	102.46	103.20	104.03	105.18	105.96	106.30	106.47
18	101.16	102.20	102.77	103.37	104.06	104.37	104.41	104.37
24	100.96	101.74	102.13	102.50	102.83	102.85	102.74	102.60
30	100.70	101.20	101.40	101.54	101.49	101.22	100.94	100.69
36	100.59	100.96	101.09	101.14	100.97	100.62	100.30	100.02

Table 6. Average one-step ahead forecast MSE obtained by applying $\hat{\beta}_C$ for annual US stock returns ($\times 1000$)

Sample Size	g value								
	0	0.1	0.5	1	2	5	10	15	20
5	1.858	1.781	1.715	1.680	1.644	1.604	1.586	1.583	1.584
10	1.610	1.542	1.488	1.461	1.435	1.408	1.396	1.393	1.393
15	1.475	1.437	1.412	1.402	1.394	1.389	1.392	1.395	1.398
20	1.612	1.557	1.516	1.497	1.479	1.458	1.447	1.442	1.439

Table 7. Percentage average relative efficiency of $\hat{\beta}_C$ to $\hat{\beta}$ for the annual US stock returns.

Sample Size	g value							
	0.1	0.5	1	2	5	10	15	20
5	104.35	108.36	110.63	113.02	115.83	117.14	117.41	117.33
10	104.47	108.26	110.26	112.24	114.42	115.38	115.59	115.58
15	102.64	104.44	105.21	105.81	106.15	105.98	105.73	105.48
20	103.57	106.31	107.67	109.01	110.55	111.44	111.82	112.04

for small sample sizes. We note from Tables 2 and 3 that the annual returns can be assumed to be normally distributed in many cases while the monthly returns are not normally distributed. This suggests that $\hat{\beta}_C$ can also be applied for both normally distributed and non-normally distributed data. In both situations $\hat{\beta}_C$ is more efficient than $\hat{\beta}$ as illustrated in our study.

As shown in (7), g is the precision of the prior density of β . The larger the value of g , the less is the prior uncertainty about β ; and consequently, the estimate $\hat{\beta}_C$ puts heavier weight on the prior location. The results in Table 5 and Table 7 show that in general the relative efficiency is higher for greater g values and for smaller

sample sizes. This suggests that our choice of prior information is appropriate and the prior information contributes significantly in the estimation.

The results in Table 5 and Table 7 show that the relative efficiency is lower for large sample sizes. This implies that $\hat{\beta}_C$ is not much better than $\hat{\beta}$ for large sample sizes. Perhaps, it is because the portfolio of the US stock returns is not stable in time or it is because the estimate of $\hat{\beta}$ is sufficiently good enough. The results in the tables also show that the relative efficiency is lower for small g values. This makes sense because $\hat{\beta}_C$ tends to $\hat{\beta}$ when g tends to zero.

Table 5 shows that $\hat{\beta}_C$ is up to 12% more efficient than $\hat{\beta}$, while Table 7 shows that $\hat{\beta}_C$ is up to 17% more efficient than $\hat{\beta}$. These empirical results illustrate that $\hat{\beta}_C$ is uniformly better than $\hat{\beta}$ in the estimation of the parameters in the CAPM.

6. Conclusion

Bian and Dickey (1996) developed a robust Bayesian estimator for the vector of regression coefficients using a Cauchy-type g -prior. This estimator is an adaptive weighted average of the least squares estimator and prior location, and is of great robustness with respect to wild and extreme observations. In this paper, we apply the robust Bayesian estimator to financial regression models of stock returns in which the error is well-known to be ‘flat-tail’ distributed. To compare this estimator with the traditional least squares estimator, we apply both estimators to analyze the Capital Asset Pricing Model of the US annual and monthly stock returns. In our empirical study, we find that the robust Bayesian estimate is uniformly more efficient than the least squares estimate in terms of the relative efficiency of one-step ahead forecast mean square error, especially for small samples. Our study supports that the robust Bayesian estimator is more appropriate in the CAPM estimation.

The approach in our paper is based on regression modeling technique. One may apply the technique in Wong and Miller (1990) and Wong et al (1999) to investigate the fundamental component and the error component for each portfolio. One may also use the modified maximum likelihood estimation approach, see Tiku, et. al. 1999a,b,c and Tiku and Wong 1998, to relax the normality assumption on the CAPM.

Another possible area for further research is to compare the beta in this study with the equity cost of capital for each portfolio. For the estimation of the equity capital cost, for example see Thompson and Wong (1991, 1996). One may also apply the approach in this paper in studying the difference of the beta between risk averters and risk lovers, see Li and Wong (1999) and Wong and Li (1999).

Acknowledgments

We thank Professor Campbell Harvey for providing us with the US stock data and for his helpful comments. We also thank Kok-Phun Yap for assistance with the

calculations. Special thanks also to the editor and the referees for their valuable comments that have significantly improved this manuscript.

References

1. Berger, J.O., 1980, A robust generalized bayes estimator and confidence region for a multivariate normal mean, *Annals of Statistics*, 8, 716-761.
2. Berger, J.O., 1984, The robust Bayesian viewpoint, in studies in *Bayesian econometrics*, 4, edited by A. Zellner and J.B. Kadane, North-Holland, Amsterdam, 63-115.
3. Bian, G., 1995, Robust Bayesian estimators in a one-way ANOVA model, *Test*, 4, 115-135.
4. Bian, G. and J.M. Dickey, 1996, Properties of multivariate Cauchy and poly-Cauchy distributions with Bayesian g-prior applications. in *Bayesian Analysis in Statistics and Econometrics: Essay in Honor of Arnold Zellner*, edited by D.A. Berry, K.M. Chaloner and J.K. Geweke, John Wiley & Sons, New York, 299-310.
5. Bian, G. and M.L. Tiku, 1997, Bayesian inference based on robust priors and MML estimators: Part I, symmetric location-scale distributions, *Statistics*, 29, 317-345.
6. Blattberg, R.C. and N.J. Gonedes, 1974, A comparison of stable and student distribution as statistical models for stock prices, *Journal of Business*, 47, 244-280.
7. Bodurtha, Jr. and C.M. Nelson, 1991, Testing the CAPM with time-varying risks and returns, *Journal of Finance*, 46, 1485-1505.
8. Boness A., Chen A. and S. Jatusipitak, 1974, Investigations of nonstationarity in prices, *Journal of Business*, 47, 518-537.
9. Breeden, D.T., Gibbons M. and R.H. Litzenberger, 1989, Empirical tests of the consumption based on CAPM, *Journal of Finance*, 44, 231-262.
10. Blume, M.E., 1975, Betas and their regressions tendencies, *Journal of Finance*, 30, 785-795.
11. Campbell, J.Y., Lo A.W. and A.C. MacKinlay, 1997, *The econometrics of financial markets*, Princeton University Press, New Jersey.
12. Christie, A., 1983, On information arrival and hypothesis testing in event studies, Working paper, University of Rochester, 1983.
13. Clark, P.K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 37, 135-155.
14. Clements, M.P. and D.F. Hendry, 1997, An empirical study of seasonal unit roots in forecasting, *International Journal of Forecasting*, 13, 341-355.
15. Dickey, J.M., 1974, Bayesian alternatives to the F-test and least-squares estimate in the normal linear model, in *Studies in Bayesian econometrics and statistics*, 4, edited by S.E. Finberg and A. Zellner, North-Holland, Amsterdam, 515-554.
16. Drost, F.C. and T.E. Nijman, 1995, Temporal aggregation of GARCH processes, in R.F. Engle, ed. ARCH: Selected Readings, Oxford University Press.
17. Fama, E.F., 1963, Mandelbrot and the stable Paretian hypothesis, *Journal of Business*, 36, 420-429.
18. Fama, E.F., 1965a, The behaviour of stock market prices, *Journal of Business*, 38, 34-105.
19. Fama, E.F., 1965b, Portfolio analysis in a stable Paretian market, *Management Science*, 11, 401-419.
20. Fielitz, B.D. and J.P. Rozelle, 1983, Stable distributions and mixtures of distributions hypotheses for common stock return, *Journal of American Statistical Association*, 78, 28-36.
21. Gibbons, M.R., Ross, S.A. and J. Shanken, 1989, A test of efficiency of a given portfolio, *Econometrica*, 57, 1121-1152.
22. Harvey, M.C. and G. Zhou, 1990, Bayesian inference in asset pricing tests, *Journal of Financial Economics*, 26, 221-254.
23. Harvey, M.C. and G. Zhou, 1993, International asset pricing with alternative distributional specifications, *Journal of Empirical Finance*, 1, 107-131.

24. Hamada, R.S., 1972, The effects on the firm's capital structure on the systematic risk of common stocks, *Journal of Finance*, 27, 435-452.
25. Kandel, S., R. McCulloch and R. Stambaugh, 1995, Bayesian inference and portfolio efficiency, *Review of Financial Studies*, 8, 1-53.
26. Kon, S.J., 1984, Models of stock returns – a comparison, *Journal of Finance*, 39, 147-165.
27. Leavy, R.A., 1971, On the short term stationarity of Beta coefficient, *Financial Analysis Journal*, 27, 55-62.
28. Li, C.K. and W.K. Wong, 1999, A Note on Stochastic Dominance for Risk Averters and Risk Takers, *RAIRO Recherche Opérationnelle*, 33, 509-524
29. Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 47, 13-37.
30. MacKinlay, A.C. and M. Richardson, 1991, Using generalized methods of moments to test mean-variance efficiency, *Journal of Finance*, 46, 511-527
31. Maritz, J.S. and T. Lwin, 1989, *Empirical Bayes methods*, second edition, Chapman and Hall, New York.
32. Mandelbrot, B., 1963, The valuation of certain speculative prices, *Journal of Business*, 36, 394-419.
33. McCulloch, R. and P.E. Rossi, 1991, A Bayesian approach to testing arbitrage pricing theory, *Journal of Econometrics*, 49, 141-168.
34. Pettit, R.R. and R. Westerfield, 1974, Using the capital asset pricing model and returns, *Journal of Financial and Quantitative Analysis*, 9, 579-605.
35. Press, S.J., 1989, *Bayesian statistics: principles, models and applications*, John Wiley & Sons, New York.
36. Ramsay, J.O. and M.R. Novick, 1980, PLU robust Bayesian decision theory, *Journal of American Statistical Association*, 75, 901-907.
37. Shanken, J., 1987, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics*, 19, 195-216.
38. Sharpe, W., 1963, A simplified model for portfolio analysis, *Management Science*, 9, 277-293.
39. Sharpe, W., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, 19, 425-442.
40. Sharpe, W., 1982, Factors in New York stock exchange security returns, 1931-1979, *Journal of Portfolio Management*, 8, 5-19.
41. Tiku, M.L., and W.K. Wong, 1998, Testing for a unit root in an AR(1) model using three and four moment approximations: symmetric distributions, *Communications in Statistics: Simulation and Computation*, 27 (1), 185-198.
42. Tiku, M L, W.K. Wong, D.C. Vaughan and G. Bian, 1999a, Time series models with nonnormal innovations: symmetric location–scale distributions, *Journal of Time Series Analysis*, (submitted).
43. Tiku, M.L., W.K. Wong, and G. Bian, 1999b, Estimating parameters in autoregressive models in non-normal situations: symmetric Innvottions, *Communications in Statistics: Theory and Methods*, 28(2), 315-341.
44. Tiku, M.L., W.K. Wong, and G. Bian, 1999c, Time series models with asymmetric innovations, *Communications in Statistics: Theory and Methods*, 1999, 28(6), 1331-1360.
45. Treynor, J., 1961, Towards a theory of the market value of risky assets, Unpublished Manuscript.
46. Thompson H.E. and W.K. Wong, 1991, On the unavoidability of 'scientific' judgement in estimating the cost of capital, *Managerial and Decision Economics*, 12, 27-42.
47. Thompson H.E. and W.K. Wong, 1996, Revisiting 'Dividend Yield Plus Growth' and Its Applicability, *Engineering Economist*, Winter, Vol 41, No. 2, 123-147.
48. Tse, Y.K., 1991, Price and volume in the Tokyo stock exchange: an exploratory study, in Ziemba, W.T., Bailey, W. and Hamao, (ed), *Japanese Financial Market Research*, 91-119.
49. Vasicek, O.A., 1973, A note on using cross sectional information in Bayesian estimation of security betas, *Journal of Finance*, 28, 1233-1239.

50. Wong W.K. and C.K. LI, 1999, A note on convex stochastic dominance theory, *Economics Letters*, 62, 293-300.
51. Wong, W.K. and R.B. Miller, 1990, Analysis of ARIMA-Noise models with repeated time series, *Journal of Business and Economic Statistics*, 8, no. 2, 243-250.
52. Wong, W.K., R.B. Miller and K. Shrestha, 1999, Maximum Likelihood Estimation of ARMA Model with Error Processes for Replicated Observation, *Journal of Applied Statistical Science*, (forthcoming).
53. Zellner, A., 1986, On assessing prior distributions and Bayesian regression analysis with g -prior distribution, in *Bayesian inference and decisions techniques*, edited by P. Goel and A. Zellner, Elsevier Science Publishers B.V., North-Holland, Amsterdam, 233-243.