Thai Journal of Mathematics : (2016) 71–82
Special Issue on Applied Mathematics : Bayesian Econometrics
http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

Robust Regression for Capital Asset Pricing Model Using Bayesian Approach

K. Autchariyapanitkul[†,1], **K. Kunasri**[‡] **and S. Sriboonchitta**[§]

[†]Faculty of Economics, Maejo University
e-mail : kittawit_a@mju.ac.th
[‡]Faculty of Management Sciences, Chiang mai Rajabhat University
[§]Faculty of Economics, Chiang Mai University

Abstract : This study investigate the performance of a portfolio based on capital asset pricing model using a Bayesian statistics approach. We use a hierarchical model robustly to estimate the systematic risk of an asset. We assume that the returns follow independent normal distributions. MCMC sampling is applied to calculate all the parameters in the model. Finally, the Bayesian method gives us the probability of every possible asset returns, given the market returns and also the posterior predictions is a clue that the model could be improved.

Keywords : Bayesian Approach; CAPM; Gibb sampling; Prior Distribution; Robust Regression.
2010 Mathematics Subject Classification : 62P20; 91B84.

1 Introduction

We consider situations in financial economics such as predicting a stock return from the market return, namely, capital asset pricing model (CAPM), CAPM is a very well-known model to a measure of risk in financial analysis content. The estimated parameter reflects the sensitivity of asset returns to the market returns. A number of applications in financial issues are usually using CAPM as the based model, such as the works from [1–7]. In their papers, they used different methods to estimate parameter in CAPM model.

Thanks! This research was supported by the National Research Council of Thailand.
[1]Corresponding author.

In finance research, The CAPM parameter is typically estimated by the least squares method, which coincides with the maximum likelihood estimator under normality. However, such estimation method has at least two significant limitations. First, no available prior information on the CAPM parameters is used. And secondly, it is well known that outliers and gross errors can distort the beta estimate. The literature contains several references to problems such as the non-normality of returns and the lack of robustness of simple estimates (see, e.g., [8–10]. [11] suggest the use of student-t distributions as a robust alternative to normality, illustrating its use in multivariate analysis and regression. More recently, [12] discuss estimation of linear asset pricing models, including the CAPM, under elliptical distributions. [13] consider classical estimation of beta assuming independently student-t distributed share returns in the Chilean stock market. [14] study conditional skewness modeling for stock returns. An alternative formulation using quantile regression under non-normal errors has been described in [14].

In applications, the independent variable, let say, x. and the dependent variable, y values are provided by some real world process. In the real world process, there might or might not be any direct causal connection between x and y. It could be that x causes y, or y causes x, or some third factor causes both x and y, or x and y have no causal connection, or some combination of any or all of those. The simple linear model makes no claims about causal connections between x and y. But the simple linear model merely describes a tendency for y values to be linearly related to x values, hence "predictable" from the x values.

As a CAPM model, suppose we have measurements of stock returns and market return. The data appear to indicate that as market return increases, stock returns tends to increase. This covariation between stock returns and market return does not imply the one attribute causes the other.

The main reasons for the marginal of the Bayesian method including (1) conceptional violation to the use of prior distributions in precise and the biased procedures to probability in general, and (2) the lack of computing technology for doing realistic Bayesian analyses. Second, the growth in availability of panel data and the rise in the use of hierarchical modeling made the Bayesian method more engaging, because Bayesian statistics offers a natural approach to constructing hierarchical models. Third, there has been a growing recognition both that the enterprise of statistics is a subjective process in general and that the use of prior distributions need not influence results substantially. A prior can be determined from past information. Additionally, in many problems, the use of a prior distribution turns out to be advantageous.

The rest of this article is prepared as follows: Section 2 gives the mathematical proof of estimation of a normal likelihood, while section 3 shows the empirical application to stock market. Section 4 reports the empirical results, and final section gives conclusions.

2 Mathematical Proof of Distribution Function

2.1 A Normal Likelihood

The normal distribution indicates the probability density of a variable Y for a possible y, given the values of the mean μ and standard deviation σ are defined by

$$p(y|\mu,\sigma) = \frac{1}{Z}\exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) \tag{2.1}$$

where Z is the normalizer. i.e., a constant that makes the probability density integrate to 1. The probability density of any single datum, given parameter values, is $p(y|\mu,\sigma)$ as mentioned in Equation (2.1). The probability of all set of independent data values is $\prod_i p(y_i|\mu,\sigma) = p(Y|\mu,\sigma)$, where $Y = \{y_1, y_2, ..., yn\}$. Given a set of Y, we estimate the parameters with Bayes'rule, thus we have

$$p(\mu,\sigma|Y) = \frac{p(Y|\mu,\sigma)p(\mu,\sigma)}{\int\int \mathrm{d}\mu\mathrm{d}\sigma p(Y|\mu,\sigma)p(\mu,\sigma)} \tag{2.2}$$

It is acceptable to examine the case in which the standard deviation of the likelihood function is fixed at a particular value. In other hands, the prior distribution on σ is a spike over that exact value. We will express the value as $\sigma = S_y$. With this facilitate assumption, we are only estimating μ because we are concluding perfectly certain prior knowledge about σ

When σ is fixed, then the prior distribution on μ in Equation (2.2) can be easily chosen to be conjugate to the normal likelihood. It turns out that the product of normal distributions is again a normal distribution. Specifically, if the prior on μ is normal, then the posterior on μ is normal.

Let the prior distribution on μ be normal with mean X_μ and standard deviation S_μ. Then the likelihood times the prior is

$$\begin{aligned}
p(y|\mu,\sigma)p(\mu,\sigma) &= p(y|\mu,S_y)p(\mu)\\
&\propto \exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{S_y^2}\right)\exp\left(-\frac{1}{2}\frac{(\mu-X_\mu)^2}{S_\mu^2}\right)\\
&= \exp\left(-\frac{1}{2}\left[\frac{(y-\mu)^2}{S_y^2}+\frac{(\mu-X_\mu)^2}{S_\mu^2}\right]\right)\\
&= \exp\left(-\frac{1}{2}\left[\frac{S_\mu^2(y-\mu)^2+S_y^2(\mu-X_\mu)^2}{S_y^2S_\mu^2}\right]\right)\\
&= \exp\left(-\frac{1}{2}\left[\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2}\left(\mu^2-2\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}\mu+\frac{S_y^2X_\mu^2+S_\mu^2y^2}{S_y^2+S_\mu^2}\right)\right]\right)\\
&= \exp\left(-\frac{1}{2}\left[\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2}\left(\mu^2-2\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}\mu\right)\right]\right)\\
&\quad\times\exp\left(-\frac{1}{2}\left[\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2}\left(\frac{S_y^2X_\mu^2+S_\mu^2y^2}{S_y^2+S_\mu^2}\right)\right]\right)
\end{aligned}$$

$$\propto \exp\left(-\frac{1}{2}\left[\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2}\left(\mu^2-2\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}\mu\right)\right]\right) \tag{2.3}$$

where the transition to the last line was valid because the term that was dropped was merely a constant. Notice that the standard prior is also a quadratic expression in μ, from the Equation (2.3) we have

$$p(y|\mu,S_y)p(\mu) \propto = \exp\left(-\frac{1}{2}\left[\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2}\left(\mu^2-2\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}\mu+\left(\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}\right)^2\right)\right]\right)$$

$$= \exp\left(-\frac{1}{2}\left[\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2}\left(\mu-\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}\right)^2\right]\right) \tag{2.4}$$

Equation (2.4) is the numerator of Bayes'rule. When it is normalized by the evidence in the denominator, it becomes a probability density function. So, the mean is $\frac{S_y^2X_\mu+S_\mu^2y}{S_y^2+S_\mu^2}$ and the standard deviation is $\sqrt{\frac{S_y^2S_\mu^2}{S_y^2+S_\mu^2}}$. A wide distribution has low precision, because the posterior standard deviation is $\sqrt{\frac{S_y^2S_\mu^2}{S_y^2+S_\mu^2}}$. So, the posterior precision is

$$\frac{S_y^2+S_\mu^2}{S_y^2S_\mu^2} = \frac{1}{S_\mu^2}+\frac{1}{S_y^2} \tag{2.5}$$

Thus, the posterior mean can also rewrite in terms of precision as follow

$$\frac{1/S_\mu^2}{1/S_y^2+1/S_\mu^2}X_\mu+\frac{1/S_y^2}{1/S_y^2+1/S_\mu^2}y$$

From above equation, the posterior mean is a weighted average of the prior mean and the datum. Our purpose to determine what regression lines are most believable, given the information. We want to infer what combinations of β_0, β_1 and ϕ are most credible, given the data. We use Bayes'rule as follow

$$p(\beta_0,\beta_1,\phi|y) = p(y|\beta_0,\beta_1,\phi)p(\beta_0,\beta_1,\phi)\Big/\int\int\int d\beta_0 d\beta_1 d\phi\, p(y|\beta_0,\beta_1,\phi)p(\beta_0,\beta_1,\phi) \tag{2.6}$$

Diagnostical forms for the posterior can be obtained for appropriate priors.

2.2 Posterior Distribution

Given the prior, the posterior can be written as

$$\begin{aligned} p(\beta, \sigma^2|y, X) &\propto p(y|X, \beta, \sigma^2)p(\beta|\sigma^2)p(\sigma^2) \\ &\propto (\sigma^2)^{-n/2} \exp\Big(- \frac{1}{2\sigma^2}(y - X\beta)^{'}(y - X\beta)\Big) \\ &\times (\sigma^2)^{-k/2} \exp\Big(- \frac{1}{2\sigma^2}(\beta - \mu_0)^{'}\Lambda_0(\beta - \mu_0)\Big) \\ &\times (\sigma^2)^{(-a_0+1)} \exp\Big(- \frac{b_0}{\sigma^2}\Big) \end{aligned} \tag{2.7}$$

We can show the posterior mean μ_n in the format of least squares estimator $\hat{\beta}$ and prior mean μ_0 with the prior precision matrix Λ_n given by

$$\mu_n = (X^{'}X + \Lambda_0)^{-1}(X^{'}X\hat{\beta} + \Lambda_0\mu_0) \tag{2.8}$$

We can re-expressed this equation above to the quadratic format of $\beta - \mu_n$ as follows

$$(y-X\beta)^{'}(y-X\beta)+(\beta-\mu_0)^{'}\Lambda_0(\beta-\mu_0) = (\beta-\mu_0)^{'}(X^{'}X+\Lambda_0)(\beta-\mu_n)+y^{'}y-\mu_n^{'}(X^{'}X+\Lambda_0)\mu_n+\mu_0^{'}\Lambda_0\mu_0 \tag{2.9}$$

Given two densities of $N(\mu_n, \sigma^2\Lambda_n^{-1})$ and $Inverse-Gamma(a_n, b_n)$ distributions where the parameters are updated to the following formulas:

$$\begin{aligned} \mu_n &= (X^{'}X + \Lambda_0)^{-1}(\Lambda_0\mu_0 + X^{'}y) \\ \Lambda_n &= (X^{'}X + \Lambda_0) \\ a_n &= a_0 + \frac{n}{2} \\ b_n &= b_0 + \frac{1}{2}(y^{'}y + \mu_0^{'}\Lambda_0\mu_0 - \mu_n^{'}\Lambda_n\mu_n). \end{aligned}$$

3 Real World Applications

In this section, we show that the robust regression using Bayesian method is a more appropriate estimator of the parameters in the CAPM by investigating several shares in the US stock market.

3.1 Capital Asset Pricing Model

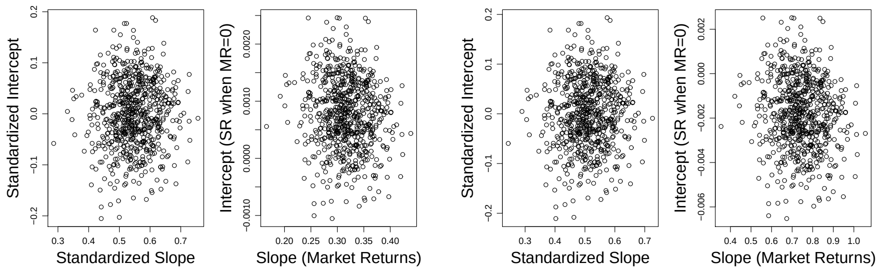
The Capital Asset Pricing Model (CAPM) was introduced by Sharpe (,1963,1964), Lintner (1965) and Mossin (1966), This model shows that the expected return on interested security (R_a) is correlated to the excess return on the market (R_m), the excess returns are investment returns from a security or portfolio that exceed the riskless rate on a security generally perceived to be risk free, such as a certificate of deposit or a government−issued bond.

$$R_a = \beta_0 + \beta_1 R_m + \epsilon, \tag{3.1}$$

where R_a is the excess return on asset, R_m is the excess return on market, and ϵ is the random error component.

3.2 Standardize Process for MCMC Sampling

The left and right panel of Figure 1 exhibits a correlation in believable values. The plausible slopes and intercepts are correlated for both MSFT and WMT. Sampling from such a tightly correlated distribution is typically tough to do directly. It is hard to discover a point in the narrow zone in the first place. Then, having created a viable point, the chain does not move efficiently. Metropolis algorithms often are not intelligent enough to automatically tune a proposal distribution to match a diagonal posterior.

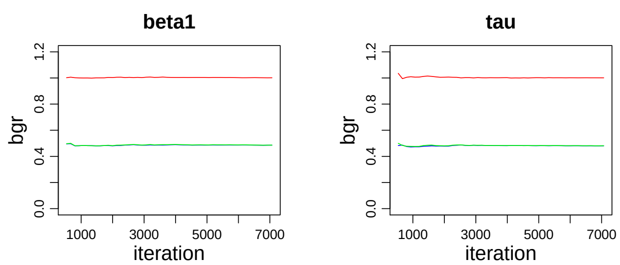


(a) MSFT (b) WMT

Figure 1: Believable regression lines, instandardized and raw scales.

3.3 MCMC Burn-in and Thinning

When using a Markov chain to generate a Monte Carlo sample from a distribution, we want to be sure that the resulting chain is a genuinely representative sample from the distribution. There are several ways in which the chain might fail to be representative. Fortunately, the conditional posterior distributions for the regression parameters and the error variance parameter are well known, and so Gibbs sampling provides a more efficient alternative.



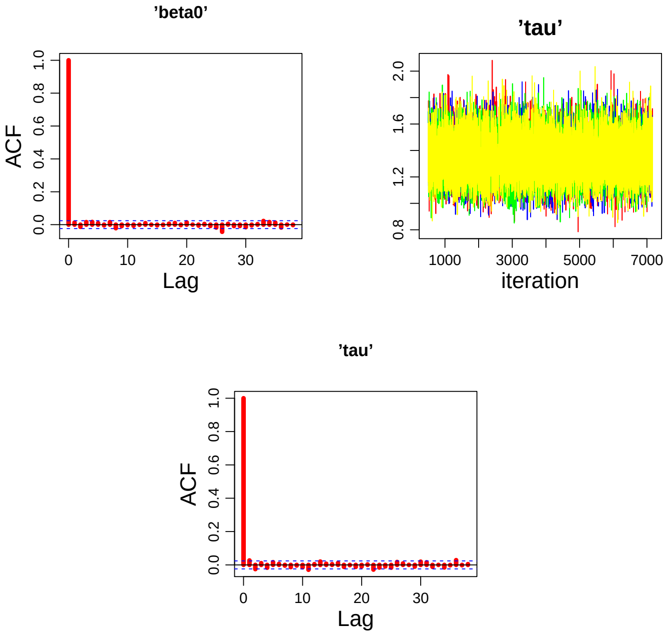


Figure 2: Burn-in and thinning on MCMC chains with 7,000 steps and thinning to 1 every 500 steps for WMT

The top of Figure 2 shows the bgr statistic, which measures how well mixed the chains are, over a limited length of the chains. As mentioned above, the bgr should be around 1.0. The plots show the bgr as a curve hovering near a value of 1.0. The two other curves, lower in graphs show the between chain and within-chain variances. In this particular application, the chains converge quickly and well. Figure 2 shows the chains after 10,000 burn-in steps (which is overkill for this particular application). The chains appear to be converged, without systematically increasing or decreasing. The bottom row of panels in Figure 2 shows the autocorrelation function, ACF. The left and middle columns show results when there is no thinning of the chains, and you can see that the ACF remains high for lags up to 15 or 20 steps. The autocorrelation is visible in the chains themselves as sustained plateaus during which the values from step to step barely change. The right column showed results when the chains were thinned, keeping only 1 step in every 500. You can see that after even one of the thinned steps, ACF(L=1) is nearly zero.

3.4 Initializing the Chains

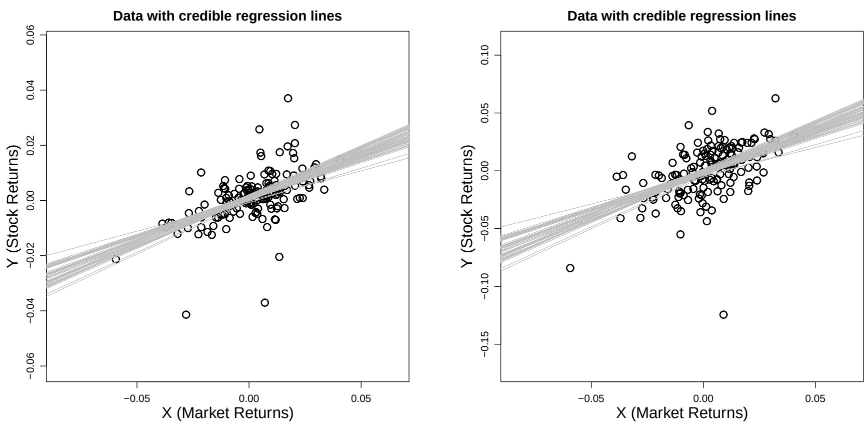
For the MCMC chain to randomly sample from the posterior, the random walk mus first get into the modal region of the posterior at the beginning. We may only start the chain at any point in parameter space, randomly selected from the prior distribution, and wait through the burn-in period until the chain randomly wanders into the bulk of the posterior.

3.5 Data and Parameter Estimation

In this study, we used the stock returns in $S\&P500$. We applied this procedure to several stocks namely, Microsoft Corporation (MSFT)and Wal-Mart Stores Inc. (WMT). All the weekly data are extracted from Yahoo in the period of 2013 until 2015 with a total of 156 observations for each selected shares.

Table 1: Parameter Estimation

		β_0	β_1	ϕ
WMT	Mean	-0.0007	0.5120	1.3460
		(0.0698)	(0.0700)	(0.1553)
	Median	-0.1395	0.3748	1.062
MSFT	Mean	-0.0007	0.5475	1.4180
		(0.0680)	(0.0682)	(0.1636)
	Median	-0.1395	0.4139	1.1190



(a) MSFT (b) WMT

Figure 3: Data points with credible regression lines covered

Figure 3 indicates that as market return increases, stock performance also tends to grow for both of MSFT and WMT. But this covariation between market return and capital return does not imply that one attribute causes the other. Despite the lack of direct causal relationship, the two values do covary, and one can be predicted from the other. Now, we can see various believable lines that go through the scatter plots. Notice that if a line has a steep slope, its intercept must be small, but if a line has a lower slope, its intercept must be higher. Thus, there is a trade-off in slope and intercept for the believable lines.

3.6 Posterior Prediction

Bayesian statistic provides the probability of every possible value R_a given the covariate R_m and the historical data: $p(R_a|R_m, D)$. The distribution of R_a values has uncertainty stemming from the inherent noise ϕ and also from uncertainty in the estimated values of the regression coefficients. A simple way to get a good approximation of $p(R_m|R_a, D)$ is by generating random values of y for every step in the MCMC sample of credible parameter values. Thus, at any step in the chain, there are particular values of β_0, β_1 and ϕ, which we use to generate predicted representative values of R_a according to $R_a \sim N(\mu = \beta_0 + \beta_1 R_m, \sigma = 1/\sqrt{(\phi)}$

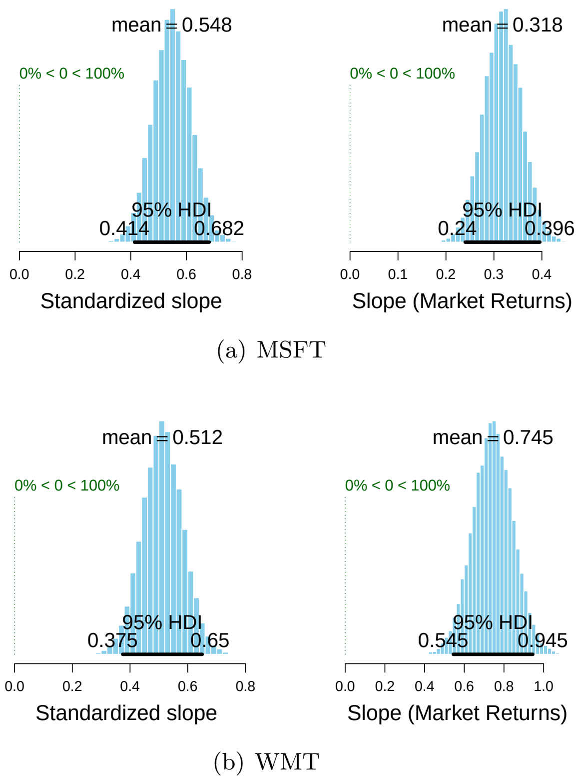
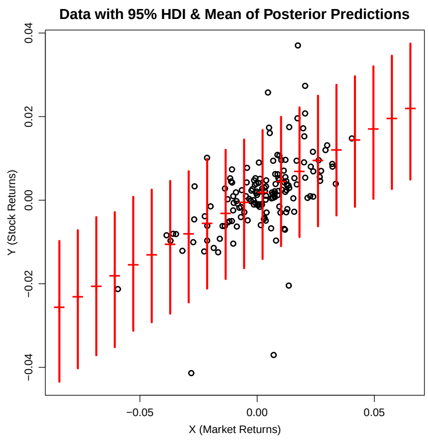
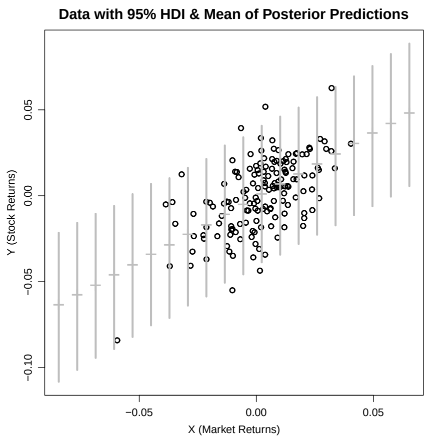


Figure 4: Posterior distribution of slopes.

Figure 4 exhibits the posterior distribution of slope values. The standardized and original scale slopes indicate the same relationship on different scales, and therefore the posterior distributions are identical except for a change of scale. The posterior distribution tells us about the believable slopes. We see that return on the market increases by 0.318 and 0.745 percentage for every 1 percentage growth in stock returns for MSFT and WMT, respectively.



(a) MSFT



(b) WMT

Figure 5: Bayesian Highest density interval (BHDI)

A summary of the posterior predictions is displayed in Figure 5 Each vertical red and gray segment, over particular market returns, shows the extent of the 95% BHDI for the distribution of randomly generated market returns at the market returns. The dash across the middle of the gray segment indicates the mean of the posterior predicted stocks returns. We see that the data do fall mostly within the range of the posterior predictions. However, the distribution of the data is systematically discrepant from the linear spine of the model: At high and low values of R_m, the data fall well above the long spine, but at traditional values of R_m, the data fall well below the straight spine. This sort of systematic discrepancy from the posterior predictions is a clue that the model could be improved.

4 Concluding Remarks

In this paper, we used the Bayesian approach to the capital asset pricing model by using a Gibbs samplers to calculate all the parameters in the model. The methodology in our study relies on regression modeling technique. The empirical study shows that the Bayesian estimator is more appropriate for a model prediction. Roughly speaking, a model is a good example if (1) the data fall mostly within the predicted zone, (2) the data are distributed approximately the way the model says they should be, e.g., regularly, and (3) the discrepancies of data from predictions are random and not systematic. Finally, the Bayesian method gives us the probability of every possible asset returns, given the market returns and also the posterior predictions is a clue that the model could be improved.

Acknowledgements : The authors thank Prof. Dr. Hung T. Nguyen for his helpful comments and suggestions.

References

[1] A.J. Patton, A. Timmermann, Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts, Journal of Financial Economics 98 (3) (2010) 605-625.

[2] H. Levy, The CAPM is alive and well: A review and synthesis, European Financial Management 16 (1) (2010) 43-71.

[3] P. Maio, Intertemporal CAPM with conditioning variables, Management Science 59 (1) (2013) 122-141.

[4] K. Autchariyapanitkul, S. Chanaim, S. Sriboonchitta, T. Denoeux, Predicting stock returns in the capital asset pricing model using quantile regression and belief functions, In International Conference on Belief Functions (2014) 219-226.

[5] K. Autchariyapanitkul, C. Somsak, S. Sriboonchitta, Quantile regression under asymmetric Laplace distribution in capital asset pricing model. Econometrics of Risk (2015) 219-231.

[6] S. Piamsuwannakit, K. Autchariyapanitkul, S. Sriboonchitta, R. Ouncharoen, Capital Asset Pricing Model with Interval Data, In International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (2015) 163-170.

[7] K. Autchariyapanitkul, S. Piamsuwannakit, S. Chanaim, S. Sriboonchitta, Optimizing Stock Returns Portfolio Using the Dependence Structure Between Capital Asset Pricing Models: A Vine Copula-Based Approach, In Causal Inference in Econometrics (2016) 319-331.

[8] E.F. Fama, The behavior of stock-market prices, The journal of Business 38 (1) (1965) 34-105.

[9] R.C. Blattberg, N.J. Gonedes, A comparison of the stable and student distributions as statistical models for stock prices, The journal of business 47 (2) (1974) 244-280.

[10] Z. Guofu, Assetpricing Tests under Alternative Distributions, The Journal of Finance 48 (5) (1993) 1927-1942.

[11] K.-L. Lange, R.-J.A. Little, J.-M.G. Taylor, Robust statistical modeling using the t distribution, Journal of the American Statistical Association 84 (408) (1989) 881-896.

[12] D.J. Hodgson, O. Linton, K. Vorkink. Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach, Journal of Applied Econometrics 17 (6) (2002) 617-639, International Journal of Approximate Reasoning 54 (2002) 793-808.

[13] D. Cademartori, C. Romo, R. Campos, M. Galea, Robust estimation of systematic risk using the t distribution in the chilean stock markets, Applied Economics Letters 10 (7) (2003) 447-453.

[14] K. Brännäs, N. Nordman, Conditional skewness modelling for stock returns, Applied Economics Letters 10 (11) (2003) 725-728.

(Received 2 August 2016)
(Accepted 20 October 2016)