jnaSci

An Application of Bayes Regression for a Small Sample

N. Abbasi* , M. Shokrizadah

Department of Statistics, Payame Noor University, I. R. Iran

Corresponding author Email : n_abbasi@pnu.ac.ir

ABSTRACT: Main purpose of this article is to have a posterior distribution of the model structural parameters in the regression model with elliptical measurement error based on prior distribution and considering the previous one was gained multivariable β it is not easy to produce data from it to measure the previous distribution mean. So we use Gibbs sampling method stated by Arellano and Bolfarine(1998). Finally we stated an application of the regression model.

Keywords: Elliptical distribution, Regression model with measurement error, Byes estimation, Gibbs algorithm

INTRODUCTION

In this article we examine one variable regression of response 'y' on the random variable ξ under observations of $\{(y_1, x_1), \dots, (y_n, x_n)\}$ which $x = (x_1, \dots, x_n)'$ measures the quantity $\xi = (\xi_1, \dots, \xi_n)'$ with the measurement error. Now we suppose that $\{(y_1, x_1), \dots, (y_n, x_n)\}$ observations come from a model with a measurement error shown by following equations :

$y = \beta_0 \mathbb{1}_n + \beta_1 \xi + \epsilon$

$x = \xi + u$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$, $y = (y_1, \dots, y_n)'$, $u = (u_1, \dots, u_n)'$ and $\mathbb{1}_n$ is component n of a vector as $\mathbb{1}_n = (1, \dots, 1)'$.

Having supposed that the measurement errors have been distributed normally Lindley & El-Sayyad (1968) examined the model. Also the state was discussed by many scientists. You may refer to Kendal and Stuart (1961), Johnston(1963) and Mandansky(1959) for more details. The essential supposition is that the measurement errors have elliptical distribution. Hereafter $El_n(\mu, \Sigma, h)$ is symbol of n-variable elliptical distribution with $\mu \epsilon \mathbb{R}^n$ mean and $n \times n$ dispersion matrix $\sum$, and density materials function $h^{(n)}$, if and only if its probable density is:

$$f(y|\mu, \Sigma, \gamma) = |\Sigma|^{-\frac{1}{2}} h^{(n)}\big(q(y)\big)$$

where $q(y) = (y - \mu)'|\Sigma|^{-1}(y - \mu)$ and $h^{(n)}$ should be nonnegative function and be right in the condition:

$$\int_0^\infty \pi^{\frac{n}{2}} u^{\frac{n}{2}-1} h^{(n)}(u)du = \Gamma\left(\frac{n}{2}\right).$$

Now, we suppose that the regression model errors have following conditional distribution:

$(\epsilon, u)|\varphi_1, \varphi_2, \sim EL_{2n}(0, diag(\varphi_1^{-1}I_n, \varphi_2^{-1}I_n), h)$

Having supposed that (φ_1, φ_2) has squared radial distribution Vidal and Arellano-Valle did the Bayes inference on the model (1) and (2). Now suppose following the elliptical model :

$y, x \mid \beta, \phi, \xi \sim El_{2n}\left[(\gamma\beta, \xi), \phi^{-1}I_{2n}, h_{\alpha\phi}^{(2n|2n+d)}\right]$

$\pi(\xi, \beta|\phi, h) \propto 1$

$\phi \propto \dfrac{(\pi\alpha)^{\frac{d}{2}}}{\Gamma\frac{d}{2}} \phi^{\frac{d}{2}-1} h^{(d)}(\alpha\phi)$

where h is generating function and $\alpha > 0$ and $d \in N$ are clear and $h_{\alpha\phi}^{(2n|2n+d)}$ is the conditional generating function of the next 2n and previous distribution ϕ is a squared radial function with 'd' freedom degree. By virtue of the model (5)-(3) Vidal & Arellano-Valle proved that the posterior distribution $(\beta_0, \beta_1, \varphi, \varepsilon)$ is as follows:

$$f(\xi, \beta|y, x, h) \propto (a + \|y - \gamma\beta\|^2 + \|x - \xi\|^2)^{-\frac{2n+d}{2}} \pi(\xi, \beta|\phi, h)$$

where $\beta = (\beta_0, \beta_1)$ and $\gamma = [\mathbb{1}_n, \xi]$ .By integrating ξ on parameter it is possible to have marginal distribution β as follows :

$$f(\beta|y, x, \alpha) \propto \left(\alpha + \frac{1}{(1+\beta^2)}\|y - X\beta\|^2\right)^{-\frac{n+d}{2}} (1 + \beta^2)^{-\frac{n}{2}} \qquad (6)$$

Posterior Average of Structural Parameters

In this section supposing is unknown and the usual a priori distribution the Bayes model has been given in (3)-(5):

$$\pi(\alpha) \propto \alpha^{-p}, p > 0$$

Now we should find complete conditional algorithm for Gibbs algorithm and estimate a priori average estimation β. Supposing the non-information posterior distribution $\pi(\alpha) \propto \alpha^{-p}, p > 0$ for α we have:

$$f(\alpha|y, x, \beta) \propto \alpha^{\frac{d}{2}-p}(\alpha + \frac{1}{1 + \beta_1^2}\|y - X\beta\|^2)^{\frac{-n+d}{2}}$$

which is the gamma – gamma distribution kernel with following parameters:
$Gg\left[n/2 + p - 1, (1 + \beta_1^2)^{-1}\|y - X\beta\|^2, \frac{d}{2} - p + 1\right]$.
If $> -2(1 - p), n > 2(1 - p)$, $f(\alpha|y, x, \beta)$ is appropriate distribution. Now considering the β is multivariable it is uneasy to produce data from it to be able to estimate the posterior distribution average . So we use Gibbs sampling method. Algorithm is a device to produce random data from multivariable distribution from which it is uneasy to produce data. Here we may use Gibbs algorithm in detail. We put $\beta_0 = \hat{\beta}_0$ for $t = 1,2, \ldots$, and produce two following steps::

1. $\alpha_t \sim f(\alpha|y, x, \beta_{t-1})$
2. $\beta_t \sim f(\beta|y, x, \alpha_t)$

This algorithm is stated in Vidal and Bolfarine (2011). In next section we examine the findings on real data.

Examining scientific production data

In this section we examine the findings on real data. Here we use the model defined in (1) . The data in following Table show two variables: Nos. of the articles and score of participation in M.A. (M.S.) thesis. The data are about 17 members of scientific board in one centre of Payame Noor University in 2011. The members scientific productions depend on several factors , but their professional and internal motives to do and create science vary . It should be emphasized that if such activities are always protected, related participation and scientific productions increase, too. So we consider financial factors with participation variable in M.A. (M.S.) theses like X and ξ relation in the (1) relations we may have a more exact model for ξ and Y.

Table 2. Raw Data and Statistical Descriptive

Row	1	2	3	4	5	6	7	8	9
Y	29	17	5	7	27	23	1	4	4
X	34.5	28	22	22	61.5	44.0	24.0	21.5	12.0
Row	10	11	12	13	14	15	16	17	
Y	18	2	4	2	11	0	4	0	
X	42.5	55.5	5.0	0.0	0.0	4.0	4.0	7.5	

Variable	Minimum	1st quartile	Medium	Average	3rd quartile	Maximum
X	0	7.50	24	28.62	44	75.50
Y	0	2	4	9.23	17	29

X: Number of articles   Y: Score of participation in thesis

In first step the correlation between X and Y was gained as 0.4201769. The model in the example is the elliptical model defined in (3)-(5). Also the samples were gained from posterior distribution β by Gibbs algorithm. The figure 1 is the dispersion of all data with three fitting lines according to three maximum likelihood methods with the total least errors squares and Bayes method .
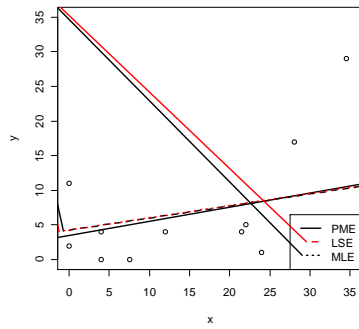
Figure 1. Scientific Productions Rate

As you see in Figure (6-4) two maximum likelihood estimation fitting lines and total least errors squares are intersected; it might be seen in previous example, too .In Table (5-4) it may be seen the estimations by three mentioned methods. In third column of the Table (5-4) low error square average MSE indicates the model gained by PME with a very little difference from the Model LSE is better than two other models.

Table 4. The rates estimated by regression parameters

	B_0	B_1	MSE
MLE	3.43951	0.20458	321.1932
LSE	4.25631	0.17603	296.1852
PME	4.27037	0.17367	176.0832

As it is clear from Table (4) the Bayes estimation rate is near the total least errors squares described analytically in section (3-4).

Table 5. Regression Parameters Summary :

		Average	Medium	Standard deviation
Metropolis- Within-Gibbs	B_0	4.27037	4.23148	0.01468
	B_1	0.17367	0.17459	0.00052

We drew the B_0 posterior distribution histogram by software 'F' and showed that it has symmetric distribution .
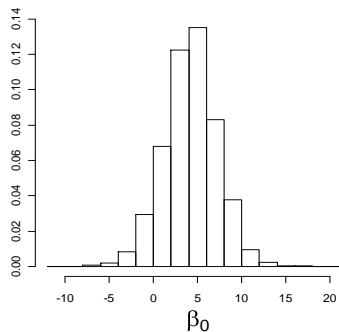

Figure 2. B_0 posterior distribution histogram

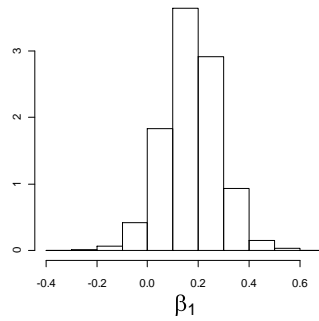Figure (2–4): We drew B_1 posterior distribution histogram; here it may be seen the symmetry.

Figure 3: B$_1$ posterior distribution histogram

Figure (4) shows the B$_0$ changes and rates. Here it may be seen the B$_0$ distribution symmetry, too.
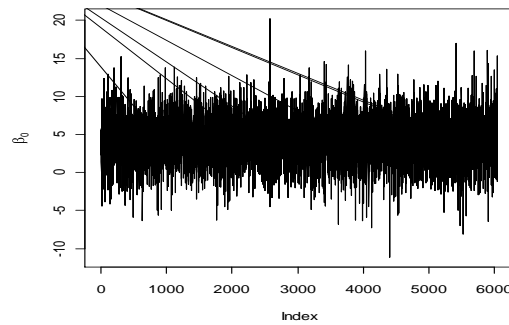

Figure 4. T the B$_0$ changes and rates

Figure (5) shows the B$_1$ changes and rates. Here it may be seen the B$_1$ distribution symmetry, too.
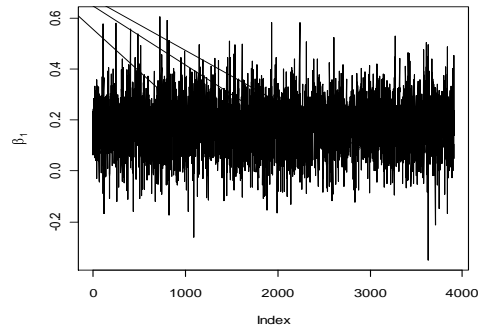

Figure 5. The B$_1$ changes and rates

Findings

The main purpose is to find a regression model by two visible variables with an average in which no common relation with third variable is visible. In recent applied subject it was shown that the regression model gained by posterior data distribution may be a linear model with the least error squares method. If there are more comprehensive and better knowledge about the variables B$_1$ and B$_1$ behavior and the posterior distribution is more proper, the findings are more exact.

REFERENCES

Bernardo, J.M., Smith, A.F.M., 1994, Bayesian Theory. Wiley, Chichester.

Bolfarine, H., Arellano-Valle, R.B., 1998. Weak non differential measurement error models. Statistics & Probability Letters 40, 279–287.

Bolfarine, H., Arellano-Valle, R.B., 2005. Elliptical measurement error models—a Bayesian approach. In: Dey, D.K., Rao, C.R. (Eds.), Bayesian Thinking:

Modeling and Computation. In: Handbook of Statistics, vol. 25. Elsevier B.V., North-Holland, pp. 669–688.

Carrol, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective. Chapman & Hall, CRC, Boca Raton.

Chen, M.H., Shao, Q.M., 1999. Monte Carlo estimation of Bayesian credible and HPD intervals. Journal of Computational and Graphical Statistics 8, 69–92.

Cheng, C.-L., van Ness, J.W., 1999. Statistical Regression with Measurement Error. In: Kendall's Library of Statistics, vol. 6. Arnold, London.

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, London, New York.

Fuller, W.A., 1987. Measurement Error Models, vol. 1. John Wiley & Sons, New York.

Hobert, J.P., Robert, C.P., Goutis, C., 1997. Connectedness conditions for the convergence of the Gibbs sampler. Statistics & Probability Letters 33, 235–240.

Kelly, G.E., 1984. The influence function in the errors in variables problem. The Annals of Statistics 12, 87–100.

Lindley, D.V., El-Sayyad, G.M., 1968. The Bayesian estimation of a linear functional relationship. Journal of the Royal Statistical Society. Series B 30, 190–202.

O'Hagan, A., 1994. Bayesian Inference. In: Kendall's Advanced Theory of Statistics, vol. 2B. Edward Arnold, London.

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods, second ed. Springer, New York.

Vidal, I., Arellano-Valle, R. B., 2010. Bayesian inference for dependent elliptical measurement error models. Journal of Multivariate Analysis 101, 2587–2597.

Vidal, I., and Bolfarine, H., 2011 Bayesian estimation of regression parameters in elliptical measurement error models. Statistical and Probability Letters, 81, 1398-1406.