Bayesian Inference of a Non normal Multivariate Partial Linear Regression Model

Sarmad A.Salih and Emad H.Aboudi

Department of Statistic, College of Administration and Economics, University of Baghdad, Iraq

Article information	Abstract
	This research includes the Bayesian estimation of the parameters of the multivariate partial linear regression model when the random error follows the matrix-variate generalized modified Bessel distribution and found the statistical test of the model represented by finding the Bayes factor criterion, the predictive distribution under assumption that the shape parameters are known. The prior distribution about the model parameters is represented by non-informative information, as well as the simulate on the generated data from the model by a suggested way based on different values of the shape parameters, the kernel function used in the generation was a Gaussian kernel function, the bandwidth (Smoothing) parameter was according to the rule of thumb. It found that the posterior marginal probability distribution of the location matrix θ and the predictive probability distribution is a matrix-t distribution with different parameters, the posterior marginal probability distribution of the scale matrix Σ is proper distribution but it does not belong to the conjugate family, Through the Bayes factor criterion, it was found that the sample that was used in the generation process was drawn from a population that does not belong to the generalized modified Bessel population.

1. Introduction

Most of the classical literature on multivariate estimation and hypotheses testing for a multivariate partial linear regression model when the random error limit is normally distributed, but there are cases in which the random error observations may be dependent but uncorrelated, or the data distribution belongs to probability distributions it has heavy tails that are heavier than the tails of the normal distribution. In such a case, the mixed distributions are more appropriate, and one of these distributions is a matrix-variate generalized modified Bessel distribution. The Matrix-variate generalized modified Bessel distribution belongs to the family of symmetric heavy-tailed probability distributions and is considered a continuous probability distribution. In addition, this distribution has special applications in the market of securities, random signal analysis, quality control, and filtering (Thabane, L., & Drekic, S. 2003). (Thabane, L., & Drekic, S. 2003) studied the characteristics of the generalized multivariate modified Bessel distribution of the with its special cases and confirmed that the mixed probability distributions such as the mixed multivariate normal distribution and the multivariate student-t distribution as special cases of it, and that the mixed distribution resulted from the Multivariate normal distribution with a generalized inverse Gaussian distribution as well as its applications in the Bayesian analysis of the normal multiple linear regression model assuming a generalized inverse Gaussian distribution as a prior distribution of the variance parameter. (Thabane, L., & Haq, M. S. 2004) generalized multivariate modified Bessel distribution (symmetric multivariate generalized hyperbolic distribution) to the matrix-variate generalized modified Bessel distribution and its special case studies as well as its applications in the Bayesian analysis of the normal multivariate linear regression model assuming the matrix generalized inverse Gaussian distribution as a prior distribution of the scale matrix. (Choi, et al. 2009) tested a statistical hypothesis in the Bayesian

technique of the normal multiple partial regression model and assumed that the parametric part of the model is a linear multidimensional function while the nonparametric part is an infinite series of trigonometric functions and deduced upon increasing the sample size that the Bayes factor is under the null hypothesis of the linear function is consistent, that is, it approaches infinity while it approaches zero under the alternative hypothesis of the partial linear function.The second section deals with the description of the multivariate partial linear regression model when the error follows the matrix-variate generalized modified Bessel distribution. The third section shows some types of kernel functions. Some methods of selecting the bandwidth parameter were presented in the fourth section. The fifth section includes the Bayesian estimate of the model parameters when non-informative prior information is available. The sixth section includes finding the Bayes factor criterion, The predictive distribution of future observations in the section seventh. While the eight section include a simulation of generated data from the model. The last section shows the most important conclusions and future studies.

2. Description of a Multivariate Partial Linear Regression Model

The multivariate partial linear regression model is described according to the following equation: (Przystalski, M. 2014) (You, J., et al 2013)

$$Y_{im} = X_i'\beta_m + g_m(T_i) + \epsilon_{im} \quad i = 1,2,\dots,n \ , m = 1,2,\dots,k \tag{1}$$

Where $X_i'\beta_m$ represents the parametric part of the model, and β_m is estimated by one of the parametric methods, such as the method of least squares, the maximum likelihood, moments, or Bayes ..., and $g_m(T_i)$ represents the nonparametric part of the model, which is an unknown smoothing function that is estimated by one of the nonparametric methods, such as the kernel smoother, nadaraya-watson smoother, and the Gasser-muller smoother ....., It is possible to write the model defined in equation (1) in the form of matrices as follows: (AL-Mouel, A. S., & Mohaisen, A. J. 2017)

$$Y = X\beta + W\gamma + \epsilon \tag{2}$$

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nk} \end{bmatrix}_{n\times k}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}_{n\times p+1}$$

$$\beta = \begin{bmatrix} \beta_{01} & \cdots & \beta_{0k} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pk} \end{bmatrix}_{p+1\times k}, W = \begin{bmatrix} ker_h(t_1 - T_{11}) & \cdots & ker_h(t_s - T_{1s}) \\ \vdots & \ddots & \vdots \\ ker_h(t_1 - T_{n1}) & \cdots & ker_h(t_s - T_{ns}) \end{bmatrix}_{n\times s}$$

$$\gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1k} \\ \vdots & \ddots & \vdots \\ \gamma_{s1} & \cdots & \gamma_{sk} \end{bmatrix}_{s\times k}, \epsilon = \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1k} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{nk} \end{bmatrix}_{n\times k}$$

where:

Y: Matrix of response variables of dimension $(n \times k)$ and n represents the number of observations and k represents the number of response variables.

X: A non-random matrix representing the observations of the parametric explanatory variables of dimension $(n \times p + 1)$ and that p represents the number of the parametric explanatory variables.

β: Matrix of model parameters for the parametric part of the dimension $(p + 1 \times k)$.

W: The design matrix indicates the kernel weights. It can be taken with other weights such as the spline, wavelet, and k-nearest neighbor weights. It is of dimension $(n \times s)$, and s represents the number of nonparametric explanatory variables, and $ker_h(.)$ represents the kernel function is as follows:

$$ker_h(.) = \frac{1}{h} \ ker(\frac{.}{h})$$

And that this function is a real, symmetric, and continuous function and that h represents the bandwidth parameter, they will be mentioned later.

γ : Matrix of parameters of the nonparametric part (additive parameters) of the dimension $(s \times k)$.

ϵ: Matrix of random errors of dimension $(n \times k)$.

It is possible to rewrite the form defined in Equation (2) as follows: (AL-Mouel, A. S., & Mohaisen, A. J. 2017)

$$Y_{n\times k} = C_{n\times(p+s+1)}\theta_{(p+s+1)\times k} + \epsilon_{n\times k} \tag{3}$$

Where: $C = [X \quad W] \quad , \quad \theta = [\beta \quad \gamma]^T$

Assume that the random error matrix of the model has a probability distribution is symmetric heavy tails represented by the matrix-variate generalized modified Bessel distribution, where the probability density function can be found using the mixed distributions from the mixed matrix normal distribution (normal variance mixture) and the generalized inverse Gaussian distribution as follows: (Gallaugher, M. P.B., & McNicholas, P.D. 2019) (Thabane, L., & Drekic, S. 2003) (Thabane, L., & Haq, M. S. 2004)$\epsilon|Z \sim MN_{n,k}(0, Z\Sigma, I_n)$, $Z \sim GIG(\lambda, \psi, v)$

As the probability density function for $\epsilon|Z$ is as follows:

$$f(\epsilon|Z) = \frac{1}{(2\pi Z)^{\frac{nk}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2Z} tr\, (\epsilon)^T (\epsilon)\Sigma^{-1}} \quad , \quad -\infty < \epsilon < \infty \qquad (4)$$

The probability density function of the random variable Z that follows the generalized inverse Gaussian distribution is as follows: (Lemonte, A. J., & Cordeiro, G. M. 2011)

$$P(Z) = \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{v}{2}}}{2\, K_v\left(\sqrt{\lambda\psi}\right)} Z^{v-1} exp\left[-\frac{1}{2}\left(\left(\frac{\psi}{Z}\right) + \lambda Z\right)\right] \qquad , Z > 0 \qquad (5)$$

Where: λ, ψ: scale parameters. v: shape parameter.

$K_v(.)$: represents the modified Bessel function of the third kind of order v which takes the following equation: (Gallaugher, M. P.B., & McNicholas, P.D. 2019) (Koudou, A. E., & Ley, C. 2014)

$$K_v(x) = 0.5 \int_0^\infty t^{v-1} \exp\left(-0.5\, x(t + t^{-1})\right) dt \qquad x > 0 \qquad (6)$$

The space of the scale parameters of the distribution is defined according to the following equation: (Thabane, L., & Drekic, S. 2003)

$$\begin{array}{llll} \psi > 0 & , & \lambda \geq 0 & for\ v < 0 \\ \psi > 0 & , & \lambda > 0 & for\ v = 0 \\ \psi \geq 0 & , & \lambda > 0 & for\ v > 0 \end{array} \qquad (7)$$

Therefore, the probability distribution of the matrix of random errors (ϵ) and according to the concept of mixed distributions is as follows:

$$f(\epsilon) = \int_0^\infty f(\epsilon|Z)\, P(Z)\ dZ$$

$$= \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{nk}{4}}\ K_{\frac{2v-nk}{2}}\left(\sqrt{\lambda\psi\left(1 + \frac{tr\ \epsilon^T\epsilon\Sigma^{-1}}{\psi}\right)}\right)}{(2\pi)^{\frac{nk}{2}}\ |\Sigma|^{\frac{n}{2}}\ K_v\left(\sqrt{\lambda\psi}\right)} * \left(1 + \frac{tr\ \epsilon^T\epsilon\Sigma^{-1}}{\psi}\right)^{\frac{2v-nk}{4}} \qquad (8)$$

Equation (8) represents the probability density function of the matrix-variate generalized modified Bessel distribution and is described as follows:

$$\epsilon \sim MGMB_{n,k}(0, \Sigma, I_n, \lambda, \psi, v) \ \leftrightarrow\ vec(\epsilon) \sim MGMB_{nk}(vec(0), \Sigma \otimes I_n, \lambda, \psi, v)$$

Since the matrix Y defined in equation (3) is a linear combination in terms of the matrix ϵ that follows the matrix-variate generalized modified Bessel distribution, and accordingly, the probability distribution of the response observations matrix (Y) follows the matrix-variate generalized modified Bessel distribution. It can be found in the same way as follows: (Thabane, L., & Haq, M. S. 2004)

$$E(Y|Z) = C\theta + E(\epsilon|Z)$$

$$E(Y|Z) = C\theta$$

$$V(Y|Z) = V(\epsilon|Z)$$

$$\therefore (Y|Z) \sim MN_{n,k}(C\theta, Z\Sigma, I_n)$$

Therefore, the probability density function of the matrix of (Y) distribution conditional by (Z) $(Y|Z)$ that follows the mixed matrix normal distribution is as follows:

$$f(Y|Z) = \frac{1}{(2\pi Z)^{\frac{nk}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2Z} tr\, (Y-C\theta)^T (Y-C\theta)\Sigma^{-1}} \qquad (9)$$

The probability distribution of (Y) unconditional by (Z) and depending on the concept of mixed distributions is as follows:

$$f(Y) = \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{nk}{4}} K_{\frac{2v-nk}{2}}\left(\sqrt{\lambda\psi\left(1 + \frac{tr\ (Y-C\theta)^T(Y-C\theta)\Sigma^{-1}}{\psi}\right)}\right)}{(2\pi)^{\frac{nk}{2}}\ |\Sigma|^{\frac{n}{2}}\ K_v\left(\sqrt{\lambda\psi}\right)}$$

$$* \left(1 + \frac{tr\ (Y-C\theta)^T\ (Y-C\theta)\Sigma^{-1}}{\psi}\right)^{\frac{2v-nk}{4}} \quad (10)$$

Express this distribution descriptively as follows:

$Y \sim MGMB_{(n,k)}\big(C\theta, \Sigma, I_n, \lambda, \psi, v\big) \leftrightarrow vec(Y) \sim MGMBH_{(nk)}\big(vec(C\theta), \Sigma \otimes I_n, \lambda, \psi, v\big)$

Where:θ: The location matrix with a dimension $(p + s + 1 \times k)$. Σ: The scale matrix with a dimension $(k \times k)$.

λ, ψ, v: shape parameters.

3. Kernel Functions

The kernel functions are used in estimating the regression functions, the spectral functions, and the probability density functions. These functions can be distinguished through two series, namely, the optimal kernel functions which reduce the AMISE criterion, and the kernel functions with the least variance and works to reduce the asymptotic variance, meaning the MISE derivation relative to the kernel function. (Schucany, W. R., & Sommers, J. 1977). The kernel function has other names, including (shape, weight, and window function), and the kernel function is a real, symmetric, continuous, and definite function, and its integral is equal to one. The following table reviews some types of the kernel function: (Langrene, N. & Warin, X. 2019)

Table (1): Some kernel functions

Kernel	Ker(x)	
Epanchnikov	$(3/4)(1 - x^2)$	$I(|x| \le 1)$
Quartic	$(15/16)(1 - x^2)^2$	$I(|x| \le 1)$
Triweight	$(35/32)(1 - x^2)^3$	$I(|x| \le 1)$
Triangular	$(1 - |x|)$	$I(|x| \le 1)$
Gauss	$(2\pi)^{-0.5}\exp(-x^2/2)$	$I(|x| < \infty)$
Uniform	*0.5*	$I(|x| \le 1)$
Tricube	$(70/81)(1 - |x|^3)^3$	$I(|x| \le 1)$
Cosine	$(\pi/4)\cos\left(\frac{\pi}{2}x\right)$	$I(|x| \le 1)$

4. Some Methods of Selecting the Bandwidth Parameter

The selection of the bandwidth parameter (*h*) is an essential part of estimating the nonparametric and semi-parametric regression curve, and that the selection of the bandwidth parameter is more important than the choice of the kernel function, and there are several names for this parameter, including (constraint amplitude - bandwidth parameter-smoothing parameter-variance parameter), one of its characteristics is a non-random, symmetric, and positive parameter, and usually the selection of the bandwidth parameter is based on the researcher's experience or iterative methods to obtain the best bandwidth parameter and that this parameter greatly affects the variance and bias, as increasing the bandwidth parameter leads to a decrease in variance and increase the bias and vice versa, The researcher must estimate it in a way that balances variance and bias, as well as it represents a function in terms of sample size so that it meets the following conditions: (Hardle, W. 1991) (Silverman, B.W. 1986) $\lim_{n\to\infty} h = 0$; $\lim_{n\to\infty} n\,h = \infty$

There are several ways to choose the bandwidth parameter, including:

4.1 Cross Validation Method

This method is considered one of the most used methods for selecting the smoothing parameter if gradually excludes one value from the response and explanatory variables to determine the parameter (*h*) that makes the sum of squares error at its lower end and is sometimes called the (Leave-one-out) method, The parameter (*h*) is given by the following formula: (Aydin, D. & Tuzemen, M. 2010)

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{g}_h^{(-1)}T_i\right)^2 \qquad (11)$$

Therefore, the bandwidth parameter gives the smallest value for the criterion $CV(h)$ as follows:

$$h_{CV} = argmin \; CV(h) \tag{11a}$$

4.2 Rule of Thumb Method

This method goes back to the world of **Deheuvels**, which depends on replacing the unknown smoothing function with the distribution function. It was circulated by **Silverman** and it is also called the normal distribution rule. And write its general formula as follows:

$$h_{thumb} = \hat{\sigma} \; CV(k) \; n^{-\frac{1}{2v+1}} \tag{12}$$

Where (v) denotes the kernel degree and the moments of the odd degree are equal to zero, the kernel degree (v) represents the first non-zero moment, meaning that the kernel degree must be an even number, while ($\hat{\sigma}$) indicates the standard deviation of the sample .and $CV(k)$ is constant, as shown in Table (2) below, and depends on the degree of the kernel. (Silverman, B.W. 1986)

Table (2): The value of $CV(k)$ for some kernel functions and by degree of kernel

Kernel degree	v = 2	v = 4	v = 6
Quartic	2.78	3.39	3.84
Gaussian	1.06	1.08	1.08

5. Bayesian Estimation of the Parameters of the Multivariate Partial Linear Regression Model

In this section, the parameters of the model defined in equation (3), represented by the location matrix (θ) and the scale matrix (Σ) are estimated under the assumption that they are unknown matrices and that the prior distributions of these parameters are non-informative. The joint prior distribution of (θ, Σ) is found from Fisher's information by taking the natural logarithm of the two sides of the probability density function of (Y) conditional by (Z) ($Y|Z$) and knowing in equation (9) and taking the second partial derivative relative to (θ, Σ), the joint prior distribution of ($\theta, \Sigma|Z$) is as follows: (Press, S. J. 2003)

$$P(\theta|\Sigma, Z) = |\Sigma|^{-\frac{(p+s+1)}{2}} \tag{13}$$

$$P(\Sigma|Z) \propto |\Sigma|^{-\frac{k+1}{2}} \tag{14}$$

$$P(\theta, \Sigma|Z) \propto P(\theta|\Sigma, Z) \; P(\Sigma|Z) \tag{15}$$

$$P(\theta, \Sigma|Z) \propto |\Sigma|^{-\frac{p+s+k+2}{2}} \tag{16}$$

By merging the joint prior probability distribution defined in equation (16) with the probability function of (Y) conditional by (Z) defined in equation (9), we obtain the kernel of the joint posterior probability distribution for (θ, Σ) conditional by the random variable (Z) as follows:

$$P(\theta, \Sigma|Y, Z) \propto P(\theta, \Sigma|Z) f(Y|\theta, \Sigma, Z) \propto |\Sigma|^{-\frac{n+p+s+k+2}{2}} \; e^{-\frac{1}{2Z} tr (Y-C\theta)^T (Y-C\theta)\Sigma^{-1}} \tag{17}$$

By adding and subtracting the amount ($C\hat{\theta}^*$) to the exponential function in equation (17) and which ($\hat{\theta}^*$) represents the maximum likelihood estimator conditional by the random variable Z, it is found by the partial derivation of the natural logarithm of equation (9) relative to θ:

$$\hat{\theta}^* = (C^T C)^{-1} C^T Y \tag{18}$$

And by performing some mathematical operations, we get the following:

$$P(\theta, \Sigma|Y, Z) \propto |\Sigma|^{-\frac{n+k+1}{2}} |\Sigma|^{-\frac{(p+s+1)}{2}} \; \exp\left(-\frac{1}{2Z} tr \; B_3 \; \Sigma^{-1}\right)$$

$$* \exp\left(-\frac{1}{2Z} tr \left(\theta - \hat{\theta}^*\right)^T C^T C \left(\theta - \hat{\theta}^*\right) \Sigma^{-1}\right) \tag{19}$$

Where:

$$B_3 = \left(Y - C\hat{\theta}^*\right)^T \left(Y - C\hat{\theta}^*\right) \quad, \quad \hat{\theta}^* = (C^T C)^{-1} C^T Y$$

From equation (19) we notice that $\left[|\Sigma|^{-\frac{(p+s+1)}{2}} \exp\left(-\frac{1}{2Z} tr \left(\theta - \hat{\theta}^*\right)^T C^T C \left(\theta - \hat{\theta}^*\right) \Sigma^{-1}\right)\right]$ it represents kernel of the matrix normal variance mixture distribution of (θ) and the distribution parameters are$(\hat{\theta}^*, Z \Sigma, (C^T C)^{-1})$, that $\left[|\Sigma|^{-\frac{n+k+1}{2}} e^{-\frac{1}{2Z} tr \; B_3 \; \Sigma^{-1}}\right]$ it represents kernel of the inverse Wishart distribution by parameters$\left(\frac{B_3}{Z}, n\right)$ of Σ.

Therefore, the joint posterior probability distribution of (θ, Σ) conditional by the random variable (Z)$(\theta, \Sigma|Y, Z)$ is as follows:

$$P(\theta,\Sigma|Y,Z) = \frac{|B_3|^{\frac{n}{2}} \; Z^{-\frac{nk}{2}}}{|\Sigma|^{\frac{n+k+1}{2}} \; 2^{\frac{nk}{2}} \; \Gamma_k\left(\frac{n}{2}\right)} \; \exp\left(-\frac{1}{2Z} \; tr \; B_3 \; \Sigma^{-1}\right)$$

$$* \frac{|C^TC|^{\frac{k}{2}}}{(2\pi Z)^{\frac{(p+s+1)k}{2}}|\Sigma|^{\frac{(p+s+1)}{2}}} \; e^{-\frac{1}{2Z} tr \, (\theta-\hat{\theta}^*)^T C^T C \, (\theta-\hat{\theta}^*) \, \Sigma^{-1}} \qquad (20)$$

Where: $\Gamma_k\left(\frac{n}{2}\right)$: Multivariate Gamma function is calculated as follows: (Nagar, D. K. & Gupta, A. K. 2013)

$$\Gamma_k(x) = (\pi)^{\frac{(k-1)k}{4}} \prod_{j=1}^{k} \Gamma(x + (1-j)/2)$$

In order to obtain the joint posterior probability distribution of (θ,Σ) that is not conditioned by the random variable Z, we integrate equation (20) relative to (Z) as follows:

$$P(\theta,\Sigma|Y) = \int_0^\infty P(\theta,\Sigma|Y,Z) \, P(Z) \, dZ$$

$$\boldsymbol{P(\theta,\Sigma|Y)} = \frac{|B_3|^{\frac{n}{2}} \; |C^TC|^{\frac{k}{2}} \left(\frac{\lambda}{\psi}\right)^{\frac{k(n+(p+s+1))}{4}}}{(2\pi)^{\frac{(p+s+1)k}{2}}|\Sigma|^{\frac{n+(p+s+1)+k+1}{2}} \; 2^{\frac{nk}{2}} \; \Gamma_k\left(\frac{n}{2}\right) K_v(\sqrt{\lambda\psi})}$$

$$* \; K_{\frac{2v-k(n+(p+s+1))}{2}}\left(\sqrt{\lambda\psi\left(1 + \frac{tr\left(B_3 + (\theta-\hat{\theta}^*)^T C^T C(\theta-\hat{\theta}^*)\right)\Sigma^{-1}}{\psi}\right)}\right)$$

$$* \left(1 + \frac{tr\left(B_3 + (\theta-\hat{\theta}^*)^T C^T C(\theta-\hat{\theta}^*)\right)\Sigma^{-1}}{\psi}\right)^{\frac{2v-k(n+(p+s+1))}{4}} \qquad (21)$$

To find the posterior marginal probability distribution of the location matrix (θ) conditioned by the random variable (Z), we integrate equation (20) relative to the scale matrix (Σ) as follows:

$$P(\theta|Y,Z) = \int_\Sigma P(\theta,\Sigma|Y,Z) \; d\Sigma$$

$$P(\theta|Y,Z) = \frac{|B_3|^{\frac{n}{2}} \; |C^TC|^{\frac{k}{2}} \; \Gamma_k\left(\frac{n+(p+s+1)}{2}\right)}{(\pi)^{\frac{(p+s+1)k}{2}} \; \Gamma_k\left(\frac{n}{2}\right)} \; \left|B_3 + (\theta-\hat{\theta}^*)^T C^T C(\theta-\hat{\theta}^*)\right|^{-\frac{n+(p+s+1)}{2}} \qquad (22)$$

It is possible to rewrite equation (22) as follow:

$$P(\theta|Y,Z) = R(k,(p+s+1),r) * \frac{|B_3|^{-\frac{(p+s+1)}{2}} \; |C^TC|^{\frac{k}{2}}}{\left|I_k + (\theta-\hat{\theta}^*)^T C^T C(\theta-\hat{\theta}^*)B_3^{-1}\right|^{\frac{n+(p+s+1)}{2}}} \qquad (23)$$

Where:

$$R(k,(p+s+1),r) = \frac{\Gamma_k\left(\frac{n+(p+s+1)}{2}\right)}{(\pi)^{\frac{(p+s+1)k}{2}} \; \Gamma_k\left(\frac{n}{2}\right)} \qquad , \; r \; is \; degree \; of \; freedom$$

And using property of Box and Tiao. (Box, G. P., & Tiao, G. C. 1973)

$$\left|I_k + P_{k\times(p+s+1)} \, Q_{(p+s+1)\times k}\right| = \left|I_{(p+s+1)} + Q_{(p+s+1)\times k}P_{k\times(p+s+1)}\right|$$

Thus:

$$P(\theta|Y,Z) = \frac{R(k,(p+s+1),r) \quad |B_3|^{-\frac{(p+s+1)}{2}} \; |C^TC|^{\frac{k}{2}}}{\left|I_{(p+s+1)} + (\theta-\hat{\theta}^*)B_3^{-1}(\theta-\hat{\theta}^*)^T C^T C\right|^{\frac{n+(p+s+1)}{2}}} \qquad (24)$$

And:

$$R(k,(p+s+1),r) = R((p+s+1),k,r)$$

$$\frac{\Gamma_k(\frac{n+(p+s+1)}{2})}{(\pi)^{\frac{(p+s+1)k}{2}} \ \Gamma_k(\frac{n}{2})} = \frac{\Gamma_{(p+s+1)}(\frac{n+(p+s+1)}{2})}{(\pi)^{\frac{(p+s+1)k}{2}} \ \Gamma_{(p+s+1)}(\frac{n-k+(p+s+1)}{2})}$$

Hence, the posterior marginal probability distribution of the parameter matrix θ unconditional of the variable Z is as follows:

$$P(\theta \mid Y) = \frac{|B_3|^{-\frac{(p+s+1)}{2}} \ |C^T C|^{\frac{k}{2}} \ \Gamma_{(p+s+1)}(\frac{n+(p+s+1)}{2})}{(\pi)^{\frac{(p+s+1)k}{2}} \ \Gamma_{(p+s+1)}(\frac{n-k+(p+s+1)}{2})}$$

$$* \left| I_{(p+s+1)} + (\theta - \hat{\theta}^*) B_3^{-1}(\theta - \hat{\theta}^*)^T C^T C \right|^{-\frac{n+(p+s+1)}{2}} \tag{25}$$

We notice from equation (25) that the posterior marginal probability distribution of θ is a matrix-t distribution of the parameters $(\hat{\theta}^*, B_3, (C^T C)^{-1}, r = n-k+1)$ and is described as follows:

$$\theta \sim Mt_{(p+s+1),k}(\hat{\theta}^*, B_3, (C^T C)^{-1}, r) \ , vec(\theta) \sim Mt_{(p+s+1)k}(vec(\hat{\theta}^*), B_3 \otimes (C^T C)^{-1}, r)$$

And the estimate Bayes under quadratic loss function of (θ) is:

$$\hat{\theta}_B = \hat{\theta}^* = (C^T C)^{-1} C^T Y \tag{25a}$$

This estimator is similar if the error of the model follows the matrix normal and matrix-t distribution.

To find the posterior marginal probability distribution of the scale matrix (Σ) conditioned by (Z), we integrate equation (20) relative to the location matrix (θ) as follows:

$$P(\Sigma \mid Y, Z)$$

$$= \int_\theta P(\theta, \Sigma \mid Y, Z) \ d\theta = \frac{|B_3|^{\frac{n}{2}} \ Z^{-\frac{nk}{2}}}{|\Sigma|^{\frac{n+k+1}{2}} \ 2^{\frac{nk}{2}} \ \Gamma_k(\frac{n}{2})} \ \exp\left(-\frac{1}{2Z} \ tr \ B_3 \ \Sigma^{-1}\right) \tag{26}$$

Equation (26) represents the inverse Wishart distribution of parameters $\left(\frac{B_3}{Z}, n\right)$ of the posterior marginal probability distribution of (Σ) conditional by the random variable (Z). Based on Bayes theorem we conclude that the posterior marginal probability distribution of the scale matrix (Σ) unconditioned by (Z) is proper distribution but doesn't belong to the conjugate family, as follows:

$$P(\Sigma \mid Y) = \int_0^\infty P(\Sigma \mid Y, Z) \ P(Z) \ dZ$$

$$P(\Sigma \mid Y) = \frac{|B_3|^{\frac{n}{2}} \ \left(\frac{\lambda}{\psi}\right)^{\frac{nk}{4}}}{|\Sigma|^{\frac{n+k+1}{2}} \ 2^{\frac{nk}{2}} \ \Gamma_k(\frac{n}{2}) \ K_v(\sqrt{\lambda\psi})} \ K_{\frac{2v-nk}{2}}\left(\sqrt{\lambda\psi(1+\frac{tr \ B_3 \ \Sigma^{-1}}{\psi})}\right)$$

$$* \left(1 + \frac{tr \ B_3 \Sigma^{-1}}{\psi}\right)^{\frac{2v-nk}{4}} \tag{27}$$

And the estimate Bayes of Σ under quadratic loss function is:

$$\hat{\Sigma}_B = E_Z E_{\Sigma \mid Z}(\Sigma \mid Y, Z) = \frac{(Y - C\hat{\theta}^*)^T (Y - C\hat{\theta}^*)}{n-k-1} \frac{K_{v-1}(\sqrt{\lambda\psi})}{K_v(\sqrt{\lambda\psi})} \left(\frac{\lambda}{\psi}\right)^{0.5} \tag{28}$$

6. Bayesian Hypotheses Testing of a Multivariate Partial Regression Model

The Bayes factor criterion is considered one of the important criteria applied Bayesian hypotheses testing and is defined as the proportion between two statistical hypotheses. It results from dividing the posterior probability distribution relative to the null hypothesis (H_0) on the posterior probability distribution relative to the alternative hypothesis (H_1). This criterion is expressed mathematically as follows: (Jefferys, H. 1961)

$$BF = \frac{P(Y \mid H_0)}{P(Y \mid H_1)} \tag{29}$$

To test the model, we know the following statistical hypothesis: (AL-Mouel, A. S., & Mohaisen, A. J. 2017)

$$H_0: Y = C \ \theta_0 + \epsilon \quad , \quad \theta_0 = [\beta_0 \quad \gamma_0]^T \quad , \Sigma > 0$$

$$H_1: Y = C\,\theta + \epsilon \quad , \Sigma > 0 \tag{30}$$

Under the above statistical hypothesis and based on equation (29), the Bayes factor criterion is as follows:

$$BF = \frac{\int_Z \int_\Sigma f(Y|\theta_0, \Sigma, Z)\, P(\Sigma)\, P(Z)\, d\,\Sigma\, dZ}{\int_Z \int_\Sigma \int_\theta f(Y|\theta, \Sigma, Z)\, P(\Sigma)\, P(\theta)\, P(Z)\, d\,\theta\, d\,\Sigma\, dZ} \tag{31}$$

We represent the numerator with the quantity (L_0) in equation (31) which represents the probability function of (Y) conditional by (Z) and the knowledge in equation (9) under the null hypothesis multiplied by the prior distribution of the scale matrix (Σ) defined in equation (14) and the generalized inverse Gaussian distribution defined in equation (5).

$$L_0 = \int_Z \int_\Sigma f(Y|\theta_0, \Sigma, Z)\, P(\Sigma)\, P(Z)\, d\,\Sigma\, dZ$$

$$L_0 = \int_Z \frac{P(Z)}{(2\pi Z)^{\frac{nk}{2}}} \int_\Sigma |\Sigma|^{-\frac{n+k+1}{2}} e^{-\frac{1}{2Z} tr\,(Y-C\theta_0)^T (Y-C\theta_0)\Sigma^{-1}}\, d\,\Sigma\, dZ \tag{32}$$

The integral of equation (32) relative to the matrix (Σ) represents the kernel of inverse Wishart distribution by the parameters $(\frac{B_1}{Z}, n)$.

Where: $B_1 = (Y - C\theta_0)^T (Y - C\theta_0)$

Accordingly:

$$L_0 = \int_Z \frac{P(Z)\,(2)^{\frac{nk}{2}}\,\Gamma_k\left(\frac{n}{2}\right)}{(2\pi Z)^{\frac{nk}{2}}\; \left|\frac{B_1}{Z}\right|^{\frac{n}{2}}}\; dZ \tag{33}$$

$$L_0 = \frac{\Gamma_k\left(\frac{n}{2}\right)}{(\pi)^{\frac{nk}{2}}\; |B_1|^{\frac{n}{2}}} \tag{34}$$

In the same procedure we find the denominator of the equation (31)

$$L_1 = \int_Z \int_\Sigma \int_\theta f(Y|\theta, \Sigma, Z)\, P(\Sigma)\, P(\theta)\, P(Z)\, d\,\theta\, d\,\Sigma\, dZ$$

$$L_1 = \int_Z \frac{P(Z)}{(2\pi Z)^{\frac{nk}{2}}} \int_\Sigma \int_\theta |\Sigma|^{-\frac{n+k+1}{2}} e^{-\frac{1}{2Z} tr\,(Y-C\theta)^T (Y-C\theta)\Sigma^{-1}}\, d\theta\, d\Sigma\, dZ \tag{35}$$

By adding and subtracting $(C\hat{\theta}^*)$ to the exponential function in equation (35) and which $(\hat{\theta}^*)$ represents the maximum likelihood estimator which was previously defined in equation (18) and by performing some mathematical operations, we get the following:

$$L_1 = \int_Z \frac{P(Z)}{(2\pi Z)^{\frac{nk}{2}}} \int_\Sigma \frac{e^{-\frac{1}{2Z} tr\,(Y-C\hat{\theta}^*)^T (Y-C\hat{\theta}^*)\Sigma^{-1}}}{|\Sigma|^{\frac{n+k+1}{2}}} \int_\theta e^{-\frac{1}{2Z} tr\,(\theta-\hat{\theta}^*)^T C^T C\,(\theta-\hat{\theta}^*)\,\Sigma^{-1}}\, d\theta\, d\Sigma\, dZ \tag{36}$$

The last exponential function of equation (36) represents the kernel of the matrix normal distribution conditioned by the variable (Z) relative to (θ) by the parameters $(\hat{\theta}^*, (C^T C)^{-1} Z\Sigma)$ and therefore:

$$L_1 = \int_Z \int_\Sigma \frac{P(Z)\, e^{-\frac{1}{2Z} tr\,(Y-C\hat{\theta}^*)^T (Y-C\hat{\theta}^*)\Sigma^{-1}}}{(2\pi Z)^{\frac{nk}{2}}\, |\Sigma|^{\frac{n+k+1}{2}} |C^T C|^{\frac{k}{2}}} \frac{(2\pi Z)^{\frac{(p+s+1)k}{2}}}{|\Sigma|^{-\frac{(p+s+1)}{2}}}\, d\Sigma\, dZ \tag{37}$$

$$L_1 = \int_Z \frac{P(Z)\,(2\pi Z)^{\frac{(p+s+1)k}{2}}}{(2\pi Z)^{\frac{nk}{2}}\, |C^T C|^{\frac{k}{2}}} \int_\Sigma |\Sigma|^{-\frac{n-(p+s+1)+k+1}{2}} e^{-\frac{1}{2Z} tr\,(Y-C\hat{\theta}^*)^T (Y-C\hat{\theta}^*)\Sigma^{-1}}\, d\Sigma\, dZ \tag{38}$$

The integral in equation (38) represents the kernel of inverse Wishart distribution by the parameters $(\frac{B_3}{Z}, n-(p+s+1))$.

Accordingly:

$$L_1 = \int_Z \frac{P(Z) \; \Gamma_k \left(\frac{n-(p+s+1)}{2} \right)}{(\pi)^{\frac{k(n-(p+s+1))}{2}} \; |C^T C|^{\frac{k}{2}} \, |B_3|^{\frac{n-(p+s+1)}{2}}} \; dZ \qquad (39)$$

$$L_1 = \frac{\Gamma_k \left(\frac{n-(p+s+1)}{2} \right)}{(\pi)^{\frac{k(n-(p+s+1))}{2}} \; |C^T C|^{\frac{k}{2}} \, |B_3|^{\frac{n-(p+s+1)}{2}}} \qquad (40)$$

To find the Bayes factor criterion, we divide equation (34) on the equation (40) as follows:

$$BF = \frac{\Gamma_k \left(\frac{n}{2} \right) \; |C^T C|^{\frac{k}{2}} \, |B_3|^{\frac{n-(p+s+1)}{2}}}{(\pi)^{\frac{(p+s+1)k}{2}} \; |B_1|^{\frac{n}{2}} \Gamma_k \left(\frac{n-(p+s+1)}{2} \right)} \qquad (41)$$

7. Predictive Distribution of a Multivariate Partial Linear Regression Model

The predictive distribution represents the probability density function for future observations Y_f that is conditioned by a set of current observations Y, so we have future observations (f) for all response variables, which represent the matrix (Y_f). Depending on future observations, the multivariate partial linear regression model is as follows: (Thabane, L., & Haq, M. S. 2004)

$$Y_f = C_f \theta + \epsilon_f \qquad (42)$$

Where:

Y_f: The matrix of future observations (f) has a dimension $(n_f \times k)$.

C_f: Matrix with dimension $(n_f \times (p + s + 1))$.

θ: The parameter matrix with dimension $((p + s + 1) \times k)$.

ϵ_f: The matrix of future random errors with dimension $(n_f \times k)$.

Since the error matrix (ϵ_f) follows the matrix–variate generalized modified Bessel distribution with the parameters $\left(0, \Sigma, I_{n_f}, \lambda, \psi, \upsilon \right)$, we know that (Y_f) is a linear combination in terms of the future error matrix, therefore (Y_f) follows the matrix–variate generalized modified Bessel distribution by the parameters $\left(C_f \theta, \Sigma, I_{n_f}, \lambda, \psi, \upsilon \right)$.

Using the Bayes theory, the predictive distribution of the future matrix Y_f is defined by the following formula:

$$f(Y_f|Y) = \int_\Sigma \int_\theta f(Y_f|\theta, \Sigma) \; P(\theta, \Sigma|Y) d\theta \; d\Sigma \qquad (43)$$

Due to the difficulty of finding a predictive distribution from equation (43), we use the concept of mixed distributions, that is, probability distribution conditioned by the random variable (Z).

$$f(Y_f|Y) = \int_Z \int_\Sigma \int_\theta f(Y_f|\theta, \Sigma, Z) \; P(\theta, \Sigma|Y, Z) \; P(Z) \; d\theta \; d\Sigma \; dZ \qquad (43a)$$

We know that the probability density function (Y_f) conditional by (Z) is described as follows:

$$(Y_f|\theta, \Sigma, Z) \sim MN_{n_f \times k} \left(C_f \theta , Z\Sigma, I_{n_f} \right)$$

Therefore, the probability density function for the conditional (Y_f) is as follows:

$$f(Y_f|\theta, \Sigma, Z) = \frac{1}{(2\pi Z)^{\frac{n_f k}{2}} |\Sigma|^{\frac{n_f}{2}}} e^{-\frac{1}{2Z} tr \, (Y_f - C_f \theta)^T (Y_f - C_f \theta) \Sigma^{-1}} \qquad (44)$$

The joint posterior probability distribution of (θ, Σ) conditional by (Z) and previously defined in equation (20) will be combined with the conditional probability density function (Y_f) defined in equation (44). We obtain the predictive distribution of (Y_f) conditional by the random variable (Z) as follows:

$$f(Y_f|Y, Z)$$

$$\propto \int_\Sigma \int_\theta |\Sigma|^{-\frac{n_f + n + (p+s+1) + k + 1}{2}} e^{-\frac{1}{2Z} tr \, B_3 \, \Sigma^{-1}} \; e^{-\frac{1}{2Z} tr \left\{ (Y_f - C_f \theta)^T (Y_f - C_f \theta) + (\theta - \hat{\theta}^*)^T C^T C \, (\theta - \hat{\theta}^*) \right\} \Sigma^{-1}} \; d\theta \; d\Sigma \qquad (45)$$

Assume that:

$$B_{111} = \left(Y_f - C_f\theta\right)^T\left(Y_f - C_f\theta\right) + \left(\theta - \hat{\theta}^*\right)^T C^T C \left(\theta - \hat{\theta}^*\right)$$

By adding and subtracting $\left(C_f\hat{\theta}^*\right)$ to the parentheses of the first term of B_{111} and that $(\hat{\theta}^*)$ represents the maximum likelihood estimator conditional by the random variable (Z) which was previously defined in equation (18) and by performing some mathematical operations, we get the following:

$$B_{111} = \left(Y_f - C_f\hat{\theta}^*\right)^T\left[I_{n_f} + C_f\phi_1^{-1}C_f^T\right]\left(Y_f - C_f\hat{\theta}^*\right) + \left(\theta - \hat{\theta}^* + \phi_2\right)^T\phi_1\left(\theta - \hat{\theta}^* + \phi_2\right)$$

As:

$$\phi_1 = C_f^T C_f + C^T C$$

$$\phi_2 = \phi_1^{-1}C_f^T\left(Y_f - C_f\hat{\theta}^*\right)$$

Accordingly:

$$f\left(Y_f|Y,Z\right) \propto \int_\Sigma |\Sigma|^{-\frac{n_f+n+k+1}{2}}e^{-\frac{1}{2Z}tr\left[B_3+\left(Y_f-C_f\hat{\theta}^*\right)^T\left[I_{n_f}+C_f\phi_1^{-1}C_f^T\right]\left(Y_f-C_f\hat{\theta}^*\right)\right]\Sigma^{-1}}$$

$$\int_\theta |\Sigma|^{-\frac{(p+s+1)}{2}} \quad e^{-\frac{1}{2Z}tr\left\{\left(\theta-\hat{\theta}^*+\phi_{22}\right)^T\phi_1\left(\theta-\hat{\theta}^*+\phi_{22}\right)\right\}\Sigma^{-1}} \quad d\theta \, d\Sigma$$

The result of integration relative to (θ) is $\left(\dfrac{(2\pi Z)^{\frac{(p+s+1)k}{2}}}{|\phi_1|^{\frac{k}{2}}}\right)$ and this expression will cancel out with the constant of proportionality, and therefore:

$$f\left(Y_f|Y,Z\right) \propto \int_\Sigma |\Sigma|^{-\frac{n_f+n+k+1}{2}}e^{-\frac{tr}{2Z}\left[B_3+\left(Y_f-C_f\hat{\theta}^*\right)^T\left[I_{n_f}+C_f\phi_1^{-1}C_f^T\right]\left(Y_f-C_f\hat{\theta}^*\right)\right]\Sigma^{-1}}\,d\Sigma \qquad (46)$$

The integral result of the above equation relative to the matrix Σ is the reciprocal of the constant of inverse Wishart distribution by the parameters $\left(\dfrac{B_{112}}{Z}, n_f + n\right)$.

Where:

$$B_{112} = B_3 + \left(Y_f - C_f\hat{\theta}^*\right)^T\left[I_{n_f} + C_f\phi_1^{-1}C_f^T\right]\left(Y_f - C_f\hat{\theta}^*\right)$$

$$f\left(Y_f|Y,Z\right) \propto \left|B_3 + \left(Y_f - C_f\hat{\theta}^*\right)^T\left[I_{n_f} + C_f\phi_1^{-1}C_f^T\right]\left(Y_f - C_f\hat{\theta}^*\right)\right|^{-\frac{n_f+n}{2}}$$

$$f\left(Y_f|Y,Z\right) \propto \left|I_k + \left(Y_f - C_f\hat{\theta}^*\right)^T\left[I_{n_f}\right.\right.$$

$$\left.\left. + C_f\phi_1^{-1}C_f^T\right]\left(Y_f C_f\hat{\theta}^*\right) B_3^{-1}\right|^{-\frac{n_f+n}{2}} \qquad (47)$$

And using property of Box and Tiao. (Box, G. P., & Tiao, G. C. 1973)

$$\left|I_k + P_{k\times n_f}\, Q_{n_f\times k}\right| = \left|I_{n_f} + Q_{n_f\times k}P_{k\times n_f}\right|$$

By performing the same steps as when finding the posterior marginal probability distribution of θ conditional by (Z), the kernel of predictive distribution of (Y_f) conditional by (Z) is as follows:

$$f\left(Y_f|Y,Z\right) \propto \left|I_{n_f} + \left(Y_f - C_f\hat{\theta}^*\right) B_3^{-1}\left(Y_f - C_f\hat{\theta}^*\right)^T\left[I_{n_f} + C_f\phi_1^{-1}C_f^T\right]\right|^{-\frac{n_f+n}{2}} \qquad (48)$$

Equation (48) represents the kernel of a matrix-t distribution by degree of freedom $(n - k + 1)$ and the parameters $\left(C_f\hat{\theta}^*, B_3, \left[I_{n_f} + C_f\phi_1^{-1}C_f^T\right]^{-1}\right)$.

The predictive distribution of (Y_f) that is not conditioned by (Z) is as follows:

$$f\left(Y_f|Y\right) = \frac{|B_3|^{-\frac{n_f}{2}}\left|I_{n_f} + C_f\phi_1^{-1}C_f^T\right|^{\frac{k}{2}}\Gamma_{n_f}\left(\dfrac{n+n_f}{2}\right)}{(\pi)^{\frac{n_f k}{2}}\Gamma_{n_f}\left(\dfrac{n+n_f-k}{2}\right)}$$

$$* \left|I_{n_f} + \left(Y_f - C_f\hat{\theta}^*\right) B_3^{-1}\left(Y_f - C_f\hat{\theta}^*\right)^T\left[I_{n_f} + C_f\phi_1^{-1}C_f^T\right]\right|^{-\frac{n_f+n}{2}} \qquad (49)$$

Where:

$$B_3 = \left(Y - C\hat{\theta}^*\right)^T \left(Y - C\hat{\theta}^*\right) \quad , \quad \hat{\theta}^* = (C^T C)^{-1} C^T Y$$

Accordingly, the Bayesian prediction is as follows:

$$E\left(Y_f \middle| Y\right) = C_f \hat{\theta}^* \tag{50}$$

8. The Experimental Sample

In this section, we discuss the simulation of the mechanism reached in the theoretical side to data generated by a suggested method from a multivariate partial linear regression model when random error follows the matrix-variate generalized modified Bessel distribution.

8.1) The Suggested Method for Generating Data

It is difficult to generate random observation from the multivariate partial linear regression model when the random error follows the matrix-variate generalized modified Bessel distribution. Therefore, it is resorted to generate these data through mixed distributions, as the matrix normal variance-mean mixture distribution and the generalized inverse Gaussian distribution previously mentioned were used.

While, random data were generated from the distribution of the multivariate standard normal $\mathbb{Z}$, and since $\mathbb{Z} = \left(\epsilon \middle| Z\right) * \left(Z\Sigma\right)^{-0.5}$ and (ϵ) represents the matrix of random errors of the model from which the observations are to be generated and through the concept of mixed distributions and as follow:

$$\epsilon | Z = \mathbb{Z} * (Z \, \Sigma)^{0.5} \tag{51}$$

$$\epsilon = \int_Z \epsilon | Z \;\; P(Z) \; dZ$$

$$\epsilon = \mathbb{Z} * \Sigma^{0.5} * \frac{K_{\frac{2v+1}{2}}\left(\sqrt{\lambda\psi}\right)}{K_v\left(\sqrt{\lambda\psi}\right)\left(\frac{\lambda}{\psi}\right)^{\frac{1}{4}}} \tag{52}$$

Equation (52) represents the matrix of random errors, which follows the matrix-variate generalized modified Bessel distribution. The following algorithm shows the suggested method for generating random observation from a matrix-variate generalized modified Bessel distribution:

Step1: Assume we have the number of observations ($n = 100$), the number of response variables ($k = 2$).

Step2: Generate random numbers from the multivariate standard normal distribution with (***n***) observations, let the multivariate standard normal random matrix be $\mathbb{Z}$.

Step3: Put $\epsilon | Z = \mathbb{Z} * (Z \, \Sigma)^{0.5}$, $\mathbb{Z}$ represents step (2).

Step4: Find the generated data (ϵ) which represent the random error observations generated from the matrix-variate generalized modified Bessel distribution taking into account the assumed values of the shape parameters$\left(\lambda, \psi, v\right)$ defined in table (3).

Step5: For the purpose of generating data from the multivariate partial linear regression model, we generate the data of the two explanatory variables (p, $s = 2$) for the parametric and non-parametric part (X_1, X_2) and (t_1, t_2) through the following equation:

$$X_j = 2\,\overline{X_j}\,u_j$$

Where u_j represents the standard uniform distribution, $\overline{X_j}$ represents the arithmetic mean and they are usually assumed values, the non-parametric part W represents the kernel weights by represents the Gaussian kernel function and depending on the rule of thumb to choose the bandwidth parameter, the non-parametric variables (t_1, t_2) is a standard normal variable.

Step6: Randomly assumed values are given for (θ) and (Σ) and for shape parameters they are given random values based on the state of the studied distribution $(\lambda, \psi, v) > 0$ and as in table (3) below.

Table (3): Approved default values for all parameters

	λ	ψ	v	$\theta_{(p+s+1) \times k}$	$\Sigma_{k \times k}$
1	5	0.5	2.5	$\begin{bmatrix} 2 & 1.4 & 5 & 3 & 4.1 \\ 1.5 & 3 & 2.5 & 1 & 3.4 \end{bmatrix}^T$	$\begin{bmatrix} 1.5 & 2 \\ 2 & 3 \end{bmatrix}$
2	7		3.5		

Step7: Substituting step (4) and observations of the two explanatory variables for the parametric part, the Gaussian kernel weights matrix for the non-parametric part defined in step (5), the assumed values of the parameters defined in step (6), we obtain (4) models based on the combination between the assumed values of the response matrix (*Y)*.

8.2) Estimation of Model Parameters

The location matrix (θ) and the scale matrix (Σ) were estimated in a Bayes technique when non-informative prior information was available and under the quadratic loss function. The comparison for the estimates was made using the **MSE** and depending on all the combinations between the default values shown in table (3) by using a program Matlab-R2016a.

Table (4): MSE for the estimator of location matrix (*θ)* and scale matrix (Σ).

Models (λ, ψ, v)	Gaussian kernel function $h = 1.06\,\hat{\sigma}\,n^{-\frac{1}{5}}$		Rank	
	θ	Σ	θ	Σ
first$(5, 0.5, 2.5)$	0.0686	**0.0014**	3	**1**
second$(5, 0.5, 3.5)$	0.0739	0.0131	4	3
Third $(7, 0.5, 2.5)$	**0.0510**	0.0065	**1**	2
fourth $(7, 0.5, 3.5)$	0.0557	0.0255	2	4

We notice from table (4) that the best estimator for (θ) and (Σ) it was at the third and first model respectively, this estimate is as follows:

$$\hat{\theta}_B = \begin{bmatrix} 1.8864 & 1.4106 & 4.8197 & 2.9095 & 4.4682 \\ 1.3652 & 3.0134 & 2.3788 & 1.0046 & 3.8588 \end{bmatrix}^T \quad ; \quad \hat{\Sigma}_B = \begin{bmatrix} 1.4743 & 1.9521 \\ 1.9521 & 3.0067 \end{bmatrix}$$

The following figure shows the generated and estimated response variables matrix for third model that are chosen according to the lowest MSE for (θ).
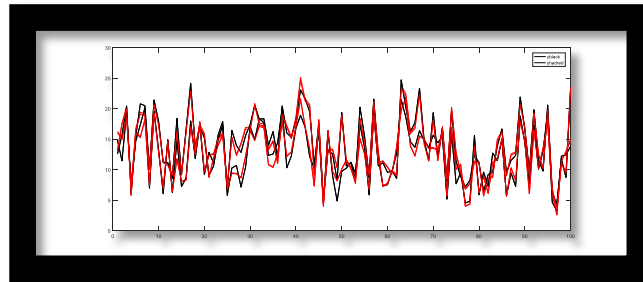

Figure (1): generated and estimated observations of the third model

8.3) Bayesian hypotheses testing

After estimating the parameters of the multivariate partial linear regression model, we usually test the statistical hypotheses, and this section includes a test about the validity of the limitations imposed on the model using the Bayes factor criterion (B.F.) and compares it with the values that definite by the Jeffreys (Jeffreys, H. 1961), as follows:

Table (5): Bayes factor criterion (B.F.) based on non-informative prior information and for the Gaussian kernel function

(λ, ψ, v)	B.F.	The decision
$(5, 0.5, 2.5)$	8.0394×10^{-89}	*Strongly favors* H_1
$(7, 0.5, 2.5)$	3.5478×10^{-88}	*Strongly favors* H_1
$(5, 0.5, 3.5)$	1.8421×10^{-89}	*Strongly favors* H_1
$(7, 0.5, 3.5)$	8.8190×10^{-89}	*Strongly favors* H_1

We notice from tables (5) that the values of the Bayes factor criterion are less than one, and this means that the alternative hypothesis is accepted, that is, the sample was drawn from a population that does not belong to a generalized modified Bessel population.

9. Conclusions and Future Works

A multivariate partial regression model is used when the error follows the matrix –variate generalized modified Bessel distribution as an alternative to the model in which the error follows the matrix normal distribution to find the Bayesian estimations of the model parameters. It found that the posterior marginal probability distribution of the location matrix θ follows the matrix-t distribution by the

parameters $(\hat{\theta}^*, B_3, (C^T C)^{-1}, r = n - k + 1)$ defined in equation (25). The posterior probability distribution of the scale matrix (Σ) is proper distribution defined in equation (27) as well as finding the predictive probability distribution of the matrix of future observations which follows the matrix-t distribution by the parameters $(C_f \hat{\theta}^*, B_3, \left[I_{n_f} + C_f \phi_1^{-1} C_f^T\right]^{-1}, r = n - k + 1)$ defined according to equation (49), Steadfastly the parameters (ψ, v) as the value of the parameter (λ) increases, we get the smaller value of the criterion MSE for the matrix estimator θ, and with its decrease, we get the best Bayesian estimator for Σ through the criterion MSE. Through the Bayes factor criterion, it was found that the sample that was used in the generation process was drawn from a population that does not belong to the generalized modified Bessel population. The two researchers recommend conducting an application side to implement what was reached in the research, depending on the kernel functions and bandwidth parameter defined in Sections (3) and (4) respectively and under different loss functions.

10. References

1. AL-Mouel, A. S., & Mohaisen, A. J. (2017) **"Study on Bayes Semiparametric Regression"** Journal of Advances in Applied Mathematics, Vol. 2, No. 4**,** pages 197-207 **.**
2. Aydin, D., & Tuzemen, M. (2010) **" Estimation in semi-parametric and additive regression using smoothing and regression spline",** Second International Conference on Computer Research and Development, Computer Society.
3. Box, G. P., & Tiao, G. C. (1973)**"Bayesian Inference in Statistical Analysis"** Addison Wesley publishing company, Inc. London, U.K.
4. Choi, T., Lee, J. & Roy, A. (2009) **" A note on the bayes factor in a semiparametric regression model",** Journal of Multivariate Analysis (100), p.p.1316-1327.
5. Gallaugher, M. P.B., & McNicholas, P.D. (2019) **"Three Skewed Matrix Variate Distributions "** Statistics & Probability Letters, Elsevier, vol.145, p.p. 103-109.
6. Hardle, W. (1991)**" Smoothing Technique with Implementation in S"** Spring-Verlage, New York.
7. Jefferys, H. (1961) **"Theory of Probability",** Clarenden press, Oxford, London U.K.
8. Koudou, A. E., & Ley, C. (2014) **" Characterizations of GIG Laws: A survey Complemented with Two New Result "** Proba. Surv., vol. 11, p.p. 161-176.
9. Langrene, N., & Warin, X. (2019)**" Fast and Stable Multivariate Kernel Density Estimation by Fast Sum Updating"** Journal of Computational and Graphical Statistics, vol.28, p.p. 596-608.
10. Lemonte, A. J., & Cordeiro, G. M. (2011) **"The Exponentiated Inverse Gaussian"** Journal of Statistics and Probability letters, vol.81, p.p.506-517, journal homepage: www.elsevier.com.
11. Nagar, D. K., & Gupta, A. K. (2013) **" Extended Matrix Variate Gamma and Beta function"** Journal of Multivariate Analysis, vol. 122, p.p. 53-69.
12. Press, S. J. (2003) **"Subjective and Objective Bayesian Statistics Principles, Models and Application"** John Wiley & Sons, Canada.
13. Przystalski, M. (2014) **"Estimation of the Covariance Matrix in Multivariate Partially Linear Models"** Journal of Multivariate Analysis, Vol.123, pages 380 -385.
14. Schucany, W. R., & Sommers, J. (1977**) "Improved of Kernel Type Density Estimators"** JASA, vol.72, No. 353, p.p. 420-423.
15. Silverman, B.W. (1986) **"Density Estimation for Statistics and Data Analysis"** Chapman and Hall, London.
16. Thabane, L., & Drekic, S. (2003) **" Hypothesis testing for the generalized multivariate modified Bessel model "** Journal of Multivariate Analysis, vol. 86, p.p.360-374.
17. Thabane, L., & Haq, M. S. (2004) **"On the Matrix-variate Generalized Hyperbolic Distribution and its Bayesian Applications "** Journal of Theoretical and Applied Statistical Science, vol. 38(6), p.p.511-526.
18. You, J., et al (2013) "**Statistical Inference for Multivariate Partially Linear Regression Models"** the Canadian journal of statistics, vol.41, No.1, pages1-22.

الاستدلال البيزي لانموذج الانحدار الخطي الجزئي متعدد المتغيرات غير الطبيعي

سرمد عبدالخالق   **صالح عماد حازم عبودي**

المستخلص

تم في هذا البحث التقدير البيزي لمعلمات انموذج الانحدار الخطي الجزئي متعدد المتغيرات عندما يتوزع الخطأ العشوائي توزيع مصفوفة بسل المحور المعمم وايجاد المختبر الاحصائي للانموذج والمتمثل بايجاد معيار عامل بيز والتوزيع التنبؤي بافتراض أن تكون معلمات الشكل معلومة. تمثلت المعلومات حول التوزيع السابق لمعلمات الانموذج بمعلومات قليلة.

وقد تمت محاكاة البيانات المولدة من الانموذج بطريقة مقترحة اعتماداً على قيم مختلفة لمعلمات الشكل، وان دالة النواة المستخدمة في التوليد كانت دالة نواة طبيعية، وان معلمة عرض الحزمة (التمهيد) كانت وفقاً لقاعدة الابهام (قاعدة التوزيع الطبيعي). واستنتج الباحثان أن التوزيع الاحتمالي الهامشي اللاحق لمصفوفة الموقع θ وتوزيع التنبؤ البيزي هو توزيع (Matrix–t) ولكن بمعلمات مختلفة وأن التوزيع الاحتمالي الهامشي اللاحق لمصفوفة القياس Σ هو توزيع مناسب ولكن لاينتمي الى العائلة المتآلفة ومن خلال معيار عامل بيز تبين بان العينة التي استخدمت في عملية التوليد سحبت من مجتمع لاينتمي الى مجتمع بسل المحور المعمم.

الكلمات المفتاحية: انموذج الانحدار الخطي الجزئي متعدد المتغيرات، توزيع مصفوفة بسل المحور المعمم، دوال اللب، معلمة عرض الحزمة، اسلوب بيز .