IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Bayesian Technique to Perform the Prediction Process for a Multivariate Mixed Compact Regression Model

Sarmad Abdulkhaleq Salih [1], Omar Ramzi Jassim [2], Emad Hazim Aboudi [3] and Qutaiba N. Nayef AL-Qazaz[4]

[1] Department of Mathematics, College of Education for Pure Sciences, University of Hamdaniya, Mosul, Iraq.
[2]Department of Accounting, College of Administration and Economics, University of Hamdaniya, Mosul, Iraq.
[3,4]Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq.

ARTICLE INFO

Article history:
Received 21/10/2024
Revised 21/10/2024
Accepted 13/01/2025
Available online 15/05/2025

Keywords:
Multivariate compact regression model,
Bayesian technique,
Matrix-variate extension hyperbola distribution,
Kernel function,
Smoothing parameter.

ABSTRACT

Matrix-variate extension hyperbola distribution (m-vehd) belongs to the family of probability distributions with heavy tails. It is considered a mixed continuous probability distribution and a twisted probability distribution. It is the result of mixing the matrix-varite Gaussian variance-mean mixture distribution with the generalized inverse normal distribution (gind). And this distribution has wide applications in the field of economics. On this basis, the paper will study a multivariate compact regression model that follows a (m-vehd).

Assuming that the shape parameters, scale matrix, and the twisted matrix are known, the parameters of the multivariate compact regression model will be estimated using the Bayesian technique depending on informative prior information. In addition, the smoothing parameter is selected by a normal distribution rule (rule of thumb) and the kernel function based on the Gauss kernel function and Quartic kernel function, and then finding Bayesian predictive distributions based on informative prior information and estimating the model parameters under balanced and unbalanced loss functions, and application to real data related by the reality of financial inclusion in Arab countries for the year 2014. The researchers concluded the superiority of the Bayes estimator under the balanced quadratic loss function at weight 0.75 and for Gauss kernel function. In addition, the predictive distribution of future observations is an uncommon distribution, but it is an appropriate distribution.

1. Introduction:

Multivariate regression models are statistical models that have great importance in different areas of life, especially in economic fields. Among these models is the multivariate compact regression model. Most of the economic models are compact regression, as is known from their name, and they are a mixture of multivariate parametric (linear) regression model and multivariate nonparametric (nonlinear) regression model; the parametric part is the regression function that is supposed to be linear in the observations of its explanatory variables, which explains some important economic phenomena, while the nonparametric part is an unknown smoothing function and is a nonlinear function, the multivariate compact regression model provides an intermediate solution between the linear and nonlinear model.

Therefore, many researchers have been interested in estimating multivariate compact regression models in which the random error follows a matrix-variate Gaussian distribution, but there are cases in which the random error follows a heavy tails distribution or the error observations are not independent

* Corresponding author.E-**mail** address: dr.qutaiba@coadec.uobaghdad.edu.iq
https://doi.org/10.62933/q222v981

and uncorrelated, in the two cases above, it is more appropriate to pay attention to alternative probability distributions than the matrix-variate Gaussian distribution, meaning that the mixture distributions are more fits as the (m-vehd).

(Thabane & Haq, 2004) presented a generalization the multivariate extension hyperbola distribution to the (m-vehd) as a mixed distribution resulting from the distribution of the matrix-variate Gaussian variance-mean mixture with the matrix generalized inverse normal distribution. They also studied some of its properties and special cases. They studied the Bayesian approach of the multivariate Gaussian linear regression model, assuming that the prior distribution of the scale matrix is a matrix generalized inverse normal distribution. Found **(Thabane & Kibria, 2007)** the Bayesian prediction of the multiple general regression model in the case of equal correlation when the random error follows a extension multivariate transmuted Bessel distribution (symmetric multivariate extension hyperbola distribution), as it was assumed that the prior probability function for the scale parameter follows the generalized inverse normal distribution. The prediction distribution was obtained that follows a symmetric multivariate extension hyperbola distribution. This result contrasts the T-distribution obtained using the prior inverse chi-square distribution for the scale parameter. Study **(AL-Mouel & Mohaisen,2017)** Bayesian estimation based on the MCMC algorithm for the Gaussian multiple compact regression model represented by the Gaussian multiple compact linear regression model with the conditional ratio between the scale of the parametric part on the scale of the model error, and the scale of the nonparametric part on the scale of the model error. The weights matrix was the penalized spline and the statistical laboratory formation based on the criterion Bayes factor and the application of the findings to experimental data, two nonparametric functions. They concluded that the values generated from the data do not belong to a Gaussian society for different samples. Estimated **(Hmood & Hassan, 2020)** the multiple compact linear regression model when random error term is distributed normally with a mean of zero and variance σ^2 using two smoothing methods, namely, wavelet smoothing and kernel smoothing, which the researchers concluded through the experimental study and under different functions and sample sizes that show the wavelet smoother is better than the kernel smoother based on the (MASE) criterion.

The paper was divided into seven sections. First section included the introduction. Second section dealt with the description of the multivariate mixed compact regression model. The third section discusses some kernel functions and the smoothing parameter. Using Bayesian technique to estimate the parameters of a multivariate mixed compact regression model depending on informative prior information, and under different loss functions in the fourth section. Finding the Bayesian predictive distribution based on informative prior information in the fifth section. While the sixth section dealt with an applied study as the application was made on real data related by the reality of financial inclusion in Arab countries for the year 2014, and the study analysis process is carried out based on the MATLAB R2022a program. The seventh section shows the most important conclusions.

2. Multivariate Mixed Compact Regression Model:

Multivariate compact linear regression model is more important models symbolized by the symbol (MCLM), and it is considered a special case of additive models. It is one of the models that depend on linear parametric and nonparametric variables. Usually, the variables of this model are continuous. These linear and nonlinear variables affect the response variable. It is a generalization of standard linear regression techniques, so it is better than nonlinear models because it avoids the problem of dimensionality.(Przystalski, 2014)

The multivariate compact linear regression model is described according to the following equation :(You et al., 2013)

$$Y_{ij} = X_i'\beta_j + m_j(T_i) + \epsilon_{ij} \quad i = 1,2,\dots,n \quad ,j = 1,2,\dots,k \qquad \dots(1)$$

Since $X_i'\beta_j$ represents the linear part of the model, which β_j is estimated by one of the parametric methods, such as the method of ordinary least squares, the maximum likelihood, moments, or Bayes ..., and $m_j(T_i)$ represents the nonlinear part of the model, which is an unknown smoothing function that is estimated by one of the nonparametric methods, such as the Nadaraya-Watson smoother, Priestley-Chao, Gasser-muller smoother, kernel smoother, local linear regression smoother, spline smoother ....., It is possible to write the model defined in equation (1) in the form of matrices as follows: (AL-Mouel & Mohaisen, 2017)

$$Y = X\theta + W\delta + \epsilon \qquad \dots(2)$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix}_{n\times k}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times p}$$

$$\theta = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1k} \\ \vdots & \ddots & \vdots \\ \theta_{p1} & \cdots & \theta_{pk} \end{bmatrix}_{p\times k} ,$$

$$W = \begin{bmatrix} k_h(t_1 - T_{11}) & \cdots & k_h(t_s - T_{1s}) \\ \vdots & \ddots & \vdots \\ k_h(t_1 - T_{n1}) & \cdots & k_h(t_s - T_{ns}) \end{bmatrix}_{n\times s}$$

$$\delta = \begin{bmatrix} \delta_{11} & \cdots & \delta_{1k} \\ \vdots & \ddots & \vdots \\ \delta_{s1} & \cdots & \delta_{sk} \end{bmatrix}_{s\times k} ,$$

$$\epsilon = \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1k} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{nk} \end{bmatrix}_{n\times k}$$

Where Y is a Matrix of response variables of degree ($n \times k$) and $\boldsymbol{n}$ represents the number of observations, and $\boldsymbol{k}$ represents the number of response variables. X is a non-random matrix representing the observations of the parametric explanatory variables of degree ($n \times p$) and $\boldsymbol{p}$ represents the number of the parametric explanatory variables. θ is a matrix of model parameters for the parametric (linear) part of the degree ($p \times k$). W is a 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 degree ($n \times s$), and $\boldsymbol{s}$ represents the number of nonparametric explanatory variables, and $k_h(x)$ represents the kernel function and as follows: (Hmood ,2005)

$$k_h(x) = \frac{1}{h}\, k\left(\frac{x}{h}\right)$$

And that this function is a real, symmetric, and continuous function and that $\boldsymbol{h}$ represent the smoothing parameter, they will be mentioned later. δ is a matrix of parameters of the nonparametric (nonlinear) part (additive parameters) of degree ($s \times k$). ϵ is a matrix of random errors of degree ($n \times k$).

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

$$Y_{n\times k} = Z_{n\times(p+s)}\vartheta_{(p+s)\times k} + \epsilon_{n\times k} \qquad \dots(3)$$

As: $Z = [X \quad W]$, $\vartheta = \begin{bmatrix}\theta \\ \delta\end{bmatrix}$

We assume that the matrix of random errors (**ϵ**) follows a (m-vehd), that is, the matrix of errors is distributed of a non-zero matrix mean. The probability density function of the matrix (**ϵ**) can be found using the concept of mixed distributions from the matrix-variate Gaussian variance-mean mixture distribution with the generalized inverse normal distribution and as follows:

(Gallaugher & McNicholas, 2019) (Thabane & Haq, 2004)

$$\epsilon|\tau \sim M - VN_{n,k}(\gamma\tau, \tau\Sigma, I_n) \quad , \quad \tau \sim GIN(\alpha, \varphi, r)$$

As the probability density function of the matrix of random errors conditional by (τ) takes the following formula:

$$f(\epsilon|\tau) = \frac{\left(\Gamma\left(\frac{1}{2}\right)\right)^{-nk}}{(2\,\tau)^{\frac{nk}{2}}\,|\Sigma|^{\frac{n}{2}}}\; e^{-\frac{1}{2\tau} tr\,(\epsilon-\gamma\tau)^T(\epsilon-\gamma\tau)\Sigma^{-1}} \qquad \dots(4)$$

Equation (4) represents the matrix-variate Gaussian variance-mean mixture distribution, and the probability density function of the variable τ defined in the following equation:

$$P(\tau) = \frac{(\alpha)^{\frac{r}{2}}\ \tau^{-(1-r)}}{2\,(\varphi)^{\frac{r}{2}}\, H_r(sqrt(\alpha\varphi))}\ exp\left[-\frac{1}{2}\left(\left(\frac{\varphi}{\tau}\right) + \alpha\tau\right)\right] \quad ,\tau > 0 \qquad \dots(5)$$

Whereas $(\alpha, \varphi)$are scale parameters and r is shape parameter.$H_r(.)$ is a represents the transmuted Bessel function of the third kind of order r, and its formula is as follows:(Gallaugher & McNicholas, 2019) (Mora & Mata, 2013)

$$H_r\left(sqrt(\alpha\varphi)\right) = \int_0^{\infty} \frac{t^{-(1-r)}}{2\exp\left(0.5\, sqrt(\alpha\varphi)(t+t^{-1})\right) dt}, sqrt(\alpha\varphi) > 0 \quad \dots (6)$$

According to the concept of mixed distribution, the probability distribution of the unconditional error matrix is as follows:

$$f(\epsilon) = \int_0^{\infty} f(\epsilon|\tau)\, P(\tau) \quad d\tau$$

$$f(\epsilon) = \frac{(2)^{\frac{-nk}{2}} (\alpha)^{\frac{nk}{4}} \left(H_r(sqrt(\alpha\varphi))\right)^{-1} \left(1+\frac{tr\, \epsilon^T \epsilon\Sigma^{-1}}{\varphi}\right)^{\frac{2r-nk}{4}}}{(\varphi)^{\frac{nk}{4}} \left(\Gamma(\frac{1}{2})\right)^{nk} |\Sigma|^{\frac{n}{2}}\, e^{-tr\,(\epsilon)^T \gamma\, \Sigma^{-1}} \left(1+\frac{tr\, \gamma^T \gamma\, \Sigma^{-1}}{\alpha}\right)^{\frac{2r-nk}{4}}} * H_{\frac{2r-nk}{2}}\left(sqrt\left((\alpha\varphi + \alpha\, tr\, \epsilon^T \epsilon\Sigma^{-1})(\alpha\varphi + \varphi\, tr\, \gamma^T \gamma\, \Sigma^{-1})\right)\right) \dots (7)$$

As:

α, φ, r: represent shape parameters.

γ: twisted matrix of degree ($n \times k$).

Σ: scale matrix of degree ($n \times k$).

Equation (7) represents the (m-vehd) for the random error matrix, which is described as follows:

$$vec(\epsilon) \sim m - vehd_{nk}(vec(0), \Sigma \otimes I_n, \alpha, \varphi, r, vec(\gamma))$$

Since the matrix Y defined in equation (3) is a linear combination in terms of the matrix ϵ that follows the (m-vehd). Accordingly, the probability distribution of the response observations matrix Y follows the (m-vehd). It can be found in the same way as follows: (Thabane & Haq, 2004)

The probability density function of the matrix of response variables conditional by τ $(Y|\tau)$ that follows the matrix-variate Gaussian variance-mean mixture distribution is as follows:

$$f(Y|\tau) = \frac{\left(\Gamma\left(\frac{1}{2}\right)\right)^{-nk}}{(2\,\tau)^{\frac{nk}{2}} |\Sigma|^{\frac{n}{2}}}\, e^{-\frac{1}{2\,\tau} tr\,(Y - Z\vartheta - \gamma\tau)^T (Y - Z\vartheta - \gamma\tau)\Sigma^{-1}} \quad \dots (8)$$

Depending on the concept of mixed distributions, the probability distribution of Y unconditional by τ is as follows:

$$f(Y) = \frac{(2)^{\frac{-nk}{2}} (\alpha)^{\frac{nk}{4}} \left(H_r(sqrt(\alpha\varphi))\right)^{-1} \left(1+\frac{tr\,(Y-Z\vartheta)^T (Y-Z\vartheta)\Sigma^{-1}}{\varphi}\right)^{\frac{2r-nk}{4}}}{(\varphi)^{\frac{nk}{4}} \left(\Gamma(\frac{1}{2})\right)^{nk} |\Sigma|^{\frac{n}{2}}\, e^{-tr\,(Y-Z\vartheta)^T \gamma\, \Sigma^{-1}} \left(1+\frac{tr\, \gamma^T \gamma\, \Sigma^{-1}}{\alpha}\right)^{\frac{2r-nk}{4}}} * H_{\frac{2r-nk}{2}}\left(sqrt\left((\alpha\varphi + \alpha\, tr\,(Y-Z\vartheta)^T (Y-Z\vartheta)\Sigma^{-1})(\alpha\varphi + \varphi\, tr\, \gamma^T \gamma\, \Sigma^{-1})\right)\right) . (9)$$

This distribution can be expressed descriptively as follows:

$$vec(Y) \sim m - vehd_{nk}(vec(Z\vartheta), \Sigma \otimes I_n, \alpha, \varphi, r, vec(\gamma))$$

3. Kernel Functions and Smoothing Parameter:

Kernel functions) are used to estimate regression, spectral and probability density functions, 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. That the method of selecting the smoothing parameter (h) is an essential part in estimating the nonparametric regression curve, and that choosing the smoothing parameter is more important than choosing the kernel function and there are several labels for this parameter, including (constraint capacity - bandwidth - concentration parameter - contrast parameter) and its properties are a non-random, symmetrical and positive boundary parameter. Table 1 shows the kernel functions used in the paper as well as the selection of the smoothing parameter based on the thumb rule method.

(Langrene & Warin, 2019) (Hmood, 2005)

Table 1: Some Kernel Functions and Smoothing Parameter.

kernel	$k(x)$		$h = \hat{s}\, C(k)\, n^{-\frac{1}{5}}$
Quartic	$(15/16)(1-y^2)^2$	$I(\lvert y\rvert \le 1)$	$C(k) = 2.78$

Gauss	$(2\pi)^{-0.5}\exp(-y^2/2)$	$I(\|y\| < \infty)$	$C(k) = 1.06$

4. Bayesian Technique to Estimate of a Multivariate Mixed Compact Regression Model Under Different Loss Functions:

In this section, the parameters of the multivariate compact regression model defined by equation (3) are estimated when the random error follows a (m-vehd) based on informative prior information; that is, it has probability distributions that belong to the conjugate family and as follows:

Depending on mixed distributions, the posterior probability distribution of location matrix (ϑ) unconditional of τ is as follows: (Salih & Aboudi , 2022)

We know that the prior distribution of $(\vartheta|\Sigma,\tau)$ has the following form:

$$P(\vartheta|\,\Sigma,\tau) \propto exp\left(-\frac{1}{2\,\tau}\,tr\,(\vartheta-\vartheta_0)^T\,{\Psi_0}^{-1}(\vartheta-\vartheta_0)\Sigma^{-1}\right) \quad \ldots(10)$$

The posterior probability distribution of the matrix $(\vartheta|\,Y,\Sigma,\tau)$is the distribution resulting from merging the equation (8) with the equation (2.124), by adding and subtracting the amount $Z\hat{\vartheta}_{m|\tau}$ and that $\hat{\vartheta}_{m|\tau}$ represents the conditional maximum likelihood estimator:

$$\hat{\vartheta}_{m|\tau} = (Z^TZ)^{-1}Z^TY - (Z^TZ)^{-1}Z^T\,\gamma\,\tau \quad \ldots(11)$$

And by making some mathematical simplifications, we get the following:

$$P(\vartheta|Y,\Sigma,\tau) \propto P(\vartheta|\,\Sigma,\tau)\,f(Y|\vartheta,\Sigma,\tau)$$

$$P(\vartheta|Y,\Sigma,\tau) \propto\; exp\left[-\frac{1}{2\,\tau}\;tr\,(\vartheta-\vartheta^*)^T\left[{\Psi_0}^{-1}+(Z^TZ)\right](\vartheta-\vartheta^*)\Sigma^{-1}\right] \ldots(12)$$

$$\vartheta^* = \left[{\Psi_0}^{-1}+(Z^TZ)\right]^{-1}\left[{\Psi_0}^{-1}\vartheta_0 + Z^TY - Z^T\,\gamma\tau\,\right] \quad \ldots(13)$$

$$P(\vartheta|Y,\Sigma) = \int_0^\infty P(\vartheta|Y,\Sigma,\tau)\,P(\tau)\,d\tau$$

$$P(\vartheta|Y,\Sigma) = \frac{(2\,)^{\frac{(p+s)k}{2}}(\alpha)^{\frac{(p+s)k}{4}}\left|{\Psi_0}^{-1}+Z^TZ\right|^{\frac{k}{2}}\left(H_r(sqrt(\alpha\varphi))\right)^{-1}}{(\varphi)^{\frac{(p+s)k}{4}}\left(\Gamma\left(\frac{1}{2}\right)\right)^{(p+s)k}\,|\Sigma|^{\frac{(p+s)}{2}}\;e^{-\,tr\,(\vartheta-Q)^T\,Z^{*T}(-\gamma)\Sigma^{-1}}}$$

$$* H_{\frac{2r-(p+s)k}{2}}\left(\sqrt{(\alpha\varphi+\alpha tr\,Q^*\Sigma^{-1})(\alpha\varphi+\varphi\,tr\,\gamma^TZ^*\,Z^{*T}\gamma\,\Sigma^{-1})}\right) * (1+\varphi^{-1}tr\,Q^*\Sigma^{-1})^{\frac{2r-(p+s)k}{4}}\left(1+\alpha^{-1}tr\,\gamma^TZ^*\,Z^{*T}\gamma\,\Sigma^{-1}\right)^{\frac{-2r+(p+s)k}{4}} (14)$$

As:

$$Q^* = (\vartheta-Q)^T\left[{\Psi_0}^{-1}+(Z^TZ)\right](\vartheta-Q)$$

$$Q = \left[{\Psi_0}^{-1}+(Z^TZ)\right]^{-1}\left[{\Psi_0}^{-1}\vartheta_0+(Z^TZ)\,\hat{\vartheta}^*\,\right]\;,\quad \hat{\vartheta}^* = (Z^T\,Z)^{-1}Z^T\,Y$$

$$Z^{*T} = \left[{\Psi_0}^{-1}+(Z^TZ)\right]^{-1}Z^T$$

Equation (10) represents the (m-vehd) and is described as follows:

$$vec(\theta)\sim-vehd_{\,(p+s)k}\left(vec(Q)\,,\Sigma\otimes\left({\Psi_0}^{-1}+(Z^TZ)\right)^{-1},\alpha,\varphi,r,vec(\,Z^{*T}(-\gamma))\right)$$

The quadratic loss function is described according to equation below:

$$l(\hat{\vartheta}\,,\vartheta) = tr\,(\hat{\vartheta}-\vartheta)^T\,\Delta\,(\hat{\vartheta}-\vartheta) \quad \ldots(15)$$

and under the assumption that the weight matrix (Δ) is an identity matrix with dimension $((p+s)\times(p+s))$. Therefore; the quadratic loss function of the (ϑ) is written in the following formula:

$$l_q(\hat{\vartheta}\,,\vartheta) = tr\,(\hat{\vartheta}-\vartheta)^T\,I\,(\hat{\vartheta}-\vartheta) \quad \ldots(16)$$

As for the quadratic risk function of the matrix (ϑ), it is represented by the mathematical expectation of the quadratic loss function, and as follows:

$$R_q(\hat{\vartheta}\,,\vartheta) = E\left[l_q(\hat{\vartheta}\,,\vartheta)\right]$$

$$= \int_\vartheta l_q(\hat{\vartheta}\,,\vartheta)\,P(\vartheta|Y)\,d\vartheta \quad \ldots(17)$$

The Bayesian estimator for (ϑ) is found, which makes the quadratic risk function $R_q(\hat{\vartheta}\,,\vartheta)$ at its lower limit, and after substituting equation (15) into equation (16), taking the partial derivative relative to ($\hat{\vartheta}$) and equal the derivative by zero, we get the following:

$$\frac{\partial R_q(\hat{\vartheta}\,,\vartheta)}{\partial\hat{\vartheta}^T} = \int_\vartheta \frac{\partial}{\partial\hat{\vartheta}^T}tr\,(\hat{\vartheta}-\vartheta)^T\,I\,(\hat{\vartheta}-\vartheta)\,P(\vartheta|Y)\,d\vartheta$$

$$\frac{\partial R_q(\hat{\vartheta},\vartheta)}{\partial \hat{\vartheta}^T} = \int_{\vartheta} (2\hat{\vartheta} - 2\vartheta)\, P(\vartheta|Y)\, d\vartheta$$

After equality by zero, we get the Bayesian estimator:

$$\hat{\vartheta}_q = \int_{\vartheta} \vartheta\, P(\vartheta|Y)\, d\vartheta$$

$$\hat{\vartheta}_q = E_{\vartheta}(\vartheta|Y) \quad \ldots(18)$$

We conclude from equation (18) that the Bayesian estimator under the quadratic loss function represents the mean of the posterior probability distribution of (θ).

$$\hat{\vartheta}_q = \left[\Psi_0^{-1} + (Z^T Z)\right]^{-1} \left[\Psi_0^{-1}\vartheta_0 + Z^T Y - Z^T \gamma * \frac{H_{r+1}(\sqrt{\alpha\varphi})}{H_r(\sqrt{\alpha\varphi})}\left(\frac{\alpha}{\varphi}\right)^{-0.5}\right] \ldots(19)$$

To find the quadratic risk function:

$$R_q(\hat{\vartheta},\vartheta) = \int_{\theta} tr\,(\hat{\vartheta} - \vartheta)^T(\hat{\vartheta} - \vartheta)\, P(\vartheta|Y)\, d\vartheta$$

$$R_q(\hat{\vartheta},\vartheta) = tr\, E\,\left[(\hat{\vartheta} - \vartheta)^T\,(\hat{\vartheta} - \vartheta)\right] \ldots(20)$$

We notice from equation (20) that the quadratic risk function represents the variance of the posterior probability distribution of (ϑ).

The general form of the quadratic balanced loss function is known as: (Jozani et al., 2012)

$$l_{qb}(\hat{\vartheta},\vartheta) = w\, tr(\hat{\vartheta} - \hat{\vartheta}_m)^T(\hat{\vartheta} - \hat{\vartheta}_m) + (1-w)\, tr\,(\hat{\vartheta} - \vartheta)^T(\hat{\vartheta} - \vartheta) \ \ldots(21)$$

Whereas:

$\hat{\theta}_m$: The unconditional maximum likelihood estimator is defined in equation below:

$$\hat{\vartheta}_m = (Z^T Z)^{-1} Z^T Y - (Z^T Z)^{-1} Z^T \gamma * \frac{H_{r+1}(\sqrt{\alpha\varphi})}{H_r(\sqrt{\alpha\varphi})}\left(\frac{\alpha}{\varphi}\right)^{-0.5} \quad \ldots(22)$$

w: Weight value, $0 \le w \le 1$.

The quadratic balanced risk function of the parameter matrix (ϑ) is as follows:

$$R_{qb}(\hat{\vartheta},\vartheta) = E\left[l_{qb}(\hat{\vartheta},\vartheta)\right]$$

$$= \int_{\theta} l_{qb}(\hat{\vartheta},\vartheta)\, P(\vartheta|Y)\, d\vartheta \quad \ldots(26)$$

The Bayesian estimator for (ϑ) is found, making the quadratic balanced risk function at a lower limit. We partially differentiate the equation (26) relative to ($\hat{\vartheta}$) and set it equal to zero, we get the following:

$$\frac{\partial R_{qb}(\hat{\vartheta},\vartheta)}{\partial \hat{\vartheta}^T} = \int_{\vartheta} \frac{\partial}{\partial \hat{\vartheta}^T}\left[w\, tr(\hat{\vartheta} - \hat{\vartheta}_m)^T(\hat{\vartheta} - \hat{\vartheta}_m) + (1-w)\, tr\,(\hat{\vartheta} - \vartheta)^T(\hat{\vartheta} - \vartheta)\right] P(\vartheta|Y)\, d\vartheta$$

$$\frac{\partial R_{qb}(\hat{\vartheta},\vartheta)}{\partial \hat{\vartheta}^T} = \int_{\vartheta} \frac{\partial}{\partial \hat{\vartheta}^T}\left[w\ \left\{tr\,\hat{\vartheta}^T\hat{\vartheta} - 2tr\hat{\vartheta}^T\,\hat{\vartheta}_m + tr\hat{\vartheta}_m{}^T\hat{\vartheta}_m\right\} + (1-w)\ \left\{tr\hat{\vartheta}^T\hat{\vartheta} - 2tr\hat{\vartheta}^T\,\vartheta + tr\vartheta^T\vartheta\right\}\right] P(\vartheta|Y)\, d\vartheta$$

$$\frac{\partial R_{qb}(\hat{\vartheta},\vartheta)}{\partial \hat{\vartheta}^T} = \int_{\vartheta} w\ \{2tr\hat{\vartheta} - 2\,tr\hat{\vartheta}_m\} + (1-w)\ \{2tr\hat{\vartheta} - 2tr\vartheta\}\, P(\vartheta|Y)\, d\vartheta$$

Solve the equation ($\frac{\partial R_{qb}(\hat{\vartheta},\vartheta)}{\partial \hat{\vartheta}^T} = 0$) and doing the integration relative to (ϑ) we get the Bayes estimator under the quadratic balanced loss function, and as follows:

$$\hat{\vartheta}_{qb} = w\,\hat{\vartheta}_m + (1-w)E_{\vartheta}(\vartheta|Y) \quad \ldots(27)$$

To find the quadratic balanced risk function:

$$R_{qb}(\hat{\vartheta},\vartheta) = \int_{\vartheta}\left[w\, tr(\hat{\vartheta} - \hat{\vartheta}_m)^T(\hat{\vartheta} - \hat{\vartheta}_m) + (1-w)\, tr\,(\hat{\vartheta} - \vartheta)^T(\hat{\vartheta} - \vartheta)\right] P(\vartheta|Y)\, d\vartheta$$

$$R_{qb}(\hat{\vartheta},\vartheta) = w\, tr(\hat{\vartheta} - \hat{\vartheta}_m)^T(\hat{\vartheta} - \hat{\vartheta}_m) + (1-w) tr\, E\,\left[(\hat{\vartheta} - \vartheta)^T\,(\hat{\vartheta} - \vartheta)\middle| Y\right] \ldots(28)$$

As:

$E\,\left[(\hat{\vartheta} - \vartheta)^T\,(\hat{\vartheta} - \vartheta)\middle| Y\right]$: Variance of the posterior probability distribution of (ϑ).

5. Bayesian Predictive Distribution:

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 compact regression model is as follows: (Thabane & Haq, 2004)

$$Y_f = Z_f\vartheta + \epsilon_f \qquad \dots (29)$$

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

Z_f is a matrix with dimension $(n_f \times (p + s))$. ϑ is a parameter matrix with dimension $((p + s) \times k)$. ϵ_f is a matrix of future random errors of degree $(n_f \times k)$.

Since the error matrix (ϵ_f) follows a (m-vehd) with the parameters $\left(0, \Sigma, I_{n_f}, \alpha, \varphi, r, \gamma_f\right)$, we know that (Y_f) is a linear combination in terms of the future error matrix, therefore (Y_f) follows a (m-vehd) of the parameters $\left(Z_f\vartheta, \Sigma, I_{n_f}, \alpha, \varphi, r, \gamma_f\right)$.

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

$$f\left(Y_f \middle| Y\right) = \int_{\vartheta} f\left(Y_f \middle| \vartheta, \Sigma\right) \, P(\vartheta|Y) d\vartheta \qquad \dots (30)$$

Due to the difficulty of finding a predictive distribution from equation (30), we use the concept of mixed distributions, that is, probability distribution conditioned by the random variable (τ).

$$f\left(Y_f \middle| Y\right) = \int_{\tau} \int_{\vartheta} f\left(Y_f \middle| \vartheta, \Sigma, \tau\right) \, P(\theta|Y, \tau) \, P(\tau) \, d\vartheta \, d\tau \qquad \dots (31)$$

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

$$f\left(Y_f \middle| \theta, \Sigma, \tau\right) = \frac{1}{(2\pi\tau)^{\frac{n_f k}{2}} |\Sigma|^{\frac{n_f}{2}}} e^{-\frac{1}{2\tau} tr\left(Y_f - Z_f\vartheta - \gamma_f Z\right)^T \left(Y_f - Z_f\vartheta - \gamma_f Z\right)\Sigma^{-1}} \qquad \dots (32)$$

The posterior probability distribution of (ϑ) conditional by (τ) defined in equation below:

$$P(\vartheta|Y, \tau) = con * e^{-\frac{1}{2\tau} tr\,(\vartheta - \vartheta^*)^T \left[\Psi_0^{-1} + (Z^T Z)\right](\vartheta - \vartheta^*)\Sigma^{-1}} \qquad \dots (33)$$

As:

$$con = \frac{\left|\Psi_0^{-1} + (Z^T Z)\right|^{\frac{k}{2}}}{(2\pi\tau)^{\frac{(p+s)k}{2}} |\Sigma|^{\frac{(p+s)}{2}}} \, , \vartheta^* = \left[\Psi_0^{-1} + (Z^T Z)\right]^{-1} \left[\Psi_0^{-1}\vartheta_0 + Z^T Y - Z^T \gamma\tau\right]$$

Combined equation (33) and equation (26), we get the kernel of the predictive distribution of (Y_f) conditional by the random variable (τ), and as follows:

$$f\left(Y_f \middle| Y\right) \propto \int_{\tau} \int_{\vartheta} e^{-\frac{1}{2\tau} tr\left(Y_f - Z_f\vartheta - \gamma_f\tau\right)^T \left(Y_f - Z_f\vartheta - \gamma_f\tau\right)\Sigma^{-1}} * e^{-\frac{1}{2\tau} tr\,(\vartheta - \vartheta^*)^T \left[\Psi_0^{-1} + (Z^T Z)\right](\vartheta - \vartheta^*)\Sigma^{-1}} P(\tau) \, d\vartheta \, d\tau \qquad \dots (34)$$

We assume that:

$$B = \left(Y_f - Z_f\vartheta - \gamma_f\tau\right)^T \left(Y_f - Z_f\vartheta - \gamma_f\tau\right) + (\vartheta - \vartheta^*)^T \left[\Psi_0^{-1} + (Z^T Z)\right](\vartheta - \vartheta^*)$$

Adding and subtracting $(Z_f\vartheta^*)$ to the parentheses of the first term of (B) and performing some of the algebraic operations we get:

$$B = \left(Y_f - Z_f\vartheta^* - \gamma_f\tau\right)^T \left[I_{n_f} + Z_f \phi_3^{-1} Z_f^{\,T}\right] \left(Y_f - Z_f\vartheta^* - \gamma_f\tau\right) + (\vartheta - \vartheta^* + \phi_4)^T \phi_3 \, (\vartheta - \theta^* + \phi_4)$$

As:

$$\phi_3 = Z_f^{\,T} Z_f + \left[\Psi_0^{-1} + (Z^T Z)\right]$$

$$\phi_4 = \phi_3^{-1} Z_f^{\,T} \left(Y_f - Z_f\vartheta^* - \gamma_f\tau\right)$$

$$f\left(Y_f \middle| Y\right) = \frac{\left|\Psi_0^{-1} + (Z^T Z)\right|^{\frac{k}{2}}}{(2\pi)^{\frac{n_f k}{2}} |\Sigma|^{\frac{n_f}{2}} |\phi_3|^{\frac{k}{2}}} * \int_{\tau} (\tau)^{-\frac{n_f k}{2}} e^{-\frac{1}{2\tau} tr\left(Y_f - Z_f\vartheta^* - \gamma_f\tau\right)^T \left[I_{n_f} + Z_f\phi_3^{-1} Z_f^{\,T}\right] * \left(Y_f - Z_f\vartheta^* - \gamma_f\tau\right)\Sigma^{-1}} P(\tau) \, d\tau \qquad \dots (35)$$

$$f\left(Y_f \middle| Y\right) = \frac{(2)^{-\frac{n_f k}{2}} (\alpha)^{\frac{n_f k}{4}} \left(H_r(sqrt(\alpha\varphi))\right)^{-1} \left|\Psi_0^{-1} + (Z^T Z)\right|^{\frac{k}{2}}}{|\phi_3|^{\frac{k}{2}} (\varphi)^{\frac{n_f k}{4}} \left(\Gamma(\frac{1}{2})\right)^{n_f k} |\Sigma|^{\frac{n_f}{2}} e^{-tr\left(Y - Z_f\vartheta\right)^T \gamma_f \Sigma^{-1}}} * \left(1 + \frac{tr\left(Y - Z_f\vartheta\right)^T \left(Y - Z_f\vartheta\right)\Sigma^{-1}}{\varphi}\right)^{\frac{2r - n_f k}{4}} \left(1 + \frac{tr\,\gamma_f^{\,T}\gamma_f\Sigma^{-1}}{\alpha}\right)^{-\frac{2r - n_f k}{4}}$$

$$* H_{\frac{2r-n_f k}{2}}\left(sqrt\left(\left(\alpha\varphi + \alpha\, tr\, (Y - Z_f\vartheta)^T\, (Y - Z_f\vartheta)\Sigma^{-1}\right)\left(\alpha\varphi + \varphi\, tr\, {\gamma_f}^T\, \gamma_f\, \Sigma^{-1}\right)\right)\right) \quad \ldots (36)$$

We note that the predictive distribution of the matrix of future observations of the response variable Y_f is not one of the known probability distributions, but it is an appropriate (proper) distribution. So the Bayesian prediction of the response variable Y is found using the following formula:

$$E(Y_f|Y) = E_\tau E_{Y_f|\tau}(Y_f|Y, \tau)$$

$$= Z_f\left[{\Psi_0}^{-1} + (Z^T Z)\right]^{-1}\left[{\Psi_0}^{-1}\vartheta_0 + Z^T Y - Z^T \gamma \frac{H_{r+1}(\sqrt{\alpha\varphi})}{H_r(\sqrt{\alpha\varphi})}\left(\frac{\alpha}{\varphi}\right)^{-0.5}\right] + \gamma_f \frac{H_{r+1}(\sqrt{\alpha\varphi})}{H_r(\sqrt{\alpha\varphi})}\left(\frac{\alpha}{\varphi}\right)^{-0.5} \quad \ldots (36)$$

6. Application Side:

In this section, the application is made to real data related to the reality of financial inclusion in Arab countries for the year 2014, and the study analysis process is carried out based on the MATLAB R2022a program. https://jslem.journals.ekb.eg/article_182829_c01bc6098190b5bd2b70f785cdd0027f.pdf

6.1 Determine study variables and prepare data:

Arab countries are still among the lowest in the world in terms of financial inclusion. According to an International Monetary Fund report, those who have bank accounts with financial institutions represent 18% of the population, and this percentage drops to 13% for women.

On this basis, the paper studied the percentage of adults who borrowed males (Y_1) and females(Y_2), as response variables from financial institutions (X_1)and private lending (X_2) as explanatory variables for the year 2014. The following figure shows the behavior of the percentage of male and female borrowers from financial institutions and private lending.

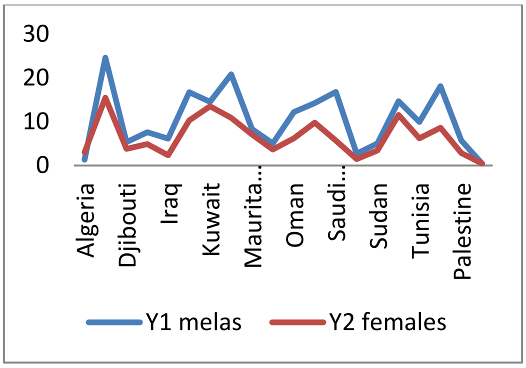


Fig. (1): behavior of the percentage of male and female borrowers from financial institutions and private lending for year 2014.

Before the data matching process was carried out for the model, it was found that the data was skewed to the right, as the value of the skewness coefficient for males (Y_1) was (0.3486) and kurtosis (2.1287), and the value of the skewness coefficient for females(Y_2), as the value of the skewness coefficient was (0.5215) and kurtosis (2.2686). In the case of converting the matrix of response variables to a vector (vector operator), i.e. the process of stacking the matrix, the value of the skewness coefficient was (0.7400) and kurtosis (2.8539).

In order to know the suitability of the financial inclusion data for the compact model used, different sample sizes were used, and through the Chi-square test, and by assuming 10 different samples for the shape parameters, it was found out that out of 10 samples, there were 7 samples that led to matching (suitability) of the data for the model used under a significance level of (α=0.05). Table below shows the values of the test.

Table (1): Values of the Chi-square test for data matching

Samples	**Chi2-calculate**	**Chi2-tab.(α=0.05)**
(2,2.5,3)	….	**….**
(3,3.5,4)	….	**….**
(5,5.5,6)	1.9854	**3.8415**
(7,7.5,8)	2.6984	**3.8415**
(9,9.5,10)	2.0041	**3.8415**
(11,11.5,12)	3.5478	**3.8415**
(13,13.5,14)	3.6541	**3.8415**
(15,15.5,16)	….	**….**

(17,17.5,18)	3.7741	**3.8415**
(19,19.5,20)	3.8010	**3.8415**

The sample data was divided into two parts, the first consisting of 16 observation (n=16), as this data was used in the estimation, and 4 observations (n=4) were used for the purpose of Bayesian prediction, represented by the countries (Tunisia - United Arab Emirates - Palestine - Yemen) as follows:

6.2 Bayesian estimation of a multivariate mixed compact regression model:

In this section, Bayesian estimation will be performed when prior information is available and based on samples that led to matching under a significance level (α=0.05) and for the Gauss and Quartic kernel function and for the quadratic, balanced quadratic loss function at the weight (w=0.25, 0.75) in addition to using traditional methods to choose the initial values for the parameter matrix ϑ as follows:

$$\vartheta_0 = \begin{bmatrix} 1.1800 & 0.9874 & 0.3541 & 0.8912 \\ 0.0192 & -0.0510 & 0.0012 & 2.2257 \end{bmatrix}^T$$

$$\Psi_0 = eye(4) \quad ,$$

Tables below show the MSE values for the multivariate mixed compact regression model.

	Samples	Gauss kernel function	Quartic kernel function	Rank
		Informative prior	Informative prior	
1	(5,5.5,6)	**5.7223**	5.7505	4
2	(7,7.5,8)	**5.744**	5.7638	5
3	(9,9.5,10)	**5.6528**	5.6879	1
4	(11,11.5,1	**5.6869**	5.693	2
5	(13,13.5,1	**5.8373**	5.8499	6
6	(17,17.5,1	**5.8597**	5.8687	7
7	(19,19.5,2	**5.7116**	5.7309	3

Table (3): MSE of the estimated multivariate mixed compact regression model under the balanced quadratic loss function

	Samples	Gauss kernel function		Quartic kernel function		Rank
		$w = 0.25$	$w = 0.75$	$w = 0.25$	$w = 0.75$	
1	(5,5.5,6)	5.1136	**4.7202**	5.2512	4.9504	3
2	(7,7.5,8)	5.1353	**4.7418**	5.2638	4.9635	5
3	(9,9.5,10)	5.0438	**4.6507**	5.1884	4.888	1
4	(11,11.5,1:	5.0868	**4.6867**	5.1933	4.8929	2
5	(13,13.5,1·	5.2291	**4.8351**	5.3486	5.0491	6
6	(17,17.5,1(	5.2516	**4.8575**	5.3661	5.0676	7
7	(19,19.5,2(	5.1309	**4.7399**	5.2396	5.0000	4

From the results of table (2) and table (3), we notice that the balanced quadratic loss function outperforms the rest of the functions at weight (w=0.75) for all cases of samples and for Gauss kernel function.

$$\hat{\vartheta}_{qb} = \begin{bmatrix} 1.2105 & 0.9064 & 0.4021 & 0.8900 \\ 0.0254 & -0.0660 & 0.0101 & 2.2457 \end{bmatrix}^T$$

7. Bayesian prediction:

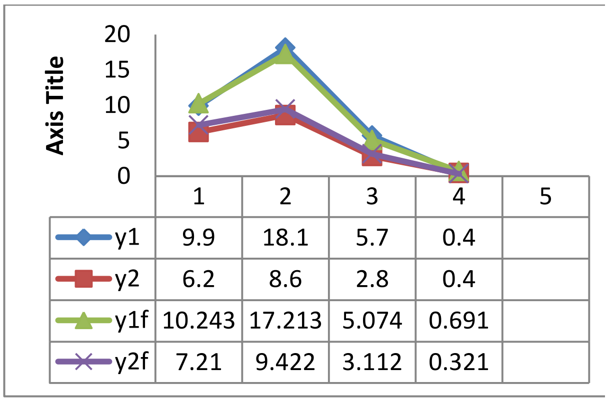
This section includes prediction will be performed when prior information is available and based on samples that led to matching under a significance level (α=0.05) and for the Gauss and Quartic kernel function and for the balanced quadratic loss function at the weight (w= 0.75). Tables below show the MSE values for the prediction multivariate mixed compact regression model:

Table (4): MSE values for the prediction multivariate mixed compact regression model under the balanced quadratic loss function

Samples		Gauss kernel function	Quartic kernel function
		$w = 0.75$	$w = 0.75$
1	(5,5.5,6)	7.9201	7.9604
2	(7,7.5,8)	7.9416	7.9645
3	(9,9.5,10)	7.8505	7.898
4	(11,11.5,12)	7.8864	7.9929
5	(13,13.5,14)	7.9753	8.0091
6	(17,17.5,18)	7.9975	8.0676
7	(19,19.5,20)	7.9396	8.0000

We note from table (4) the superiority of the Gauss kernel function in finding the Bayesian prediction for the multivariate mixed compact regression model function through the MSE criterion, and the tables below show the real and predicted values of the reality of financial inclusion in Arab countries for the year 2014.

	Real Values		Prediction Values	
	y_1	y_2	$\hat{y}_{1f}$	$\hat{y}_{2f}$
1	9.9	6.2	10.243	**7.21**
2	18.1	8.6	17.213	**9.422**
3	5.7	2.8	5.074	**3.112**
4	0.4	0.4	0.691	**0.321**



8. Conclusion:

The researchers had reached the most important theoretical and practical conclusions of the multivariate mixed compact regression model, as follows:

1. Posterior probability distribution of the matrix(ϑ) when the matrix (Σ) is known follows a matrix-variate extension hyperbola distribution.
2. Predictive distribution of the matrix of future observations of the response variable Y_f is not one of the known probability distributions, but it is an appropriate (proper) distribution.
3. Bayesian estimator for the matrix (ϑ) of the multivariate mixed compact regression model under the balanced quadratic loss function is better than the quadratic loss function at (w=0.75). Based on the criterion of the sum mean of squares error (MSE).
4. Superiority of the Gauss kernel function when estimating the location matrix (ϑ) over the Quartic kernel function for all loss functions.

References

[1] AL-Mouel, A. S. and Mohaisen A. J. (2017) "Study on Bayes Semiparametric Regression" Journal of Advances in Applied Mathematics, Vol. 2, No. 4, pages 197-207 .

[2] Gallaugher, M.P. & McNicholas, P.D. (2019) "Three skewed matrix variate distribution‹" Statistics & Probability Lettres, vol.145, p.p. 103-109.

[3] Hmood, M. Y. & Hassan, Y. A. (2020)" Estimate the Partial Linear Model Using Wavelet and Kernel Smoothers", Journal of Economics and Administrative Sciences, Volume 26, Issue 119, p.p. 428-443.

[4] Hmood, M. Y. (2005)" Comparing nonparametric estimators for probability density functions" Ph.D. dissertation, Department of Statistics, Baghdad University.

[5] Jozani, M. J., Marchand, E. & Parzian, A. (2012)" Bayesian and robust Bayesian analysis under ageneral class of balanced loss function", Statistical Papers 53, p.p. 51-60, https://doi :10.1007/s00362-010-0307-8. Source: RePEc.

[6] 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.

[7] Mora, J. A. N. & Mata, L. M. (2013) "Numerical aspects to estimate the generalized hyperbolic probability distribution", Journal of Finance & Economics, vol. 1(4), p.p. 1-9.

[8] Przystalski, M. (2014), "Estimation of the covariance matrix in multivariate partially linear models", Journal of Multivariate Analysis, Vol.123, p.p. 380 -385.

[9] Salih, S. A., & Aboudi, E. H. (2022) “Estimate the location matrix of a multivariate semiparametric regression model when the random error follows a matrix-variate generalized hyperbolic distribution”. International Journal of Nonlinear Analysis and Applications, 13(1), 2467-2482.

[10] 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.

[11] Thabane, L. & Kibria, B. M. G. (2007)” Bayesian Prediction for the Linear Model with Equi-correlated Responses: An Application of Generalized Multivariate Modified Bessel Distribution”, Journal of Applied Statistical Science, p.p. 328-335.

[12] You, J., Zhou, Y. and Chen, G. (2013)" Statistical inference for multivariate partially linear regression models", the Canadian journal of statistics, vol.41, no.1, pages1-22.