Journal of Social and Economic Statistics

*Vol. 6, No. 1, Summer 2017*


# *SELECTION OF VARIABLES INFLUENCING IRAQI BANKS DEPOSITS BY USING NEW BAYESIAN LASSO QUANTILE REGRESSION*

**Fadel Hamid Hadi ALHUSSEINI[a]**

**Abstract**

*The main focus of the paper is modelling the relationship between Iraqi banks deposits and a set of independent variables, including selecting of important independent variables that affect the Iraqi banks deposits. The approach is assigning independent scale mixture of uniform distributions for the regression parameters in the quantile regression model by building an efficient Gibbs sampler to posterior distributions through MCMC algorithms. This study contains one response variable (Iraqi banks deposits and eight independent variables. Three quantile levels (0.30, 0.60, 0.90) are utilized. The optimal quantile regression model results at high quantile level (0.90). This is clear from the pseudo-R squared value. Therefore, we will focus on the high quantile level. Five independent variables have a significant effect on the response variable (Iraqi banks deposits). At high quantile level, the result of variables selection shows six independent variables with importance in the building of the quantile regression model. The rest of the independent variables are not important.*

**Keywords**: Bayesian approach, Lasso quantile regression, scale mixture uniform, deposits of Iraqi banks, variables selection

**JEL Classification**: C11, C15, C22, E50

**Author's/Authors' Affiliation**

[a] - Department of Statistics and Economic Informatics, University of Craiova, Romania, fadhelfadhel222@yahoo.com. University of Al-Qadiseya, Iraq

## 1. Introduction

The banking sector plays an important role in the growth of the economy in any state. Banking services are considered one of the most important sources of financing for various projects and investments, especially with companies and individuals. Banks offer a set of services to their customers, including deposits, loans and financing for various projects. The banking system in Iraq, according to the Financial Stability report (2013) consists of 47 banks, which are divided into 7 state-owned banks and 40 private banks, including 10 foreign banks (Mahmoud and Ahmed, 2014). The private banks were established as an independent entity in the private sector, offering banking services to the Iraqi market. The Iraqi banks deposits are influenced by a set of factors, but these factors differ in impact from one factor to other. In this study, we will focus on response variables which represent Iraqi banks' deposits. Additionally, there are eight independent variables: the profits as an indicator for the bank's success on the banking market, the bank's reserves as an indicator for earning customer confidence, the number of bank branches, the bank's age, the amount of the interest rate, the number of debit cards as an indicator for technological evolution, advertising expenses, the number of customers as companies and individuals. The relationship between Iraqi banks' deposit and a set of factors is revealed by building a model. One of the challenges is related to selecting the most suitable model for the data under study. In this paper, we will use the quantile regression (QReg) model proposed by Koenker and Bassett (1978). This type of regression model is not required to satisfy normal distributional assumptions. The general form of the QReg is given by

$$y_i = x_i^t \beta_\tau + \varepsilon_i \qquad \tau \in (0,1) \tag{1}$$

where, $y_i$ is the response variable,
$\beta_\tau$ is a vector of unknown parameters,
$\tau$ is the quantile level belong to the interval (0,1)
$\varepsilon_i$ is the residual term with density function restricted to have the $\tau_{\text{th}}$ quantile equal to 0. It is satisfying $pro(\varepsilon_i \le 0|x_i) = p$.
$x_i^T$ is a vector of independent variables.

The estimation of the quantile regression coefficients $\beta_\tau$ is done by minimizing the following equation:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^t \beta_\tau\right) \tag{2}$$

where $\rho_\tau(\varepsilon_i)$ is the check function defined by:

$$\rho_\tau(\varepsilon_i) = \begin{cases} \tau\varepsilon_i & if\ \varepsilon_i \ge 0 \\ -(1-\tau)\varepsilon_i & if\varepsilon_i < 0 \end{cases} \tag{3}$$

Equation (2) is not differentiable at the origin; therefore, there is no exact solution for equation (2). Equation (2) can be solved by using a linear programming algorithm (Koenker and Orey, 1987). Mooney (1993) proposed to solve the equation (2) by using a bootstrap technique. The QReg model has good features, such asrobustness against outlier dataset. Furthermore, the quantile regression model can provide us the perfect picture of the full distribution of the relationship between the response variable and its independent variables. The quantile regression model can accommodate non-normal random errors. All these properties make the quantile regression model popular in many fields of knowledge. For instance, in econometrics, medicine, finance, etc. The number of quantile regression lines is infinite through $\tau_{th} \in (0,1)$. Therefore, there are infinite quantile regression lines that can be estimated. According to the logic of mathematics, there are infinite values included in the interval (0,1). But the estimation of all quantile regression lines is a difficult task. Therefore, the researchers in this field can select a set of quantile regression lines according to the data under study. In this paper, we will use three quantile regression lines described as follows: quantile regression model at the low quantile level (0.30), quantile regression model at the middle quantile level (0.60) and quantile regression model at the high quantile level (0.90). For parameters estimation of these models we used the Bayesian approach, through building the Markov Chain Monte Carlo (MCMC) algorithm for estimating the coefficients of the quantile regression model. The paper is structured as follows: in section (2), we present the Variable Selection in the Quantile Regression Model. Section (3) deals with the new Bayesian lasso in the quantile regression model. Section 4 presents a study sample and the mathematical model. Section (5) includes the analysis of the data under study, and brief conclusions are included in section (6).

## 2. Variables selection in quantile regression model

Variables selection is one important technique in the statistical models, because it is a good approach in building statistical models. The approach of variables selection has attractive properties, it provides us a flexible models. The variables selection focuses on independent variables that strongly affect on the response variable. In the same time, it excludes the independent variables that have not effect on the response variable. The philosophy of variables selection is not a modern idea, but the methods of variables selection have begun to evolve. Classical methods for variables selection are primitive, such as AIC (Akaike, 1973) and BIC (Schwarz, 1978). These classical methods have drawbacks, for example, they need a long time to achieve variables selection. Because they deal with $2^p$ models (where $p$ is a number of independent variables) for choosing the optimal model. The variables selection is a statistical tool used for discriminating between important and unimportant independent variables in statical models. The variables selection methods can improve forecasting precision. Also, variables selection is often used for identifying a small subset of independent variables from a large set of independent variables to get better explanations for the model under study. Recently, many researchers have studied the subject of variables selection. Tibshirani (1996) proposed the lasso method (least absolute shrinkage and selection operator) in the classical regression model. This technique can achieve variable selection and coefficients estimation at the same time. Lasso estimation in classical regression model is achieved through the following equation:

$$\hat{\beta}^{lasso} = \text{minimize} \sum_{i=1}^{n} \left(y - X\hat{\beta}\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \quad , \lambda \geq 0 \tag{4}$$

In this paper, we focus on coefficient estimation and variables selection in the quantile regression model by using a Bayesian approach. Koenker (2004) is the first researcher who employs the lasso penalty method in the Quantile regression model. The lasso method in the quantile regression model takes the following formula:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^t \beta_\tau\right) + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \quad (5)$$

Here, the quantity $\lambda \sum_{j=1}^{p} \left|\beta_j\right|$, is called penalty lasso and $\lambda$ is the tuning parameter responsible for the quantity of shrinkage. Furthermore, equation (5) is not differentiable at 0, but parameter estimation can be achieved through the function (rq.fit.lasso) within packages 'quantreg' (Koenker, 2013).

### 3. Bayesian quantile regression model

Yu and Moyeed (2001) noted that the random error for the quantile regression model is close to Asymmetric Laplace distribution (ALD). The likelihood function of (ALD) will take following formula:

$$f(u) = \tau^n (1-\tau)^n Exp(-\sum_{i=1}^{n} \rho_\tau (y_i - x_i^t \beta_\tau) \quad (6)$$

where $\sigma = 1$.

Dealing with Asymmetric Laplace distribution directly is leading to hard computation. Kozumi and Kobayashi (2011) provided another formula by reformulating the Asymmetric Laplace distribution as a Scale Mixture Normal family $y_i|\beta_\tau, x, \tau, m_i \sim N(x_i^t \beta_\tau + (1 - 2\tau) m_i, 2m_i)$. Here, the likelihood function takes the following formula:

$$f(y|x_i^t, \beta_\tau, m_i) = \frac{1}{\sqrt{4\pi m_i}} Exp(-\sum_{i=1}^{n} \frac{[y_i - x_i^t \beta_\tau - (1-2\tau) m_i]^2}{4m_i} \quad (7)$$

This formula gives us an efficient algorithm for coefficients estimation of quantile regression model. To implement variables selection and coefficient estimation in the quantile regression model together by using a Bayesian approach, we must choose the appropriate prior distribution with parameters of the model because the prior distribution plays a good role in determining the appropriate method for variable selection.

Tibshirani (1996) proposed Laplace distribution for parameter vector $\beta_\tau$ as prior distribution. The probability density function (p.d.f) of Laplace distribution with $\beta_\tau$ is

$$g(\beta_\tau|\lambda) = (\lambda/2)^k \operatorname{Exp}\left(-\lambda \sum_{j=1}^{k} |\beta_\tau|\right) \quad (8)$$

Hence, the posterior distribution through multiplication equation (6) in equation (8) takes the form:

$$f(\beta_\tau|y,x,\lambda) \propto exp\left\{-\sum_{i=1}^{n}\rho_\tau(T-(x_i^t\beta_\tau)\right\} - \lambda\sum_{j=1}^{k}|\beta_\tau| \qquad (9)$$

From equation (9), we will obtain the Bayesian lasso quantile regression. Hence, some parameters will shrink to zero exactly, this means, the Bayesian lasso quantile regression achieves variable selection in the quantile regression model. The next figure provides us a clear vision about the idea of variable selection in a Bayesian approach.

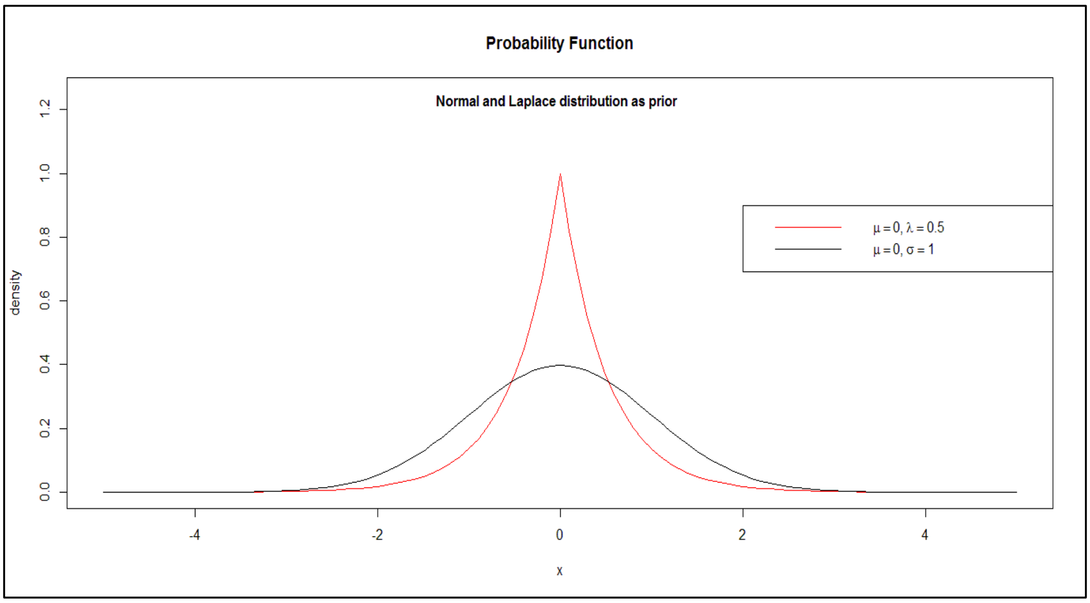


**Figure 1. The (pdf) to Normal and Laplace distributions as prior**

As indicated in figure 1, the top of the symmetric normal distribution is convex. Therefore, the projection points from its top to its base will be close to zero. But the top of the symmetric Laplace distribution takes a triangle shape. Therefore, some of projection points from its top to its base will be equal to zero exactly. Figure 1 supports the idea that Laplace distribution is necessary in the Bayesian lasso regression approach.

### *3.1. New Bayesian lasso quantile regression model*

The Bayesian lasso regression model was achieved through using the Laplace distribution as a prior (Tibshirani, 1996). The use of the general formula for the Laplace distribution is a difficult matter. Park and Casella (2008) introduced the Bayesian lasso to classical linear models by using a scale mixture of normal (SMN) priors on the parameters and independent exponential priors on their variances as the following formula:

$$g(\beta_\tau|\lambda) = (\lambda/_2)^k \,\mathrm{Exp}\left(-\lambda\sum_{j=1}^{k}|\beta_\tau|\right) = \prod_{j=1}^{k}\int_0^\infty \frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{{\beta_\tau}^2}{2s}\right)\ \frac{\lambda^2}{2}\ \ \exp\left(-\frac{\lambda^2}{2}\ s\right) \qquad (10)$$

Equation (10) is considered a reformulation of the Laplace distribution, where this formulation provides simple computation in the Bayesian lasso linear regression. Li et al. (2010) proposed the Bayesian Lasso for the linear quantile regression model through using equation (10) as a Laplace distribution prior.

Mallick and Yi (2014) proposed another formula to the Laplace prior on *βj:*

$$\frac{\lambda}{2} e^{\{-\lambda|\beta_j|\}} = \int_{u_j > |\beta_j|} \frac{1}{2u_j} \frac{\lambda^2}{\Gamma(2)} u_j{}^{2-1}\, exp\{-\lambda u_j\}\, du_j \qquad (11)$$

where $(\frac{1}{2u_j})$ is (pdf) of uniform distribution and the rest of the equation (11) represents (pdf) of the gamma distribution with scale parameter $(\lambda)$ and shape parameter (2), and $\Gamma(2) = (2-1)! = 1$. According to Mallick and Yi (2014) equation (11) represents another formula for the Laplace distribution called the scale mixture uniform (SMU). Mallick and Yi (2014) employed equation (11) for the implementation of Bayesian variable selection in a linear regression model. In this paper, we propose a new Bayesian lasso in quantile regression model for coefficient estimation and variable selection. Here in this paper, we use (SMU) as in the hierarchical prior distribution:

$$\begin{gathered} m_i \sim exp\{\tau(1-\tau)\}, \\ \beta_j|u_j \sim Uniform\,(-u_j, u_j), \\ u_j|\lambda \sim Gamma(2,\lambda)\ \sigma \sim Gamma(a,b)\ , \\ \lambda \sim Gamma(c,d)\ . \end{gathered} \qquad (12)$$

where $Exp(\tau(1-\tau))$ is the exponential distribution with rate parameter $\tau(1-\tau)$.

### *3.2. Full conditional posterior distribution*

The Bayesian quantile regression procedure is a multiplication of the likelihood function of asymmetric Laplace distribution with a set of prior distributions in equation (12). From this procedure, we will obtain the updated posterior distribution to building efficient MCMC algorithm.

- Updating $m_i{}^{-1}$: the full conditional distribution (FCD) of $m_i{}^{-1}$ is inverse Gaussian distribution, referred to as $InvGa(m_i^t, \vartheta^t)\ , where\ \vartheta^t = \frac{1}{2}\ , m_i^t = 1 \Big/ \sqrt{(y_i - x_i^t \beta)^2}$. Here, the (pdf) of the $InvGa$ is given by (Chhikara, 1988):

$$f(x|m_i^t, \vartheta^t) = \sqrt{\frac{\vartheta^t}{2\pi}}\ x^{\frac{-3}{2}}\ exp\left\{\frac{-\vartheta^t (x - m_i^t)^2}{2(m_i^t)^2 x}\right\}, x, m_i^t, \vartheta^t > 0 \qquad (13)$$

- Updating $\beta_j$: the full conditional distribution of $\beta_j$ is truncated normal with mean $\overline{\beta_J}$ and variance $s^2_{\beta_j}$, where $\overline{\beta_J} = \left( s^2_{\beta_j} \sum_{i=1}^{n} \frac{\sigma x_{ij}\ \left(y_i - (1-2\tau) m_i - \sum_{,j\neq l}^{p} x_{ij}\beta_j\right)}{2m_i} \right) I\{|\beta_j| < u_j\}$ and $s^2_{\beta_j} = \left( \sum_{i=1}^{n} \frac{\sigma x_{ij}^2}{2m_i} \right)^{-1}$.

- Updating $u_j$: the full conditional distribution of $u_j$ is a left-truncated exponential distribution given by $u_j|\beta,\lambda \sim Exp(\lambda)I\{u_j > |\beta_j|\}$. Updating $u_j$ can be done by using inversion method (Mallick and Yi, 2014) as follows:
  1. Update $u_j{}^*$ from $Exp(\lambda)$
  2. Set $u_j = u_j{}^* + |\beta_j|$
- Updating $\lambda$ the: the full conditional distribution of $\lambda$ is Gamma $\left(c + 2p, d + \sum_{j=1}^{p} |\beta_j|\right)$.

The above Gibbs sampler proceeds to sample each unknown parameter to the full conditional posterior distribution, conditional on all other unknowns. The Gibbs sampler performs thousand iterations through all elements of $(y, m, \beta, u_j, \lambda_{j)}$. Through the MCMC (Markov Chain Monte Carlo) algorithm, the Bayesian estimation for the quantile regression model implemented via mean of conditional posterior distributions is as mentioned in section 3.2.

## 4. Study sample and mathematical model

The data under study was taken from the Central Bank of Iraq, and the sample size was of 47 observations (active Banks in Iraq). This study contains one response variable which represents an Iraqi Banks deposit, and a set of independent variables. This study is focused on a set of goals, firstly the evaluation of the relationship between the response variable and independent variables. Secondly, the variables selection that affect on the Iraqi banksdeposit of active banks in Iraq. In this study, we employed the quantile regression model at three quantile levels (0.30, 0.60, 0.90) respectively, because quantile regression has attractive properties. Hence, the mathematical model for this study takes the following formula:

$$y = \alpha_\tau + \beta_{1\tau}x_{1j} + \beta_{2\tau}x_{2j} + \beta_{3\tau}x_{3j} + \beta_{4\tau}x_{4j} + \beta_{5\tau}x_{5j} + \beta_{6\tau}x_{6j} + \beta_{7\tau}x_{7j} + \beta_{8\tau}x_{8j} + \varepsilon_i \; , \quad {}_{,j=1,2,\ldots\ldots 47} \qquad (14)$$

where $(x_{1j}, x_{2j}, x_{3j}, x_{4j}, x_{5j}, x_{6j}, x_{7j}, x_{8j})$ are independent variable as in Table (1).

**Table (1): Independent variables and descriptive of statistics for each independent variable**

| Names and codes for independent variables | Type and measurement unit of independent variables | Descriptive of statistics | | | |
|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Minimum | Maximum |
| $x_{1j}$: The bank profits | Quantitative variable (Iraqi Dinars) | 38616219512 | 89541413975 | -26042000000 | 47673900000000 |
| $x_{2j}$: The banks reserve. | Quantitative variable (Iraqi Dinars) | 78263829787 | 58882448945 | 1000000000 | 1750000000000 |
| $x_{3j}$: The number of bank branches | Quantitative variable (number of banks branches) | 22 | 38.927 | 1 | 201 |
| $x_{4j}$: The banks age | Quantitative variable (number of years) | 15 | 18.101 | 2 | 78 |

| | | | | | |
|---|---|---|---|---|---|
| $x_{5j}$: The bank interest rate | Quantitative variable (%) | 5.63% | 1.85% | 3% | 9% |
| $x_{6j}$: The number of machine of debit cards | Quantitative variable (number of debit cards for each bank) | 386 | 304.59 | 0 | 987 |
| $x_{7j}$: The advertising expenditures | Quantitative variable (Iraqi Dinars) | 14155957 | 21038751 | 70000000 | 0 |
| $x_{8j}$: the number of customers | Quantitative variable (number of bank's customers) | 133121920 | 898517078 | 4778 | 898517078 |

$\beta_{1\tau}$, $\beta_{2\tau}$, $\beta_{3\tau}$, $\beta_{4\tau}$, $\beta_{5\tau}$, $\beta_{6\tau}$, $\beta_{7\tau}$, $\beta_{8\tau}$ are coefficients of independent variables at quantile level $\tau$.
$\varepsilon_i$ : is the random error at quantile $\tau$.

## 5. Analysis of data under study

In this paper, the data analysis takes two directions: the first is the coefficient estimation of the model and the second is the variable selection of the model via three quantile levels (0.30, 0.60, 0.90).

For parameters estimations of this model, we constructed a new algorithm in the statistical program R, through Markov Chain Monte Carlo (MCMC) algorithm. It was run 11000 iterations. The first 1000 were ruled out as burn in.

### *5.1. Coefficient Estimation*

We will estimate the parameters of the quantile regression model by using the Bayesian approach via three quantile levels.

**Table (2): Coefficient estimation of the quantile regression model via (0.30, 0.60, 0.90)**

| **Coefficients** | | | |
|---|---|---|---|
| Variables | Quantile level (0.30) | Quantile level (0.60) | Quantile level (0.90) |
| Intercept | 0.497 | 1.777 | 2.595 |
| $x_1$ | -0.039 | -0.264 | 0.428 |
| $x_2$ | 0.184* | 0.0046 | 0.229* |
| $x_3$ | -0.381* | -0.155 | 0.176* |
| $x_4$ | 0.131 | 0.089 | 0.076* |
| $x_5$ | 0.174 | 0.057* | 0.135* |
| $x_6$ | 0.563* | 0.236* | -0.097 |
| $x_7$ | 0.007 | -0.008 | -0.075 |
| $x_8$ | -0.081 | -0.010 | 0.534* |
| **The pseudo-R squared** | **0.503** | **0.273** | **0.872** |

* is p-value < 0.05

### *5.1.1. Coefficients estimation at low quantile level (0.30)*

The result of the pseudo-R square is 0.503, this means all independent variables ($x_{1j}$ the profit of the bank, $x_{2j}$ the bank's reserve, $x_{3j}$ the number of bank branches, $x_{4j}$ the bank's age, $x_{5j}$ the bank's interest rate, $x_{6j}$ the number of debit cards issued by the bank, $x_{7j}$ the advertising expenses of the bank, $x_{8j}$ the number of customers of the bank) can explain 50.3% of the variation in Iraqi banks deposits. From this indicator, there is a medium relationship between Iraqi banks' deposits and all the independent variables at low quantile level (0.30)for representation of the data under study. In table 2, there are three independent variables that have a significant effect on the response variable (Iraqi banks' deposits). The rest five independent variables have an insignificant effect on Iraqi banks' deposits at a low quantile level (0.30).

### *5.1.2. Coefficient estimation at middle quantile level (0.60)*

The result of the pseudo-R square is 0.273, this means all independent variables ($x_{1j}$ the profit of the bank, $x_{2j}$ the bank's reserve, $x_{3j}$ the number of bank branches, $x_{4j}$ the bank's age, $x_{5j}$ the bank's interest rate, $x_{6j}$ the number of debit cards issued by the bank, $x_{7j}$ the advertising expenses of the bank, $x_{8j}$ the number of customers of the bank) can explain 27.3% of the variation in Iraqi banks' deposits. From this indicator, there is a weak relationship between Iraqi banks' deposits and all the independent variables at middle quantile level (0.60) for representation of the data under study. In table (2), we see two independent variables that have a significant effect on the response variable (Iraqi banks' deposits). The rest six independent variables have an insignificant effect on Iraqi banks' deposits at a middle quantile level (0.60).

### *5.1.3. Coefficient estimation at high quantile level (0.90)*

The result of the pseudo-R square is 0.872, this means all independent variables ($x_{1j}$ the profit of the bank, $x_{2j}$ the bank's reserve, $x_{3j}$ the number of bank branches, $x_{4j}$ the bank's age, $x_{5j}$ the bank's interest rate, $x_{6j}$ the number of debit cards issued by the bank, $x_{7j}$ the advertising expenses of the bank, $x_{8j}$ the number of customers of the bank) can explain 87.2% of the variation in Iraqi banks' deposits. From this indicator, there is a very strong relationship between Iraqi banks' deposits and all the independent variables at high quantile level (0.90) for representation of the data under study. Therefore, we will focus on high quantile regression for explanation, as follows:

**$x_{2j}$ The bank's reserve**

The quantile regression coefficient for the bank reserve is 0.229. This indicates an increase in the bank reserve by one unit, leading to an increase in Iraqi banks deposits by 0.229 units, due to this variable being in a positive relationship with the response variable. This variable has a significant effect on Iraqi banks' deposits.

**$x_{3j}$ The number of bank branches**

The quantile regression coefficient for the number of bank branches is 0.176. An increase in the number of bank branches by one unit leads to an increase in Iraqi banks deposits by 0.176 units, due to this variable being in positive relationship with the response variable. This variable has a significant effect on Iraqi banks deposits.

**$x_{4j}$ The bank' s age**

The quantile regression coefficient for the bank's age is 0.076. An increase in the bank's age by one unit leads to an increase in Iraqi banks deposits by 0.076 units, due to this variable being in a positive relationship with the response variable. This variable has a significant effect on Iraqi banks deposits.

**$x_{5j}$ The bank interest rate**

The quantile regression coefficient for the bank interest rate is 0.135. An increase in the bank interest rate by one unit leads to an increase in Iraqi banks deposits by 0.135 units, due to this variable being in positive relationship with the response variable. This variable has a significant effect on Iraqi banks deposits.

**$x_{8j}$ The number of customers of the bank**

The quantile regression coefficient for the the number of customers of the bank is 0.534. An increase in the number of customers of the bank by one unit leads to an increase in Iraqi banks deposits by 0.534 units, due to this variable being in a positive relationship with the response variable. This variable has a significant effect on Iraqi banks deposits.

The rest independent variables ($x_{1j}$ the profit of the bank, $x_{6j}$ the number of debit cards issued by the bank, $x_{7j}$ the advertising expenses of the bank) have insignificant effect on Iraqi banks' deposits.

### *5.2. Variable selection*

In Bayesian approach, there are thousands of estimations for the model parameters via thousands of iterations (Gramacy and Lee, 2008). In this paper, the estimated coefficients are compared with the interval (-1, 1) for computing the probability value for each estimated coefficient. If the probability value for the estimated coefficient out of the interval (-1,1) is greater than 0.5, this means that the independent variable have importance in the model. If it is less than 0.5, this means the independent variables have not importance in the model (Reed, 2011).

**Table (3): The Probability value of each variable in the quantile regression model**

| The probability | | | |
|---|---|---|---|
| Variables | Low quantile level =0.30 | Middle quantile level=0.60 | High quantile level=0.90 |
| $x_1$ | 0.43 | 0.32 | 0.69 |
| $x_2$ | 0.29 | 0.51 | 0.74 |
| $x_3$ | 0.68 | 0.60 | 0.68 |
| $x_4$ | 0.57 | 0.48 | 0.57 |
| $x_5$ | 0.84 | 0.29 | 0.92 |
| $x_6$ | 0.12 | 0.18 | 0.22 |
| $x_7$ | 0.41 | 0.44 | 0.34 |
| $x_8$ | 0.26 | 0.19 | 0.57 |

### *5.2.1. At low quantile level (0.30)*

From table (3), we see three independent variables have probability values greater than 0.5, this indicator shows these independent variables ($x_{3j}$ the number of bank branches, $x_{4j}$ the bank's age, $x_{5j}$ the bank's interest rate) have importance in the construction of the quantile regression model at the low quantile level (0.30).

### *5.2.2. At middle quantile level (0.60)*

From table (3) at a middle quantile level (0.60), we see two independent variables ($x_{2j}$ the bank reserve, $x_{3j}$ the number of bank branches) have importance in the construction of the quantile regression model because these independent variables have a probability value greater than 0.5.

### *5.2.3. At high quantile level (0.90)*

At high quantile level (0.90), there are six independent variables having probability values greater than 0.5. From this result, the independent variables ($x_{1j}$ the profit of the bank, $x_{2j}$ the bank reserve, $x_{3j}$ the number of bank branches, $x_{4j}$ the bank's age, $x_{5j}$ the bank's interest rate, $x_{8j}$ the number of the customers of the bank) have high ability in building of a quantile regression model at a quantile level of 0.90. The figure 2 provides us clear a vision about probability values for each independent variables via three quantile levels.

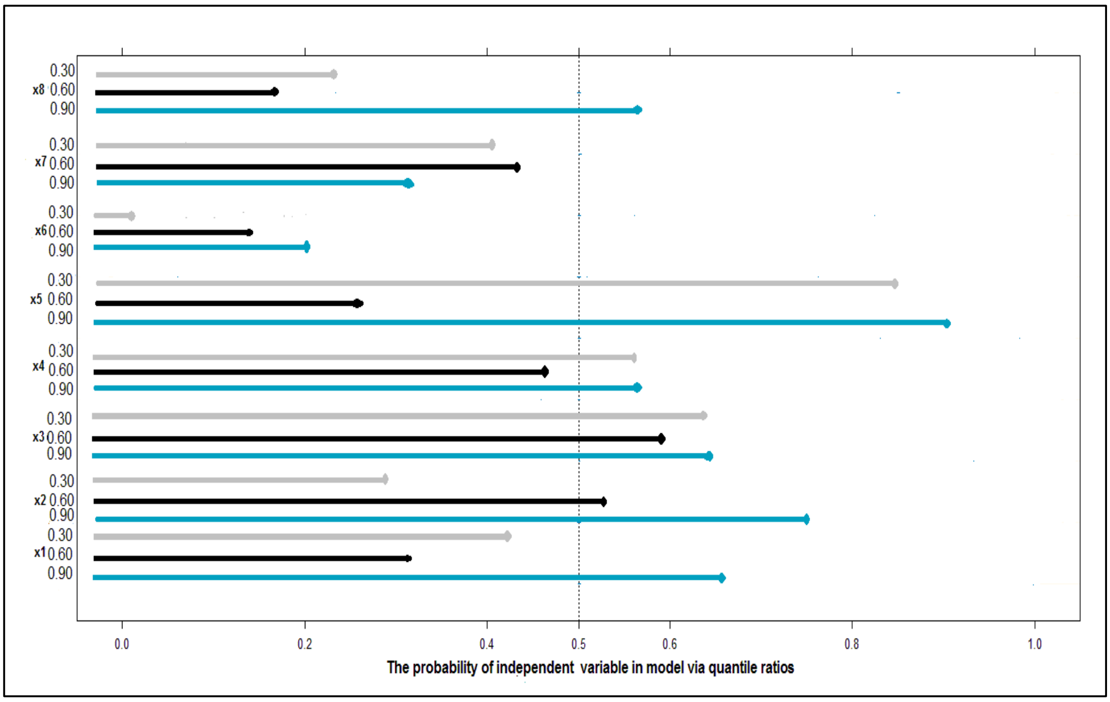


**Figure 2: The probability value for each independent variable via three quantile levels - at 0.30 (gray line), at 0.60 (black line ) and at 0.90 (blue line)**

## 6. Conclusions

To obtain the whole distribution of the relationship between the response variable and its independent variableswe will use the quantile regression model via three quantile levels. In this paper, we conclude that the quantile regression model at a high quantile level (0.90) is the best model in representation of the data under study. The quantile regression model at the low quantile level (0.30) came in second and the quantile regression model at the middle quantile level (0.60) came in third. This conclusion is built according to the amount of the pseudo-R squared.

To compare the significant variables via three quantile levels. We see the ($x_2$,$x_3$,$x_6$) have significant effect on the response variable at quantile level (0.30). At quantile level (0.60) there are two independent variables ($x_5, x_6$) that have significant effect on the response variable and ($x_2$,$x_3$,$x_4$, $x_5$, $x_7$) have significant effect on response variable at quantile level (0.90).

We see the independent variable $x_6$ has a significant effect on the response variable at both quantile levels (0.30) and (0.60). Also, we see that ($x_2$,$x_3$) have a significant effect on the response variable at both quantile levels (0.30) and (0.90). The independent variable $x_5$ has a significant effect on the response variable at both quantile levels (0. 60) and (0.90). We conclude the the independent variables ($x_6, x_2$,$x_3, x_5$ ) have high significant effect on the response variable (Iraqi banks' deposits).

The important independent variables in the construction of a quantile regression model are described as follows via three quantile levels.

**At a level of 0.30:**

$X_5$: The bank's interest rate with probability (0.84).

$X_3$: The number of bank branches with probability (0.68).

$X_4$: The bank, age with probability (0.57)

There are three independent variables ($X_5$ the bank's interest rate, $X_3$ the number of bank branches and $X_4$ the bank's age) that have importance in the construction of the quantile regression model at the low quantile level (0.30). The rest of the independent variables do not have importance in quantile regression at the same level, therefore, it is possible to remove them from the model.

**At a level of 0.60:**

$X_3$: The number of bank branches with probability (0.60).

$X_2$: The bank's reserve with probability (0.51).

There are two independent variables ($X_3$ the number of bank branches and $X_2$ the bank's reserve) which have importance in quantile regression at middle quantile level (0.60). The rest of the variables do not have importance in quantile regression at the same level, therefore, it is possible to remove them from the model.

**At a level of 0.90:**

$X_5$: The bank's interest rate with probability (0.92).

$X_2$: The bank's reserve with probability (0.51).

$X_1$: The profit of the bank with probability (0.69).

$X_3$: The number of bank branches with probability (0.68).

$X_4$: The bank's age with probability (0.57).

$X_8$: The number of customers of the bank with probability (0.57).

There are six independent variables ($X_5$ the bank's interest rate, $X_2$ the bank's reserve, $X_1$ the profits of the bank, $X_3$ the number of bank branches, $X_4$ the bank's age and $X_8$ the number of customers of the bank) which have importance in quantile regression at high quantile level (0.90). The rest of the variables are not important in quantile regression at the same level. Therefore, it is possible to remove them from the model.

We see the independent variable ($X_3$: the number of bank branches) has impact in building the quantile regression via three quantile levels. But the two independent variables ($x_{6j}$: the number of debit cards, $x_{7j}$: the advertising expenditures) are unimportant in building quantile regression model via three quantile levels. Therefore, we can cancel these independent variables from quantile regression models under study.

## References

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In: Petrov, B.N. And Caski, F. Eds. Budapest: Akademiai Kiado, pp. 267-281.

Craig Reed (2011). *Bayesian parameter estimation and variable selection for quantile regression*. Ph.D. thesis. Brunel University.

Chhikara, R. (1988). *The Inverse Gaussian Distribution: Theory: Methodology, and Applications* (Vol. 95). CRC Press.

Iraq Central Bank (2013). *Financial stability report*.

Gramacy, R.B. and Lee, H.K.H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, **103**(483), pp. 1119-1130.

Koenker, R. and Bassett Jr., G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, **46**(1), pp. 33-50.

Koenker, R.W. and d'Orey, V. (1987). Algorithm AS 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **36**(3), pp. 383-393.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**(1), pp. 74-89.

Koenker, R. (2005). *Quantile regression (No. 38).* Cambridge University Press.

Koenker, R. (2013). *Quantreg: Quantile regression*. R package version 5.05.

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81**(11), pp. 1565-1578.

Mooney, C.Z., Duval, R.D. and Duval, R. (1993). *Bootstrapping: A nonparametric approach to statistical inference (No. 94-95)*. Sage.

Mahmoud, N. and Ahmed, E.M. (2014). Factors Affecting Risk Management Practices and Financial Performance of Iraqi Private Banks. *Middle-East Journal of Scientific Research*, **22**(7), pp. 971-977.

Mallick, H. and Yi, N. (2014). A new Bayesian lasso. *S Statistics and its Interface*, **7**(4), pp. 571-582.

Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, **103**(482), pp. 681-686.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**(2), pp. 461-464.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267-288.

Yu, K. and Moyeed, R.A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54**(4), 437-447.