Bayesian Inference for Linear Regression

Daniel Ciuiu[1]

Abstract

In this paper we will perform Bayesian inference for linear regression. The normal distribution case is considered, hence the distribution of the errors and the multivariate prior distribution of the coefficients are normal. From these we obtain a multivariate normal posterior distribution for the coefficients.

1. Introduction

Opposite classical statistical inference, in the Bayesian inference all the involved parameters are random variables. We start from the formula of Bayes.

Theorem 1 *Let $(\Omega, \mathbf{K}, P)$ be a probability field, $A_1, \ldots, A_n \in \mathbf{K}$ a complete system of events, and $B \in \mathbf{K}$ an event. We have*

$$P(A_i|B) = \frac{P(B|A_i)\cdot P(A_i)}{\sum_{j=1}^{n} P(B|A_j)\cdot P(A_j)}.$$

We have a conditional distribution $X|\theta$, where X is a random variable and θ is the parameter (possible multivariate, as the coefficients of linear regression in this paper). For θ we establish a probability distribution, which is called the prior distribution of θ. Using the formula of Bayes we find the posterior distribution of $\theta|X$. Usually, as in this paper too, the parameter θ is considered to be a continuous random variable having the pdf $g(\theta)$. The posterior pdf of θ is (Preda, 1992; Gelman et al., 2000)

$$g(\theta_0|X = x_i) = \frac{g(\theta_0)\cdot P(X = x_i|\theta = \theta_0)}{\int_{\Theta} g(\theta)\cdot P(X = x_i|\theta)\,d\theta}. \tag{1}$$

if X is a discrete random variable, respectively

$$g(\theta_0|x) = \frac{g(\theta_0)\cdot f(x|\theta_0)}{\int_{\Theta} g(\theta)\cdot f(x|\theta)\,d\theta}. \tag{1'}$$

if X is a continuous random variable with the conditional pdf $f(x|\theta)$.

In the following, we present some definitions (Gelman et al., 2000; Preda, 1992) for choosing the prior distribution.

Definition 1 *The non-informative prior distribution is the prior distribution of the parameter such that the posterior distribution $g(\theta|x)$ is $g(x|\theta)$. Otherwise, the prior distribution is informative.*

[1] Technical University of Civil Engineering, Bucharest and Centre for macroeconomic Modelling, NIER, Romanian Academy. E-mail: dciuiu@yahoo.com

Definition 2 *The prior distribution according maximum entropy principle is the prior distribution with an eventual given prior information, with the maximum Shannon entropy.*

Remark 1 *If* θ *is in the bound interval with the probability one, or is a simple random variable, the non-informative prior distribution and the prior distribution according maximum entropy principle without other information are the same: uniform (continuous or discrete).*

In other cases the non-informative prior distribution is improper (we can not have constant pdf value on **R***). It can be expressed in some cases as limit of proper prior distributions.*

Definition 3 *A family of prior distribution* **P** *is called conjugated prior distribution for the family of pdfs* $\mathbf{F}=\left\{f\left(x|\theta\right)\middle|\theta\in\Theta\right\}$ *if* $(\forall g\in\mathbf{P})$ $(\forall f\in\mathbf{F})$ *we have* $g\left(\theta|x\right)\in\mathbf{P}$.

A special case of conjugated family of prior distribution is that when **P** and **F** are identical to the normal family (Preda, 1992): if $X\sim N\left(\theta,\sigma^2\right)$ with known σ^2, and $\theta\sim N\left(\mu,\tau^2\right)$ then

$$\theta|X\sim N\left(\frac{\tau^2\cdot X+\sigma^2\cdot\mu}{\sigma^2+\tau^2},\frac{\sigma^2\cdot\tau^2}{\sigma^2+\tau^2}\right). \tag{2}$$

We notice that in the above case, the non-informative prior distribution is the limit of $N\left(\mu,\tau^2\right)$ for $\tau^2\to\infty$.

Next we define the Bayes forecasted (predicted) distribution.

Definition 4 (Gelman et al., 2000) *Consider* $X|\theta$ *a discrete or continuous random variable conditioned by the random parameter* θ. *Denote by* $f\left(\theta\right)$ *the prior pdf of* θ, *respectively by* $g\left(\theta|X=x_i\right)$ *or by* $g\left(\theta|x\right)$ *the posterior pdf of* θ.

If $Y|\theta$ *is another discrete random variable, the forecast of* Y *conditioned on* X *is*

$$P\left(Y=y_j\middle|X=x_i\right)=\int_\Theta P\left(Y=y_j\middle|\theta\right)\cdot g\left(\theta|X=x_i\right)d\theta$$

if X *is discrete, respectively*

$$P\left(Y=y_j\middle|x\right)=\int_\Theta P\left(Y=y_j\middle|\theta\right)\cdot g\left(\theta|x\right)d\theta$$

if X *is continuous.*

If $Y|\theta$ *is another continuous random variable, the forecast of* Y *conditioned on* X *is*

$$f\left(y|X=x_i\right)=\int f\left(y|\theta\right)\cdot g\left(\theta|X=x_i\right)d\theta$$

if X *is discrete, respectively*

$$f\left(y|x\right)=\int_\Theta f\left(y|\theta\right)\cdot g\left(\theta|x\right)d\theta$$

if X *is continuous.*

Remark 2 *In literature (Gelman et al., 2000) we consider for a Bayesian forecasting* *model both random variables X and* Y *discrete, or both continuous. More exactly,* X *(used for inference) is even a statistics on the sample* $X_1, X_2,..., X_n$, e*ven the vectorial random variable* $(X_1, X_2,..., X_n)'$ *For instance, in the normal case* X *is the expectation, and in the Bernoulli case* X *is the number of successes/ fails from a given number of trials,* n.

The random variable Y (to forecast) is X_{n+1}. *But it is not necessary: we can have for instance a sample about the number of events in the unit time (having Poisson distribution), and we can purpose to ourselves to forecast the time* Y *until the next event (exponential with the same parameter), or vice versa.*

In the normal case, consider the sample average $\overline{X}$, which has the property that for any prior distribution of θ, we obtain the same posterior distribution for $\theta | X_1, X_2,..., X_n$ and for $\theta | \overline{X}$ if the variance σ^2 of X_i is known. Therefore, in (2) *X* is replaced by $\overline{X}$ and σ^2 by $\frac{\sigma^2}{n}$. Therefore, we take into account that

$$\overline{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right), \tag{3}$$

and from here

$$f\left(\overline{X} | \theta\right) \propto \exp\left\{-\frac{\left(\overline{X}-\theta\right)^2}{\frac{2\sigma^2}{n}}\right\}. \tag{3'}$$

The same relation we obtain for $f\left(X_1,..., X_n | \theta\right)$, hence the sample expectation is equivalent to the sample in Bayesian inference.

Applying (2), we obtain

$$\theta | \overline{X} \sim N\left(\frac{\tau^2 \cdot \overline{X} + \frac{\sigma^2}{n} \cdot \mu}{\tau^2 + \frac{\sigma^2}{n}}, \frac{\tau^2 \cdot \sigma^2}{n \cdot \tau^2 + \sigma^2}\right). \tag{4}$$

By computation, it results that the forecasted distribution of *X* is

$$X | \overline{X} \sim N\left(\frac{\tau^2 \cdot \overline{X} + \frac{\sigma^2}{n} \cdot \mu}{\tau^2 + \frac{\sigma^2}{n}}, \frac{(n+1) \cdot \tau^2 \cdot \sigma^2 + \sigma^4}{n \cdot \tau^2 + \sigma^2}\right). \tag{5}$$

We notice that instead of *X* we have used for inference a statistics ($\overline{X}$ in the normal case), and *X* is the variable to forecast.

In the above formulae, we have consider the prior distribution of $\theta \sim N\left(\mu, \tau^2\right)$. In the case of non-informative prior distribution, we have $\mu \in R$, and $\tau^2 \to \infty$. We obtain

$$\theta | \overline{X} \sim N\left(\overline{X}, \frac{\sigma^2}{n}\right), \text{ and} \tag{6}$$

$$X | \overline{X} \sim N\left(\overline{X}, \sigma^2\left(1+\frac{1}{n}\right)\right). \tag{6'}$$

There are three Bayesian estimators of the parameters (Preda, 1992; Gelman et al., 2000; Congdon, 2001): estimator-mode, estimator-expectation and estimator-median.

First one maximizes the posterior distribution of $\theta | X_1, X_2,..., X_n$. If the prior pdf of

θ is non-informative, the estimator-mode is the classical estimator of maximum likelihood method (applying logarithm). The same thing we can say in the informative case, but we add to the log-likelihood the "correction term" $\ln g(\theta)$.

The other two estimators are the estimators θ_0 that minimizes for a loss function $u(\theta_1,\theta_0)$ the integral

$$\int_\Theta u(\theta,\theta_0)\cdot g(\theta|X)\,d\theta. \qquad (7)$$

For expectation we consider the loss function $u(\theta_1,\theta_0)=(\theta_1-\theta_0)^2$, and for median we take $u(\theta_1,\theta_0)=|\theta_1-\theta_0|$. In the normal case with known variance, the three estimators are identical. This is in fact a well-known property of the normal distribution.

Consider the conditional distribution of the vector X in terms of the vector of parameters θ being normal $N(\theta,\Sigma_1)$. Consider also the prior distribution of θ being k-variate normal distribution $N(\mu,\Sigma_2)$.

The posterior distribution of $\theta|X$ is (Preda, 1992; Hamilton, 1994; Koop, 2003)

$$\theta|X \sim N\left(\left(\Sigma_1^{-1}+\Sigma_2^{-1}\right)^{-1}\left(\Sigma_1^{-1}X+\Sigma_2^{-1}\mu\right),\left(\Sigma_1^{-1}+\Sigma_2^{-1}\right)^{-1}\right). \qquad (7)$$

Therefore the multivariate normal distribution family is conjugated prior distribution for itself. Consider for the above formula instead of X the estimator $\hat{\theta}=\overline{X}$, hence $\Sigma_1=\frac{1}{n}\Sigma$, where Σ is the variance-covariance matrix of *X*, and *n* is the size of the sample. In the case of non-informative prior distribution Σ_2^{-1} vanishes in the above formula, and $\theta|\overline{X}\sim N\left(\overline{X},\frac{1}{n}\Sigma\right)$.

2. Methodology

In linear regression we write the matriceal form of linear regression

$$Y=X\cdot\theta+e, \qquad (9)$$

where θ is the vector of parameters, and e is the error. Consider $e\sim N\left(0_v,\sigma_u^2\cdot I_n\right)$ denoting by 0_v the vector with *n* components equal to zero. It results that $Y\sim N\left(X\cdot\theta,\sigma_u^2\cdot I_n\right)$. Now we take into account that if the vector $X\sim N(m,\Sigma)$, than $C\cdot X+v\sim N(C\cdot m+v,C\cdot\Sigma\cdot C')$. We obtain the above relations in the form

$$(X'X)^{-1}X'Y\sim N\left(\theta,\sigma_u^2\cdot(X'X)^{-1}\right). \qquad (10)$$

We consider the cases of known/ unknown variance of residues σ_u^2. In the first case we apply the formula (8) and we obtain

$$\theta|(X'X)^{-1}X'Y\sim N\left(\mu,\Sigma_3\right),\text{ where} \qquad (11)$$

$$\begin{cases} \hat{\mu} = \Sigma_3 \left(\frac{1}{\sigma_u^2} X'Y + \Sigma_2^{-1} \mu \right) \\ \Sigma_3 = \left(\frac{1}{\sigma_u^2} X'X + \Sigma_2^{-1} \right)^{-1} \end{cases}, \text{ and} \tag{11'}$$

the prior distribution of θ is $N(\mu, \Sigma_2)$.

If Σ_2^{-1} is neglectable with respect to $\Sigma_1^{-1} = \frac{1}{\sigma_u^2} X'X$, we obtain the non-informative prior and the non-informative posterior distributions. In this case

$$\begin{cases} \hat{\mu} = \hat{A} = (X'X)^{-1} X'Y \\ \Sigma_3 = \Sigma_1 = \sigma_u^2 (X'X)^{-1} \end{cases}. \tag{11''}$$

If the variance of the errors σ_u^2 is unknown, we apply first (Ciuiu, 2013) Bayesian inference for $\delta = \frac{1}{\sigma_u^2}$. We take into account that if we denote by S^2 the estimator of σ_u^2 presented in Jula and Jula (2012) namely the sum of squares of residues divided by $m = n - k - 1$ (the number of degrees of freedom), and we consider prior distribution $\delta \sim \Gamma(a, b)$ we obtain the posterior distribution

$$\delta | S^2 \sim \Gamma\left(\frac{m}{2} + a, \frac{2b}{m \cdot b \cdot S^2 + 2} \right). \tag{12}$$

In the case of non-informative prior distribution we have a=1 and $b \to \infty$, and from here

$$\delta | S^2 \sim \Gamma\left(\frac{m}{2} + 1, \frac{2}{mS^2} \right). \tag{12'}$$

We generate a big number of δ (for instance 10000) having the above Γ distribution, and from here $\sigma_u^2 = \frac{1}{\delta}$. Methods to generate a Gamma random variables can be found in Văduva (2004).

For each generated σ_u^2 we have to apply (11), (11') and (11''), which is difficult due to the use of σ_u^2 for estimating Σ_3.

For this reason we perform first the simultaneous diagonalization of $\Sigma_4 = (X'X)^{-1}$ and Σ_2.

We diagonalize first Σ_4 using the method of rotations:

$$U_1' * \Sigma_4 * U_1 = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_{k+1} \end{pmatrix}. \tag{13}$$

Considering $C_1 = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{\sqrt{\lambda_{k+1}}} \end{pmatrix} U'$, we obtain $C_1 \Sigma_4 C_1' = I_{k+1}$. Denoting by

$\Sigma_5 = C_1 \Sigma_2 C_1'$, we perform next the same above method to diagonalize Σ_5:

$$U_2' * \Sigma_5 * U_2 = \begin{pmatrix} \tau_1^2 & 0 & ... & 0 \\ 0 & \tau_2^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \tau_{k+1}^2 \end{pmatrix}. \quad (13')$$

Consider now $C = U_2' C_1 = U_2' * (sqrt\Lambda)^{-1} * U_1'$, which is a change of coordinates with $C^{-1} = U_1 * sqrt\Lambda * U_2$. Of course, $(sqrt\Lambda)^{-1}$ is the above diagonal matrix in (13'). We reduce in this way the problem to $k+1$ independent normal variables having the expectations $\tilde{\theta}_j$ and variance σ_u^2, and the prior distributions for $\tilde{\theta}_j$ being $N(\tilde{\mu}_j, \tau_j^2)$.

We have denoted above

$$\begin{cases} \tilde{\theta} = C * A \\ \tilde{\mu} = C * \mu \end{cases}. \quad (14)$$

3. Application

Example: *Consider the linear regression between* $X = D(SNN)$*, the difference between the current value of net income and the previous, and* $Y = D(EP)$*, the same about the household savings. The data are (Jula and Jula, 2012) from January 1997 and August 2003. The linear regression obtained in (Jula and Jula (2012) is*

$$Y = 0.903487 + 3.23 \cdot X.$$

For Bayesian inference consider for prior informative distribution of coefficients μ_0 *having the values 0.5 and 2, and for* μ_1 *having the values 2 and 5 (four possibilities). For the same prior distribution we consider* $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$ *or* $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$ *(two possibilities). For the informative prior distribution for* $\delta = \frac{1}{\sigma^2}$ *consider two cases: a=1 and b=3, and a=5 and b=100. Of course, consider also for coefficients and δ non-informative prior distributions. For forecasting Y if X is given consider X=-1.202, resulting from reducing the income in September 2003 by 25%.*

Using our C++ program we obtain the linear regression

$$Y = 0.90342 + 3.23042 \cdot X,$$

with the variance of the errors $\sigma^2 = 1.19539$. It results the variance-covariance matrix $\Sigma_1 = \begin{pmatrix} 0.01768 & -0.04565 \\ -0.04565 & 0.81747 \end{pmatrix}$.

If we consider known variance of errors $\sigma^2 = 1.19539$, we obtain the following results.

Table 1: The results in the case of known variance of the residues.

Prior μ	Prior Σ_2	Posterior equation	Posterior Σ_3	Forecasted Y	Variance of forecast
Non-informative		Y=0.90342+3.23042X	$\begin{pmatrix} 0.01768 & -0.04565 \\ -0.04565 & 0.81747 \end{pmatrix}$	-2.97954	2.5039
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	Y=0.88482+2.57542X	$\begin{pmatrix} 0.01369 & -0.01307 \\ -0.01307 & 0.3027 \end{pmatrix}$	-2.21084	1.67783
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	Y=0.90377+3.09606X	$\begin{pmatrix} 0.0172 & -0.04105 \\ -0.04105 & 0.73185 \end{pmatrix}$	-2.8181	2.36865
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	Y=0.78923+4.42205X	$\begin{pmatrix} 0.01369 & -0.01307 \\ -0.01307 & 0.3027 \end{pmatrix}$	-4.52607	1.67783
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	Y=0.88895+3.41221X	$\begin{pmatrix} 0.0172 & -0.04105 \\ -0.04105 & 0.73185 \end{pmatrix}$	-3.21253	2.36865
a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	Y=1.09977+2.19478X	$\begin{pmatrix} 0.01369 & -0.01307 \\ -0.01307 & 0.3027 \end{pmatrix}$	-1.53835	1.67783
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	Y=0.92556+3.11352X	$\begin{pmatrix} 0.0172 & -0.04105 \\ -0.04105 & 0.73185 \end{pmatrix}$	-2.81689	2.36865
a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	Y=1.00418+4.0414X	$\begin{pmatrix} 0.01369 & -0.01307 \\ -0.01307 & 0.3027 \end{pmatrix}$	-3.85359	1.67783
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	Y=0.91114+3.42967X	$\begin{pmatrix} 0.0172 & -0.04105 \\ -0.04105 & 0.73185 \end{pmatrix}$	-3.21132	2.36865

In the above table the coefficients have the prior distribution normal of expectation μ and variance-covariance matrix Σ_2. The posterior values of the coefficients are seen in the column "Posterior equation", and the posterior variance-covariance matrix is Σ_3.

In the case of unknown variance of the errors, we compute first the posterior parameters of $\delta = \frac{1}{\sigma^2}$, and next we generate 10000 variances of residues such that δ has the above posterior distribution $\Gamma(a_1, b_1)$, as in Văduva (2004). The estimator-expectation for σ^2 is equal (by simple computation of $E\left(\frac{1}{X}\right)$) to $\frac{1}{(a_1 - 1)b_1}$, and the estimator-mode for the same parameter is $\frac{1}{(a_1 + 1)b_1}$. The estimator-median is the generated value on the position 5000 after ordering the generated variances increasing.

If the prior distribution for coefficients is non-informative, we generate for each generated σ^2 a set of coefficients normal with expectation Â and variance-covariance matrix $\Sigma_1 = \sigma^2 (X'X)^{-1}$. For this we perform the Cholesky decomposition Σ_1=L*L', where $L = \begin{pmatrix} 0.12162 & 0 \\ -0.31403 & 0.76501 \end{pmatrix}$. Each normal vector is generated using this decomposition as in Văduva (2004).

If we consider informative prior distribution for coefficients, we have to diagonalize simultaneously $(X'X)^{-1}$ and Σ_2. The eigenvalues of $(X'X)^{-1}$ are 0.68602 and 0.01262, with the eigenvectors (-0.05681, 0.99839) and (0.99839, 0.05681). If $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$ we obtain $B = \Lambda_1^{-1} U_1^{'} \Sigma_2 U_1 \Lambda_1^{-1} = \begin{pmatrix} 0.72365 & 0.4574 \\ 0.4574 & 8.20719 \end{pmatrix}$. The eigenvalues of B are 8.23504 and 0.6958, with the corresponding eigenvectors (0.06078, 0.99815) and (0.99815, -0.06078). The matrix C is $C = U_2^{'} \Lambda_1^{-1} U_1^{'} = \begin{pmatrix} 8.86733 & 0.57804 \\ -0.06869 & 1.17243 \end{pmatrix}$, and $C^{-1} = \begin{pmatrix} 0.10908 & -0.05378 \\ 0.05663 & 0.82501 \end{pmatrix}$. In the other case, $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$ we obtain $B = \Lambda_1^{-1} U_1^{'} \Sigma_2 U_1 \Lambda_1^{-1} = \begin{pmatrix} 10.25817 & -1.68194 \\ -1.68194 & 76.29079 \end{pmatrix}$. The eigenvalues of B are 76.3336 and 10.21535, with the corresponding eigenvectors (-0.02545, 0.99968) and (0.99968, 0.02545). The matrix C is $C = U_2^{'} \Lambda_1^{-1} U_1^{'} = \begin{pmatrix} 8.8868 & 0.47487 \\ 0.15761 & 1.21787 \end{pmatrix}$, and $C^{-1} = \begin{pmatrix} 0.11331 & -0.04418 \\ -0.01466 & 0.82682 \end{pmatrix}$.

Other results are presented in the following tables.

Table 2: The results in the case of unknown variance of the residues having non-informative prior distribution for $\delta = \frac{1}{\sigma^2}$

Prior μ	Prior Σ_2	Type of estimator	Posterior equation	Forecasted Y
Non-informative		expectation	Y=0.902+3.23607 X	-2.98776
		mode	Y=0.76269+3.39509 X	-2.83583
		median	Y=0.90266+3.22653 X	-2.97625
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.88442+2.57631 X	-2.21231
		mode	Y=0.92771+2.5072 X	-3.212
		median	Y=0.88542+2.57442 X	-2.20748
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.9053+3.0875 X	-2.80588
		mode	Y=0.96836+2.77777 X	-2.40933
		median	Y=0.90369+3.09174 X	-2.81391
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.78922+4.40509 X	-4.5057
		mode	Y=0.63002+4.60562 X	-4.11863
		median	Y=0.79091+4.39889 X	-4.49261
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.89041+3.41234 X	-3.21122
		mode	Y=0.8138+3.90468 X	-3.87164
		median	Y=0.89103+3.40728 X	-3.20297

a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=1.09666+2.20397 X	-1.55251
		mode	Y=1.08941+2.18817 X	-1.51472
		median	Y=1.09384+2.20764 X	-1.56098
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.92411+3.11851 X	-2.82433
		mode	Y=0.84079+2.16802 X	-2.49065
		median	Y=0.92246+3.11852 X	-2.82034
a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=1.00469+4.04444 X	-3.85673
		mode	Y=0.9859+3.47216 X	-3.81219
		median	Y=1.0049+4.04686 X	-3.86468
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.9092+3.44817 X	-3.2355
		mode	Y=0.90921+4.03666 X	-3.55674
		median	Y=0.9094+3.44974 X	-3.23115

In the above table the posterior distribution of δ is $\Gamma(a_1, b_1)$, where a_1= 39.5 and b_1= 0.02173. The estimators for the variance of the errors are: expectation= 1.19539 (equal to the classical non-Bayesian estimator), mode= 1.13636 and median= 1.17769. The variance-covariance matrix of the coefficients, the variance of the forecast, the skewness, the kurtosis and the Jarque-Bera statistics are presented in next table.

Table 3: The variance, the skewness, the kurtosis and the Jarque-Bera statistics in the case of unknown variance of the residues having non-informative prior distribution for

$$\delta = \frac{1}{\sigma^2}$$

Prior μ	Prior Σ_2	Variance of forecast	Posterior Σ_3	Skewness	Kurtosis	Jarque-Bera
Non-informative		1.28481	$\begin{pmatrix} 0.01816 & 0.12054 \\ 0.12054 & 0.80052 \end{pmatrix}$	-0.00176 0.03272 -0.03719	3.05443 3.12377 3.1204	1.2397 8.16752 8.34475
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.48598	$\begin{pmatrix} 0.01395 & 0.0653 \\ 0.0653 & 0.30564 \end{pmatrix}$	-0.04607 -0.00634 0.00945	3.11167 3.0661 3.05858	8.73362 1.88722 1.57862
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.20113	$\begin{pmatrix} 0.01684 & 0.11254 \\ 0.11254 & 0.7524 \end{pmatrix}$	0.04811 -0.01344 0.01127	3.10578 3.08877 3.10605	8.52021 3.58423 4.89765
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.49775	$\begin{pmatrix} 0.01374 & 0.06527 \\ 0.06527 & 0.31033 \end{pmatrix}$	-0.08871 0.03205 -0.04917	3.10704 3.00543 3.00881	17.88942 1.72448 4.06205
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.15873	$\begin{pmatrix} 0.01691 & 0.11054 \\ 0.11054 & 0.72319 \end{pmatrix}$	-0.03908 0.02341 -0.0186	3.06402 3.00127 3.00095	4.2536 0.91397 0.57715

a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.49435	$\begin{pmatrix} 0.01407 & 0.06555 \\ 0.06555 & 0.30569 \end{pmatrix}$	0.06195 -0.0078 0.01357	3.06412 3.00247 2.99903	8.10984 0.10393 0.3072
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.15846	$\begin{pmatrix} 0.01701 & 0.11089 \\ 0.11089 & 0.72252 \end{pmatrix}$	0.03577 0.01681 -0.01171	3.08212 3.08625 3.08735	4.94265 3.5703 3.40727
a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.48921	$\begin{pmatrix} 0.01441 & 0.06699 \\ 0.06699 & 0.3078 \end{pmatrix}$	0.05217 -0.00174 0.00533	3.08159 2.98621 2.94807	7.30996 0.08428 1.17123
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.2009	$\begin{pmatrix} 0.01734 & 0.11399 \\ 0.11399 & 0.74946 \end{pmatrix}$	0.02564 -0.01123 0.0129	3.19008 3.08614 3.1019	16.15029 3.30195 4.60416

If the prior distribution of δ is Γ(a,b) with a=1 and b=3, we obtain the following results.

Table 4: The results in the case of unknown variance of the residues having informative prior distribution Γ(1,3) for $\delta = \frac{1}{\sigma^2}$

Prior μ	Prior Σ_2	Type of estimator	Posterior equation	Forecasted Y
Non-informative		expectation	Y=0.90462+3.21102 X	-2.95502
		mode	Y=0.91052+3.11991 X	-3.2143
		median	Y=0.90641+3.20315 X	-2.94888
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.88605+2.57625 X	-2.2106
		mode	Y=0.78981+2.05943 X	-1.97638
		median	Y=0.88635+2.56989 X	-2.20346
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.90253+3.1066 X	-2.83161
		mode	Y=0.99002+3.4427 X	-3.10217
		median	Y=0.90117+3.10656 X	-2.82846
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.79142+4.42331 X	-4.5254
		mode	Y=0.84444+4.33848 X	-5.13967
		median	Y=0.79151+4.4191 X	-4.52123
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.89008+3.40873 X	-3.20721
		mode	Y=0.74635+3.50713 X	-2.93714
		median	Y=0.89169+3.39708 X	-3.19727
a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=1.09851+2.20222 X	-1.54857
		mode	Y=1.0137+2.73461 X	-0.64828
		median	Y=1.09661+2.19882 X	-1.54981
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.9235+3.12162 X	-2.82869
		mode	Y=0.86342+2.96136 X	-2.69014
		median	Y=0.92134+3.13337 X	-2.83712

a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=1.00453+4.04013 X	-3.85172
		mode	Y=1.13824+4.30146 X	-3.17165
		median	Y=1.00366+4.02535 X	-3.84249
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.91068+3.42253 X	-3.20319
		mode	Y=0.86797+3.60831 X	-3.97759
		median	Y=0.91012+3.42496 X	-3.2029

In the above table the posterior distribution of δ is $\Gamma(a_1, b_1)$, where a_1= 39.5 and b_1= 0.02157. The estimators for the variance of the errors are: expectation=1.20405, mode= 1.14459 and median=1.1861. The variance-covariance matrix of the coefficients, the variance of the forecast, the skewness, the kurtosis and the Jarque-Bera statistics are presented in next table.

Table 5: The variance, the skewness, the kurtosis and the Jarque-Bera statistics in the case of unknown variance of the residues having informative prior distribution Γ(1,3) for

$$\delta = \frac{1}{\sigma^2}$$

Prior μ	Prior Σ_2	Variance of forecast	Posterior Σ_3	Skewness	Kurtosis	Jarque-Bera
Non-informative		1.36424	$\begin{pmatrix} 0.01783 & 0.1231 \\ 0.1231 & 0.85132 \end{pmatrix}$	-0.10684 0.02383 -0.02824	3.22999 3.01217 3.02781	41.06329 1.00832 1.65139
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.47879	$\begin{pmatrix} 0.01386 & 0.0647 \\ 0.0647 & 0.30223 \end{pmatrix}$	-0.00722 0.00906 -0.01588	3.04319 3.0066 2.98524	0.86395 0.15499 0.51127
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.14641	$\begin{pmatrix} 0.01721 & 0.11098 \\ 0.11098 & 0.71572 \end{pmatrix}$	0.02148 -0.00228 0.00525	2.99825 3.04413 3.03792	0.7706 0.82 0.64509
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.50128	$\begin{pmatrix} 0.01383 & 0.06563 \\ 0.06563 & 0.3118 \end{pmatrix}$	-0.05822 -0.00621 0.00504	3.28705 3.00796 3.01164	39.98154 0.09073 0.0989
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.17598	$\begin{pmatrix} 0.01723 & 0.11236 \\ 0.11236 & 0.73292 \end{pmatrix}$	-0.04468 -0.00238 -0.0023	3.03789 3.04879 3.05768	3.92547 1.00147 1.3952
a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.49841	$\begin{pmatrix} 0.01424 & 0.06636 \\ 0.06636 & 0.30997 \end{pmatrix}$	0.08333 -0.04423 0.03996	3.04671 3.02363 3.04594	12.48296 3.49345 3.54052
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.17152	$\begin{pmatrix} 0.01729 & 0.11235 \\ 0.11235 & 0.73078 \end{pmatrix}$	0.06616 -0.03702 0.03381	3.09164 3.02242 3.01972	10.79397 2.49355 2.06709
a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.48433	$\begin{pmatrix} 0.01363 & 0.06465 \\ 0.06465 & 0.30678 \end{pmatrix}$	0.0674 0.07647 -0.07687	3.05432 3.00935 3.02117	8.80024 9.78269 10.03462
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.17571	$\begin{pmatrix} 0.01743 & 0.11304 \\ 0.11304 & 0.73335 \end{pmatrix}$	0.01129 -0.01176 0.01547	3.04995 3.07646 3.06436	1.2521 2.66637 2.12478

If the prior distribution of δ is Γ(a,b) with a=5 and b=100, we obtain the following results.

Table 6: The results in the case of unknown variance of the residues having informative prior distribution Γ(5,100) for $\delta = \frac{1}{\sigma^2}$

Prior μ	Prior Σ_2	Type of estimator	Posterior equation	Forecasted Y
Non-informative		expectation	Y=0.90602+3.222 X	-2.96682
		mode	Y=0.91852+3.01862 X	-4.28484
		median	Y=0.90551+3.2231 X	-2.96909
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.88955+2.60535 X	-2.24208
		mode	Y=0.80773+2.343 X	-2.00258
		median	Y=0.88952+2.59521 X	-2.22985
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.90328+3.10354 X	-2.82718
		mode	Y=0.76664+3.30917 X	-3.14709
		median	Y=0.90341+3.09839 X	-2.82437
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.79518+4.38207 X	-4.47207
		mode	Y=0.80805+3.68571 X	-3.69925
		median	Y=0.79515+4.37829 X	-4.46633
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.88778+3.40671 X	-3.20708
		mode	Y=0.90815+3.65811 X	-3.5421
		median	Y=0.88737+3.39824 X	-3.1964
a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=1.08552+2.2447 X	-1.6126
		mode	Y=1.00806+1.6076 X	-0.95069
		median	Y=1.0845+2.24497 X	-1.60956
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.92318+3.11577 X	-2.82198
		mode	Y=0.9032+2.96871 X	-2.34143
		median	Y=0.92288+3.11475 X	-2.81843

a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	expectation	Y=0.99421+3.99631 X	-3.80935
		mode	Y=1.03466+4.36445 X	-3.58907
		median	Y=0.99387+3.9957 X	-3.80432
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	expectation	Y=0.91176+3.4123 X	-3.18983
		mode	Y=0.91484+2.1325 X	-3.35303
		median	Y=0.91135+3.40977 X	-3.18364

In the above table the posterior distribution of δ is $\Gamma(a_1, b_1)$, where a_1= 43.5 and b_1= 0.02172. The estimators for the variance of the errors are: expectation=1.08312, mode= 1.03444 and median=1.06547. The variance-covariance matrix of the coefficients, the variance of the forecast, the skewness, the kurtosis and the Jarque-Bera statistics are presented in next table.

Table 7: The variance, the skewness, the kurtosis and the Jarque-Bera statistics in the case of unknown variance of the residues having informative prior distribution Γ(5,100) for $\delta = \frac{1}{\sigma^2}$

Prior μ	Prior Σ_2	Variance of forecast	Posterior Σ_3	Skewness	Kurtosis	Jarque-Bera
Non-informative		1.16201	$\begin{pmatrix} 0.01559 & 0.10652 \\ 0.10652 & 0.72808 \end{pmatrix}$	0.02196 0.01947 -0.013	3.15801 3.06705 3.06821	11.20682 2.50522 2.22004
a=0.5, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.46006	$\begin{pmatrix} 0.0125 & 0.06007 \\ 0.06007 & 0.28883 \end{pmatrix}$	-0.00489 -0.01364 0.00361	3.02435 3.10312 3.0841	0.28699 4.74061 2.96861
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.08279	$\begin{pmatrix} 0.0158 & 0.10313 \\ 0.10313 & 0.67342 \end{pmatrix}$	0.01382 -0.00602 0.00797	3.08209 3.10942 3.10368	3.12609 5.04905 4.58495
a=0.5, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.47193	$\begin{pmatrix} 0.0125 & 0.06007 \\ 0.06007 & 0.28883 \end{pmatrix}$	-0.0207 0.04181 -0.0481	3.09895 3.04265 3.05061	4.79423 3.67185 4.92375
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.06263	$\begin{pmatrix} 0.01532 & 0.10078 \\ 0.10078 & 0.66341 \end{pmatrix}$	0.00022 0.03944 -0.03991	3.12762 3.06892 3.06407	6.78647 4.57159 4.36482
a=2, b=2	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.45979	$\begin{pmatrix} 0.01307 & 0.06198 \\ 0.06198 & 0.29414 \end{pmatrix}$	0.07034 0.0371 -0.03672	3.0376 2.97698 2.96349	8.83525 2.51423 2.80265
	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.08491	$\begin{pmatrix} 0.01589 & 0.1034 \\ 0.1034 & 0.67304 \end{pmatrix}$	0.01992 -0.04522 0.0474	3.07824 3.10464 3.12117	3.21232 7.97048 9.86196
a=2, b=5	$\begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$	0.47113	$\begin{pmatrix} 0.01277 & 0.06158 \\ 0.06158 & 0.29697 \end{pmatrix}$	0.05077 0.00831 -0.00665	3.01931 2.95357 2.96382	4.45085 1.01324 0.61904

	$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$	1.08789	$\begin{pmatrix} 0.01605 & 0.10424 \\ 0.10424 & 0.6773 \end{pmatrix}$	0.02511 0.0138 -0.00879	3.05715 3.08937 3.08589	2.4117 3.64543 3.20246

4. Conclusions

For Bayesian estimation and Bayesian forecasting we consider first X|θ for estimation and Y|θ for forecasting. The common approach is to consider a statistics instead of X and a random variable instead of Y. In the case of Bayesian regression we have considered in this paper the forecast using regression from Jula and Jula (2012). The different approach in the two cases of known/ unknown variance of residues have been considered because the classical statistics for coefficients $\left(X'X\right)^{-1}$ is not compulsory. That's why we perform the simultaneous diagonalization to reduce the multi-parameter Bayesian inference to a uni-parameter one.

Due to the Monte Carlo methods, the forecast is equal to $A_0 + A_1 \cdot X$ with X=-1.202 only for estimators-expectation, and close to this value in the case of estimators-median of coefficients. In the other case (estimators-mode) the differences arrise from the method to estimate the mode for a mixture distribution: the coefficients and forecasts are ordered increasing, and the mode is for all three values the middle of the smallest interval.

We notice that the posterior variances (the posterior variance-covariance matrix of coefficients and the variance of forecast) in Table 1 do not depend on prior Σ_2, hence posterior Σ_3 and the variances in the last column are the same with period two, except the case of non-informative prior. The explanation is that the posterior variances do not depend on prior expectation of coefficients. That's why for the same Σ_2 we have the same Σ_3 and the same variance of forecast.

The same thing in the case of informative prior coefficients can be said about Σ_3, and the small differences arrise from the Monte Carlo methods. Moreover, the values of Σ_3 do not depend on the prior distribution of the variance of errors, namely non-informative, Γ(1,3) and Γ(5,100).

We notice also that in Table 1, the results are closer to the non-informative case for $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$, when the covariance does not change the negative sign from $\Sigma_1 = \begin{pmatrix} 0.01768 & -0.04565 \\ -0.04565 & 0.81747 \end{pmatrix}$.

For the Jarque-Bera test, we compare the Jarque-Bera statistics with the quantile $\chi^2_{2;0.05} = 5.991$. For Table 3, where the prior distribution of the variance of errors is non-informative we have two cases with all Jarque-Bera statistics non-significant: prior variance-covariance matrix of coefficients being $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$ in both cases, and the prior coefficients (0.5,5), respectively (2,2). We have also one case, when the prior expectation and variance-covariance matrix of coefficients are non-informative, for which the Jarque-Bera statistics for slope A_1 and the forecast are significant, and that for intercept A_0 is non-significant. For the other six cases, the Jarque-Bera statistics is significant only for A_0.

For Table 5, where the prior distribution of the variance of errors is $\Gamma(1,3)$ we notice first that we have a case, namely prior coefficients (2,5) and $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$ with all all JB statistics significant, and four cases with JB statistics significant only for intercept A_0: prior non-informative coefficients, prior coefficients (0.5,5) and $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$, and both cases of prior coefficients (2,2). In the other four cases all the JB statistics are non-significant.

For Table 7, where the prior distribution of the variance of errors is $\Gamma(5,100)$ we have only four non-normal cases. In the case with prior coefficients (2,2) and prior variance-covariance matrix $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$ the Jarque-Bera statistics for intercept is non-significant, and the other two ore significant, and in three cases only the Jarque-Bera statistics for intercept is significant: non-informative prior coefficients and Σ_2, prior coefficients (0.5,5) and $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$, and prior coefficients (2,2) and $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$.

An explanation of decreasing the number of non-normal case could be the "distance" from non-informative prior variance of residues. If we define the distance between two prior distributions of $\delta = \frac{1}{\sigma^2}$ as the Euclidean distance between the pairs $\left(a, \frac{1}{b}\right)$, where the prior parametyers of δ are *a* and *b*. In the case of Table 3, this distance is zero, in the case of Table 3 the distance is $\frac{1}{3}$ (a=1 as in the non-informave case), and in the case of Table 7 $dist = \sqrt{16.0001} = 4.0000125$. Perhaps a consistent prior information means known variance of errors, when obviously we have normal distributions. An open problem is to test what happens if we increase the distance, if we increase only *a*, or only $\frac{1}{b}$.

For prior variance-covariance matrix of coefficients $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$ we notice different behavior, in the last case obtaining closer results to the non-informative case. The explanation is from the correlations: -0.37972 in the non-informative case with known variance of errors, 0.08944 for $\Sigma_2 = \begin{pmatrix} 0.1 & 0.02 \\ 0.02 & 0.5 \end{pmatrix}$, and -0.18898 for $\Sigma_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 7 \end{pmatrix}$.

References

Ciuiu, D. (2013), "Bayes Signification Tests in Linear Regression and Economic applications". *Scientific Journal Mathematical Modelling in Civil Engineering*, Vol. 9, No. 1, pp. 16-29.

Congdon, P. (2001), *Introduction to Bayesian Modelling*. John Wiley & Sons.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2000), *Bayesian Data Analysis*. Chapman & Hall/ CRC, Boca Raton, London, New York, Washington.

Iatan, I.F. (2005), "Classification Using Bayesian Approach: Gaussian Case". *Analele Universitatii Bucuresti*, No. 1, pp. 55-64.

Jula, D. and Jula, N. (2012), *Introducere în econometrie*, Ed. Expert, Bucharest (English: Introduction to Econometrics).

Preda, V. (1992), *Teoria deciziilor statistice*. Ed. Academiei, Bucharest (English: The Theory of Statistical Decisions).

Văduva, I. (2004), *Modele de simulare*. Ed. Universităţii Bucureşti (English: Simulation Models).