

## PalArch's Journal of Archaeology of Egypt / Egyptology

### Econometric Analysis Of Macroeconomic Indicators Of The Republic Of Karakalpakstan (On The Example Of Trading)

*Abdullaev Ulmas Alisherovich*

KCO AS RUz Applicant, Karakalpak Scientific Research Institute of Natural Sciences,  
Uzbekistan

**Abdullaev Ulmas Alisherovich: Econometric Analysis Of Macroeconomic Indicators Of The Republic Of Karakalpakstan (On The Example Of Trading) -- PalArch's Journal Of Archaeology Of Egypt/Egyptology 17(7). ISSN 1567-214x**

**Keywords: linear regression, Ridge regression, LASSO**

#### **ABSTRACT**

This article provides a correlation and regression analysis of the main economic indicators of the Republic of Karakalpakstan using Ridge regression, LASSO and least squares methods. The actual and theoretical values of the gross regional product of the Republic of Karakalpakstan were studied and the obtained statistical data were checked for the presence of a normal distribution law according to the Harke-Bar criterion.

#### **1. Introduction**

One of the urgent tasks for the comprehensive development of the Republic of Karakalpakstan includes the use of modern technologies in the field of trade, ensuring the safety of business people and their activity, liberalization of the management system of trade in a market environment [1].

The use of modern information technologies in trade enterprises is a requirement of the time. In recent years, government programs for the development of small and family businesses have been developed [2].

According to the data of the statistical office [3] of the Republic of Karakalpakstan, of the gross regional product (GRP) is agricultural production 27.3%, industrial production 39.4%, trade networks and services 33.3%. In these statistics, the average refers to the network of commerce and services. Therefore, the study and analysis of macroeconomic indicators of GRP for the development of trade and services is timely for increasing the efficiency and systematization of trade networks [4-8].

## 2. Least Square Method

Data processing in engineering, economics, medicine and other sciences often leads to the need to determine unknown parameters of a function of the form

$$y = f(x_0, x_1, \dots, x_m; \beta_0, \beta_1, \dots, \beta_m) + \varepsilon$$

In practice, the most often used is a linear relative function of unknown parameters, i.e.

$$y = \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \quad (1)$$

Where,  $x_0 = 1$ . Let us write equations (1) in matrix form

$$Y = BX + \varepsilon \quad (2)$$

where  $Y = [y_1, y_2, \dots, y_n]^T - (n \times 1)$  dimensional matrix or column vector of

values, endogenous variable,  $X = \begin{bmatrix} x_{11} & \dots & x_{1(m+1)} \\ \vdots & & \vdots \\ x_{(m+1)1} & \dots & x_{n(m+1)} \end{bmatrix} -$  deterministic

matrix of regressors,  $\beta - m + 1 -$  dimensional column vector of unknown parameters;  $\varepsilon$  - a column vector of unobservable uncorrelated observation errors.

In this case, the parameters are easily calculated using the least squares method

$$B = (X'X)^{-1} X'Y \quad (2)$$

if the matrix  $C = (X'X)$  non-degenerate. Following this, we obtain an empirical regression equation of the form

$$Y = XB. \quad (3)$$

The magnitude is the theoretical value that corresponds to the regressors  $x_m, m = 1, n$ .

OLS prediction error

$$RSS = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

Total sum of squares

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5)$$

Determination coefficient

$$R^2 = 1 - \frac{RSS}{TSS} \quad (6)$$

$R^2$  shows the quality of the forecast, the closer the coefficient of determination is to one, the better the regression model.

### 3. LASSO method and Ridge regression

Researchers are often faced with multicollinearity problems. This occurs when there is a strong correlation between the input factors. Usually, to eliminate multicollinearity, a variable is excluded from the model of one or a number of correlated variables. This exclusion of variables will require a number of economic considerations. In a number of works [9] it is proposed to use the LASSO method, which makes it possible to invalidate some variables in the model. The LASSO method is described as follows. Let a linear regression equation of the form be given:

$$Y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_m x_{mt} + \varepsilon_t \quad (7)$$

Where,  $\beta$  - unknown parameters;  $\varepsilon$  - unobservable uncorrelated observation errors.

It is required to solve the minimization problem

$$\min_a \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{k=1}^k |\beta_k| \quad (8)$$

Where,  $\lambda > 0$  regularization parameter.

We minimize not only the sum of the squares of the residuals, but we minimize the sum of the squares of the residuals plus the penalties of audible large coefficients. Regularization parameter  $\lambda$  usually found using the cross-validation method [10]. The LASSO method performs regularization of the coefficients and equates some of the coefficients to zero for a sufficiently large value  $\lambda$ . For details on the LASSO method, see the book [11]. Ridge regression is called Ridge regression in English literature. The Ridge regression method minimizes the following expression:

$$\min_a \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{k=1}^k |\beta_k|^2 \quad (9)$$

for anyone  $\lambda > 0$  exist  $f(\lambda) > 0$ , at which  $\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$ , at

$$\sum_{i=1}^n |\beta_i|^2 < f(\lambda)$$

. This means that for large regularization parameters, the vector estimation norm decreases.

#### Application of the method of least squares for the analysis of macroeconomic indicators in R

We load macroeconomic indicators into the R math package with the following command:

```
m1<- read.xlsx(file = "D:/R-programmer/STATYA.xlsx", sheetIndex=2)
```

The work uses the following notation for macroeconomic indicators:

YHM-gross regional product million soums

AS-Population million

SM- Industrial product million soums

QM - Agricultural product million soums

XZ-services million soums  
 MSS- Net tax on products million soums  
 AKI - Investment in fixed capital million soums  
 CHSA- Retail turnover mln. Soums.  
 CHSAA-Retail trade turnover per capita million soums  
 KS- Number of enterprises, thousand units  
 XIX- The volume of foreign investment million soums.  
 XIKS-Quantities of the enterprise with foreign investments units.

Let's construct a linear regression equation using the command in R:

```
> model_g<-lm(data=g, YHM~AS+.)
> summary(model_g)
Residual standard error: 96.07 on 10 degrees of freedom
Multiple R-squared: 0.9998, Adjusted R-squared: 0.9996
F-statistic: 5647 on 9 and 10 DF, p-value: < 2.2e-16
```

By analyzing the data obtained with the above package R, the estimated coefficients that meet all the criteria. It turns out that among the coefficients of indicators, the coefficient at MSS is significant. The least squares method gives a good estimate, but in cases where multicollinearity is involved, the obtained estimates of the coefficients will not give good significance. To eliminate multicollinearity and obtain an adequate model, additional ones are required, such as the LASSO method and Ridge regression. Next, let's see each method separately using the R math package.

For multicollinearity analysis, the values of the increased variance of the coefficients are calculated. For this we use the VIF command from the car library of the R package.

**Table 1. Coefficients of increasing variance**  $VIF_j, j = \overline{1,11}$ .

vif(model_g)						
AS	SM	QM	XZ	MSS	AKI	
693.52009	5423.42829	710.08299	366.16450	5303.60658	1320.13053	
CHSA	CHSAA	KS	XIH	XIKS		
532.23451	444090.93376	387434.22703	185.58201	35.36426		

The values  $VIF_j$  are presented in table-1, where we see that almost all these values  $VIF_j > 10$  great. This means that there is multicollinearity between the resulting variables.

### Application of the LASSO method for analyzing macroeconomic indicators in R

Introduce new matrices in R and define its character as follows:

```
y1<-g$YHM
x01<- model.matrix(data=g, YHM~AS+.)
```

Where,  $x01$ - regressor matrix,  $y1$  - actual GRP value.

We use glmnet command to estimate odds. This command is used to evaluate the LASSO method. Before evaluating the coefficients, we need to enter the range of variation of the regularization parameter

```

lambdas<-10^seq(7,-3, lenght=200)
g_lasso<-glmnet(x01, y1, alpha = 1, lambda=lambdas)
    
```

where the regularization parameter  $\lambda \in [10^{-3}, 10^7]$ .

The relationship between the estimates of the coefficients and the values of the regularization parameter is shown in Figure-1. The relationship between the coefficients and  $\lambda$  shown in Figure 1: a) the horizontal direction indicates the value  $\log \lambda$  and along the vertical values of the coefficients. The figure shows that if the value  $\lambda$  increases, then the value of the coefficients tends to zero;

b) horizontally specified L1Norm values, which indicates values  $\sum \beta$ . If the value of the sum is in the interval (40.50), then the values of the coefficients will be around zero; c) the relationship between variance values and coefficient values is indicated here; d) the dependence of the mean square errors on  $\log \lambda$ .

The upper part of the figure indicates the number of non-zero  $\beta_i$ .

The package contains cv.glmnet (cross validation), which, based on statistical data, will make it possible to determine the corresponding value  $\lambda$ . We describe this function as:

```

kv11_cv<-cv.glmnet(x01,y1, alpha=1)
plot(kv11_cv)
kv11_cv$lambda.min
kv11_cv$lambda.1se
    
```

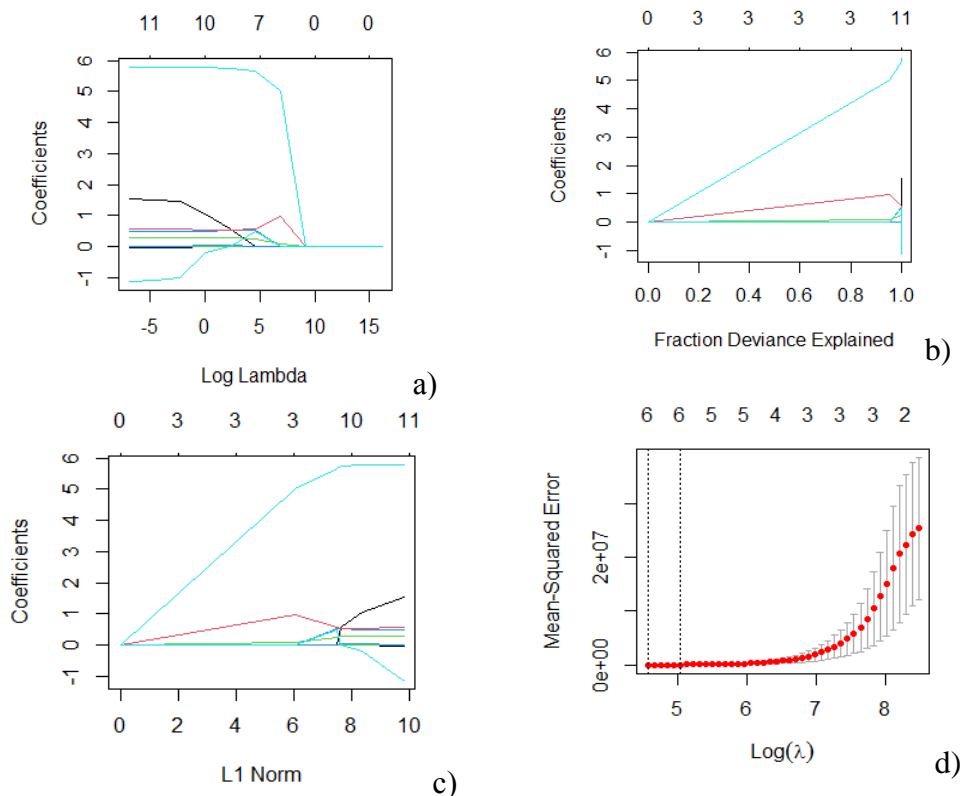


Fig. 1. Graphical interpretation using the LASSO method

Using the function `cv.glmnet`, you can determine the optimal value  $\lambda$ , which is defined as:

```
kv11_cv$lambda.min
kv11_cv$lambda.1se
> kv11_cv$lambda.min
[1] 96.15577
> kv11_cv$lambda.1se
[1] 139.5056
```

### Applying the Ridge Regression Method to Analyze Macroeconomic Indicators in R

In R, enter the following commands:

```
lambdas<-10^seq(5,-2, length=100)
g_rr<-glmnet(x01, y1, alpha = 0, lambda=lambdas)
```

The `lambdas` variable varies in range. `G_rr` stands for Ridge Regression Model. After introducing the necessary functions and commands, we received the following estimates in R.

**Table 2. Estimation of the coefficients of the model of the gross regional product of the Republic of Karakalpakstan in three methods**

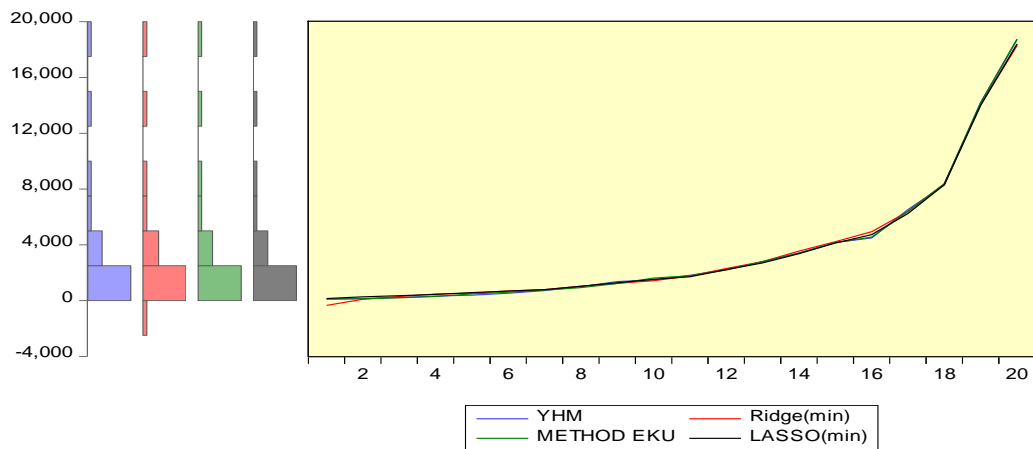
Method and coefficient	MNK	LASSO(min)	Ridge(min)
$b_0$	4234.86607	-87.15721635	-2714,33
$b_1$	-3.07876	0	1,43
$b_2$	-0.82183	0	0,23
$b_3$	0.43773	0.33047901	0,42
$b_4$	-0.02397	0.35354290	0,29
$b_5$	9.13258	4.96785683	2,08
$b_6$	-0.60035	0	0,13
$b_7$	0.40034	0	-0,11
$b_8$	6.85502	0.81405077	0,34
$b_9$	-8.07728	0	0,60
$b_{10}$	0.03954	0.01159877	0,01
$b_{11}$	-2.61116	1.83029696	12,82

After evaluating the coefficients, we were able to identify the regression equations of the model of the gross regional product of the Republic of Karakalpakstan.

**Table3. Regression Equations in Three Methods**

Method	Regression equation
MNK $R^2 = 0,9996$	$YHM = 4234,86607 - 3,07876 \cdot AS - 0,82183 \cdot SM + 0,43773 \cdot QM - 0,02397 \cdot XZ + 9,13258 \cdot MSS - 0,60035 \cdot AKI + 0,40034 \cdot CHSA + 6,85502 \cdot CHSAA - 8,07728 \cdot KS + 0,03954 \cdot XIH - 2,61116 \cdot XIKS$ <small>0,23      0,557      1,641      1,432      0,085 2,253      1,684      1,359      0,5189      0,455 0,696      0,573</small>
LASSO(min) $R^2 = 0,9999$	$YHM = -87,15721635 + 0,33047901 \cdot QM + 0,35354290 \cdot XZ + 4,96785683 \cdot MSS + 0,81405077 \cdot CHSAA + 0,01159877 \cdot XIH + 1,83029696 \cdot XIKS$
Ridge(min) $R^2 = 0,9998$	$YHM = -2714,33 + 1,43 \cdot AS + 0,23 \cdot SM + 0,42 \cdot QM + 0,29 \cdot XZ + 2,08 \cdot MSS + 0,13 \cdot AKI - 0,11 \cdot CHSA + 0,34 \cdot CHSAA + 0,60 \cdot KS + 0,01 \cdot XIH + 12,82 \cdot XIKS$

Graphic interpretation of the three methods is shown in Figure 2.



**Figure: 2. Diagram of theoretical calculations for 2000-2019**

Table 4 shows a visual analysis of GRP between actual and theoretical values with the following statistics: expectation, median, skewness, kurtosis, and Harke-Bera test.

**Table 4. Dependence of actual and theoretical values**

Variables	The average	Median	Asymmetry	Excess	Criterion Harke-Bera
YHM	3641,305	1647,900	1,950781	5,975324	20,06228 P=0,00004
MHK	3641,200	1693,988	1,950785	5,975403	20,06273 P=0.00004
LASSO	3641,305	1610,413	1,954309	5,967409	20,06902 P=0.00004
Ridge	3641,305	1591,173	1,886109	5,738871	18,10920 P=0,000117

Analysis of the information in the table shows that according to the Harke-Bera criterion, the deviation of the asymmetry coefficients and kurtosis from the confidence upper and lower values does not mean that it is not possible to

analyze the law of normal distribution. It follows from this that the actual and theoretical values have the same distribution laws. With the LASSO model, the five coefficients of the above variables are turned to zero. Regularization of the parameters of these variables will make it possible to increase the number of zeroing coefficients. Using the cross-validation method, we obtained the optimal regularization parameters for estimating the coefficients.

#### 4. Conclusion

Research and analysis, the main macroeconomic indicators of the GRP of the Republic of Karakalpakstan (RK) allows you to obtain the following results:

- the LASSO method allows you to determine essential variables and zeroing the coefficients of irrelevant variables (multicollinearity);
- the LASSO method substantially approaches the actual GRP values when processing statistical data;
- the coefficient of determination has the greatest value according to the LASSO method.
- of the above macroeconomic indicators, four variables turned out to be significant: retail turnover per capita (CHSAA), net product tax (MSS), agricultural product (QM) and foreign investment (XIH), which have positive coefficients.

In general, the results of studies using OLS, Ridge and LASSO showed the possibility of using it in studying the relationship between macroeconomic indicators of trade development in the Republic of Kazakhstan. Regression analysis of macroeconomic indicators using LASSO methods confirmed its broader capabilities compared to conventional regression analysis. The results obtained with its help are quite satisfactory in their accuracy and are consistent with the statistics of trade. The results of the studies performed allow us to draw important conclusions on the effectiveness of using these methods for statistical modeling of macroeconomic indicators in trade. Using the LASSO method reduces the initial information represented by multicollinear input variables. As a result, it seems possible to proceed to modeling the processes under study using a limited set of new variables. This makes it possible to significantly simplify the procedure of economic and statistical analysis and make it more efficient. Having the ranked sequences of the influence of generalized variables, one can choose the most effective measures to improve the performance of the trade enterprise. Economic and mathematical models of labor productivity of trade and the cost of goods, obtained using the LASSO method, can be successfully used to improve the economic and statistical analysis of the activities of a trade enterprise and the Republic of Kazakhstan in general. The accuracy of the estimates for these models is quite high (the coefficient of determination ranges from 0.96-0.99; the average approximation error is at the level of 6-16%), the transition to ordinary regression equations with the entire set of initial factors does not significantly improve these estimates.



## References

- [1] Decree of the President of the Republic of Uzbekistan "On the State Program for the Implementation of the Action Strategy in Five Priority Areas of Development of the Republic of Uzbekistan in 2017-2021 in the Year of Development of Science, Education and the Digital Economy" dated March 2, 2020. (LexUZ system).
- [2] Resolution of the President of the Republic of Uzbekistan. On additional measures to finance the program "Every family is an entrepreneur" dated May 20, 2020. (LexUZ system).
- [3] Site of statistics department of the Republic of Karakalpakstan. [www.qrstat.uz](http://www.qrstat.uz)
- [4] Abdullaev U.A. Algorithm and mathematical model for predicting GRP of the Republic of Karakalpakstan // Bulletin of KSU №3. Dec 2019
- [5] Abdullaev U.A. Mathematical model for forecasting of gross regional product of the republic of Karakalpakstan // Journal of management value and ethics. India, GMA, No. 3. 2019 y. GIF 0.626
- [6] Abdullaev U.A. The use of a regression model for self-organization of socio-economic systems of industrial trade enterprises // Modern problems of science and education. - 2013. - No. 2; <http://www.science-education.ru/ru/article/view?id=15295>, IF RSCI = 0.813
- [7] Abdullaev U.A. Identification of a mathematical model and research of decision support for self-organization of socio-economic systems of a trade enterprise // Fundamental research. - 2013. - No. 8 (part 1). - pp. 13-18.
- [8] Abdullaev U.A. Technology of product classifications using expert methods // Fundamental research. -2014, No. 11-4. S. 735-738; DOI 10.17513 / fr.35624
- [9] Linear regression analysis: theory and computing / by Xin Yan & Xiaogang Su. World Scientific Publishing Co. Pte. Ltd. 2009.
- [10] An Introduction to statistical Learning with Applications in R / G. James, D. Witten, T. Hastie, R. Tibshirani: Springer 2013.
- [11] The elements of statistical learning. Data Mining, Inference, and Prediction. T. Hastie, R. Tibshirani, J. Friedman. 2017.