IRSTI 06.71.15

Regularization of Predictors of GDP and Individual Sectors of the Economy of the Republic of Kazakhstan

Amanbay Assylbekov[1], Bayan Assylbekova[1], Roland Giese[2]

[1]*Narxoz University, Almaty, Kazakhstan*

[2]*Hochschule Zittau- Görlitz, Germany*

Abstract

The purpose of the study is to identify new, alternative directions for the development of GDP and individual industries in the Republic of Kazakhstan.

Methodology - Applying a limited number of predictors and using obvious patterns, it is possible to build statistically significant models of economic development in a country or region. In the case of Kazakhstan, such models have shown their practical failure. This study proposes an alternative approach aimed at covering as many predictors as possible and abandoning a priori theories and judgments.

In practical terms, the compression (regularization) of predictor set was carried out using the LASSO method (least absolute shrinkage and selection operator). The statistical significance of the selected predictors was investigated using the least squares method. The models were improved using the backward elimination method.

Models of development of GDP, stock market and civil aviation of the Republic of Kazakhstan have been built. The research data frame consists of 8 blocks: socio-demographic indicators, living standards, labor market and wages, prices, national economy, real sector of the economy, trade, financial system, as well as data on the capitalization of the Kazakhstan Stock Exchange.

Novelty (value) of the research - models have been developed for working with high-dimensional data, which are features of developing countries.

Results of the study - according to the results of the study, the stock market index turned out to be sensitive to a wide range of social and macroeconomic indicators: population growth, unemployment, poverty, inflation, investment, devaluation. Our conclusion: the development of the stock market does not require any specific financial measures, it is necessary to deal with the economy as a whole. The volumes of transactions in corporate securities do not have stable external predictors. The main indicators of the republic's civil aviation have stable external predictors. Passenger turnover, passenger dispatch directly depend on: unemployment rate, wages, GDP per capita, various types of services and products, money supply. There are no external predictors for a separate type of aviation work - cargo transportation. Consequently, positive results can be obtained by reforming this particular segment of services.

Key words: gross domestic product (GDP), LASSO, reverse stepwise exclusion method, high-dimensional data, Kazakhstan Stock Exchange (KASE), intellectual capital, money supply

Introduction

Applying a limited number of predictors and using explicit, obvious patterns, it is possible to build statistically significant models of economic development of a particular region or country. In the case of Kazakhstan's GDP, these include: the growth of the oil industry (according to the classification, it is included in the mining industry), the exchange rate of the national currency, and time trends - autocorrelation. At the same time, the regulation of the economy on the basis of such prerequisites has shown its practical inconsistency. The long-term focus on the development of all new oil fields ended in nothing. Numerous interventions by the National Bank of the Republic of Kazakhstan aimed at stabilizing the national currency are yielding very modest results. The implementation of many innovative projects also did not produce a tangible effect in terms of GDP growth.

This study proposes an alternative approach aimed at covering as many predictors as possible and abandoning a priori theories and judgments. In our approach, the emphasis has shifted from statistically significant models to models that are practically more consistent.

The economies of developing, emerging countries have a number of features that must be taken into account when studying them. Following the recommendations and requirements of international organizations, developing countries keep records in a relatively large number of areas (signs). In this regard, there is not much difference between developing and developed countries. On the other hand, the number of observations for each characteristic is limited in developing countries. The main reasons are the limited historical lifetime and gradual convergence with international standards. Thus, there is a need for correct interpretation of high-dimensional data.

Literature review: The theoretical basis of the study

In the context of the role that the financial system plays in a country's economic growth, Schumpeter (1935) argues that well-functioning banks help accelerate economic growth by identifying financing for entrepreneurs with the best chance of successfully supporting innovative production processes. According to research by Demirguk-Kunt and Maksimovich (1996), stock trading conveys information about the company's prospects for potential investors and lenders. According to Levin (1991), the stock market contributes to economic growth due to the ability to trade by the owners of companies without disrupting the production processes occurring within the companies and providing agents with the opportunity to diversify their portfolios. In addition, Greenwood and Smith (1997) argued that a well-developed stock market can reduce the cost of mobilizing savings and thereby encourage investment in most productive technologies.

The relationship between stock market development and economic growth has been a persistent problem in empirical research. Singh Tariqa et al. (2010) examined the random relationship between index returns, employment, exchange rate, GDP, inflation, and money supply for Taiwan. Empirical research results showed that the exchange rate and GDP seemed to influence the returns of all portfolios, while the inflation rate, the exchange rate and the money supply were negatively associated with the return on portfolios of large and medium-sized companies. Tripathi and Kumar (2014) found a positive correlation between stock returns and inflation, GDP growth, exchange rate and money supply in the BRICS countries (BRICS - Brazil, Russia, India, China, South Africa). Muhammad Aamir Ali (2014) argues that GDP per capita is highly dependent on the development of the stock market. His research is based on data from the economies of Asia - India, Pakistan, and China for the period from 1991 to 2011. According to the research results, the capitalization of the stock market has a positive effect on GDP. Kuttner and Mosser (2002) show a positive correlation between real GDP growth and US interest rates between 1950 and 2000. Lee and Werner (2018), comparing the economic performance of the United States, Great Britain with the economies of Germany and Japan, found that nominal GDP growth is positively correlated with short-term and long-term rates in all four countries. Masood and Hardaker (2012) developed an endogenous growth model for 42 countries and stated that the correlation between stock market development and economic growth in emerging economies is positive.

The aviation industry is an important part of the social economy, while at the same time it plays a vital role in terms of the mechanism of economic development. The relationship between civil aviation and economic growth has become the focus of both the industrial field and research subjects. Kalayci and Yazici (2016) examined the impact of US exports and GDP on civil aviation through the introduction of econometric models. As a result, the impact of both export volume and GDP on civil aviation was determined for the period from 1980 to 2012, where there is a long-term relationship between the variables. In addition, it was found that the volume of US GDP exports have a decisive influence on civil aviation. Dobrushkes, Lennert and Van Hamm (2011) stated that the level of the economy in terms of decision-making power, trade volume, tourism, main distance from the air market and GDP are fundamental components of air travel in major European regions.

Thus, the analysis of a large number of countries reveals a statistically significant relationship between financial development and the development of the real sector of the economy. It seems to us that the analysis for a single country is more valuable in the applied aspect.

Methodology

Data

For greater objectivity, the number of features (variables) to be analyzed, in our opinion, should be as large as possible. A priori, the subjective selection of features, biased to the author's point of view, was not performed. Relative to the time horizon; statistically, the more observations, the more reliable the research results.

The work covers up to 209 variables describing the development of the republic's economy over 15 years, from 2003 to 2017. Only in 2003, the data on the capitalization of the Kazakhstan Stock Exchange (KASE) began to be published, which explains the time interval of observations.

Our database is based on indicators of socio-economic development of Kazakhstan - 193 variables, consisting of 8 main blocks: socio-demographic indicators, standard of living, labor market and wages, prices, national economy, real sector of the economy, trade, financial system.

The financial block includes data on the state, republican and local budgets in the context of revenues, costs, deficits; foreign direct investment, average annual US dollar exchange rate[1].

[1] Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan [Electronic resource]. URL: http://stat.gov.kz/ (Date of access: 12/05/2018)

We supplemented the data on the financial system with the following indicators of the National Bank of the Republic of Kazakhstan[2]:

- monetary base (reserve money), M0 (cash in circulation), M3 (money supply), household deposits, loans to the economy, current account, capital account balance, financial account (excluding reserve assets of the NBK).

KASE contains the following indicators[3]:

- KASE capitalization on securities of corporate issuers,
- volume of deals on KASE with government securities,
- volume of deals on KASE with corporate securities, mln. tenge.

Four out of seven Kazakhstani companies included in the KASE index are represented by their own securities and on the London Stock Exchange. These are KAZ Minerals, Halyk Bank, Kcell, Kazakhtelecom. Therefore, the daily dynamics of the Kazakhstani index is significantly tied to the FTSE100. Yearly weighted average indices of KASE and FTSE100 are calculated by us according to information on the websites of KASE and https://ru.investing.com/ and are included in the table.

For the civil aviation of Kazakhstan, the following variables were available to us:

- passenger turnover of public air transport,
- transportation of passengers by public air transport,
- freight turnover of public air transport

Methods

The most common research method when there are responses (supervised learning) is the ordinary least squares method (OLS) and its various modifications and improvements.

[2] National Bank of the Republic of Kazakhstan [Electronic resource]. URL: https://nationalbank.kz/(Access date: 12/20/2018)

[3] Current state of the securities market of the Republic of Kazakhstan National Bank of the Republic of Kazakhstan, Department of the securities market. [Electronic resource]. URL: https://nationalbank.kz//(Access date: 02.20.2019)

In general terms, the procedure of the least squares method consists of finding such parameters that minimize the expression:

$$\underbrace{min}_{a_0,a_j} T = \sum_{i=1}^{n}(y_i - \hat{y})^2 \quad (1)$$

where,

$i = \overline{1,n}$ – observation index

y_i – true values of the variable Y (response),

$\hat{y}_i$ - predicted values of the variable Y, constructed according to the estimated parameters a_0, a_j:

$$\hat{y} = a_0 + \sum_{j=1}^{k} a_j * x_j \quad (2)$$

With all the prevalence of OLS and its modifications, it has application restrictions that we consider necessary to note.

We have built an OLS model for the development of the economy of Kazakhstan. A limited number of main predictors were used, and obvious patterns were considered. The dominant place in Kazakhstan's GDP is occupied by oil production and the development of minerals. Taking into account the time trend of GDP, we also included in the model, as a predictor, GDP for the previous period. The exchange rate of the US dollar against the national currency, the tenge, was also considered. Kazakhstan's GDP in this and all subsequent models was measured in US dollars. The regression output is shown below.

The characteristics of the model indicate its statistical significance. At the same time, the regulation of the economy based on such prerequisites has shown its practical inconsistency. The long-term focus on the development of all new oil fields ended in nothing. Numerous interventions by the National Bank of the Republic of Kazakhstan aimed at stabilizing the national currency are yielding very modest results. The initiation of many innovative projects also did not produce a tangible effect in terms of GDP growth.

Table 1. Modeling GDP using OLS

```
Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max 
-13150  -5463    695   3500  13399 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)             137133       2223  61.690 2.51e-15 ***
xmining.industry         43676       4799   9.102 1.88e-06 ***
xTenge.US.dollar.rate   -20251       2713  -7.465 1.25e-05 ***
xGDP.previous.period     31272       4493   6.960 2.39e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8609 on 11 degrees of freedom
Multiple R-squared:  0.9863,	Adjusted R-squared:  0.9825 
F-statistic: 263.2 on 3 and 11 DF,  p-value: 1.611e-10
```

In our study, we applied the proposed alternative approach aimed at covering as many predictors as possible. In our approach, the emphasis has shifted from statistically significant models to models that are practically more consistent. The practical problems of direct application of OLS are followed by theoretical problems. When the number of trials n, significantly exceeds the number of features p, i.e. n p, the OLS has satisfactory characteristics in terms of both bias and variance (James et al., 2013). But if the excess is insignificant, then noticeable variance can occur, which in turn will lead to the fact that the results obtained on the training data cannot be applied to new (test) data. In the case when $n < p$, the variance tends to infinity and the OLS cannot be used. In our case $209 > p > 200$ and $n = 15$, hence p >> n. Therefore, we can a priori assert that in our study, an attempt to apply OLS and multivariate linear regression will be incorrect and the results will not be statistically significant.

Another, no less important problem of multicollinearity, which is present in our data, also cannot be solved using the least squares method.

To solve these issues, more advanced and modern methods of data mining are used.

Data mining (DM) is the process of detecting patterns in large data sets using methods at the intersection of machine learning, statistics and database systems (Chakrabarti et al.,2006; Hastie et al., 2005).

The actual task of the DM is a semi-automatic or automatic analysis of large volumes of data to extract previously unknown interesting models, such as groups of data records (cluster analysis), unusual records (detection of anomalies) and dependencies (analysis of association rules, serial mining).

In our opinion, the LASSO (least absolute shrinkage and selection operator) method can equally well be used for both regularizing predictors and DM.

Unlike OLS, the expression minimized in LASSO is

$$\underbrace{min}_{a_0,a_j} LR = \sum_{i=1}^{n}(y_i - \hat{y})^2 + \lambda * \sum_{j=1}^{k}\left|a_j\right| \quad (3)$$

Where, λ is a hyperparameter.

As a result of this minimization, most of the unimportant coefficients degenerate to zero, and the most relevant remain.

Penalized regression, especially LASSO, can help researchers interested in predicting outcomes by choosing a subset of variables that minimize prediction error.

There are several other variable selection methods that have traditionally been used to build models, but their usefulness is more limited in the context of large *p*. For example:

1) *Best subset selection* includes testing each combination of variables and choosing the best model based on a set of variables that gives the best values of R^2, *AIC*, *BIC*, *AUC*, standard error, etc. The number of models to be tested, is 2^p, which can be a computational load as the number of predictors increases. For example, in our case, when $p \approx 200$, there are a total of $16069380442589 9 \cdot 10^{16}$ possible models that need to be compared.

2) *Forward selection* builds a prediction model, starting with an empty model, and then adding variables one at a time, checking the improvement that each variable adds to the fit of the model, and stopping the process when additional variables are not related to fitting the model. The maximum number of models that need to be tested is $1 + p(p+1)/2$. Thus, if p ≈ 200, then it is necessary to compare up to 20101 models.

3) *Backward elimination* creates a prediction model, starting with a model with all the variables included, then deleting the variables one at a time, checking the improvement in deleting each variable to match the model, and stopping the process as soon as deleting additional variables does not improve model fit. The number of models that need to be tested is $1 + p(p+1)/2$. As in the case of forward selection, in our case it is necessary to compare up to 20101 models.

A penalized regression can select and predict variables in the Big Data environment more efficiently and effectively than these and other methods. LASSO is based on minimizing the standard error, which is based on balancing the bias and variance, to build the most accurate and stable model.

At the end of the review of methods, we would like to note that today there are more advanced DM methods than LASSO. These include deep learning with neural networks (Goodfellow, 2015). Deep learning is widely used in finance (Heaton et al., 2017; Ding et al., 2015; Fischer and Krauss, 2018). We do not use deep learning due to the lack of available observations for training the neural network. The structure of our data belongs to the class of large dimension, when the number of features significantly exceeds the number of observations. In defense of our approach, we can note that the regularization procedure is an integral part of deep learning.

Findings and Discussion

Model - Regularization of GDP Predictors

The LASSO method is available in most of the modern platforms and languages: Python, R, SPSS, Stata, etc. We have implemented our model in the R language, because it is an open, free resource, there is a ready-made glmnet library (package) from Trevor Hasti suitable for LASSO and ridge regression. For ease of use, we used the RStudio interface. As of today, several manuals have been published on R in Russian (Robert et al., 2014; Mastitsky et al., 2014)

Our first model is designed to regularize the predictors of the GDP of the Republic of Kazakhstan in US dollars. Below is the program code of our model.

```
install.packages("glmnet")
library(glmnet)
install.packages("Matrix")
library(Matrix)
KASE_data = read.csv("Путь к файлу/Файл.csv")
X = model.matrix(KASE_data$ВВП.млн..долларов.США~.-1,KASE_data)
X_scaled = scale(X)
y = KASE_data$ВВП млн долларов США
fit.lasso = glmnet(X_scaled, y)
plot(fit.lasso, xvar='lambda', label=TRUE)
cv.lasso = cv.glmnet(X_scaled, y,nfolds = 5)
plot(cv.lasso)
coef(cv.lasso)
bestlam = cv.lasso$lambda.min
fit.lasso = glmnet(X_scaled, y, alpha = 1, lambda = bestlam)
coef(fit.lasso)
lasso.pred=predict(fit.lasso, s=bestlam, newx = X_scaled)
mean((lasso.pred - y)^2)
```

As a result of modeling, we obtained the following GDP predictors, with the corresponding coefficients (Table 2):

Table 2. Modeling GDP using LASSO

Predictor name	Reductions	Coefficient
Share of registered unemployed in the labor force	Unemployed labor	-6651.6
Youth unemployment rate 15-24 years	Youth unemployment	-14667.2

World price for Brent crude oil, USD	Brent	1221.6
Manufacture of other vehicles	Other vehicles	20582.0
Price index of imported goods	Price imported	-219.1
Number of places in commissioned preschool organizations	Preschool organizations	1844.8

The mining industry, according to the classification of the Committee on Statistics, includes oil production. The resulting model does not include any of the indicators we assumed. Nevertheless, it seems to us more useful predictively; oil production is more primary than GDPT, and the price of oil is higher than the national currency rate. The LASSO method, in this context, has successfully completed the DM task.

To pick the optimal value the hyperparameter λ we perform a grid search and report cross-validated MSE for each value of the hyperparameter λ from the grid. The results (obtained by the plot(cv.lasso) command) are shown in Figure 1.

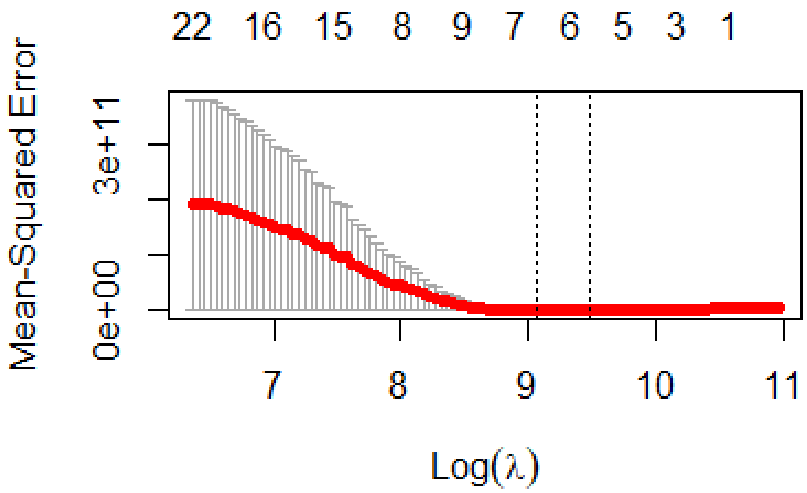

Figure 1. MSE plot by GDP model

The coefficients were obtained under the condition that the hyperparameter λ is equal to the smallest value at the smallest MSE. In the figure, the position is indicated by a left dashed vertical line. If such a condition is not set in the program code, then we get a list of coefficients for larger λ (right dashed vertical line). In this case, the composition of the predictors may change. The predictor compression procedure is shown in Figure 2.

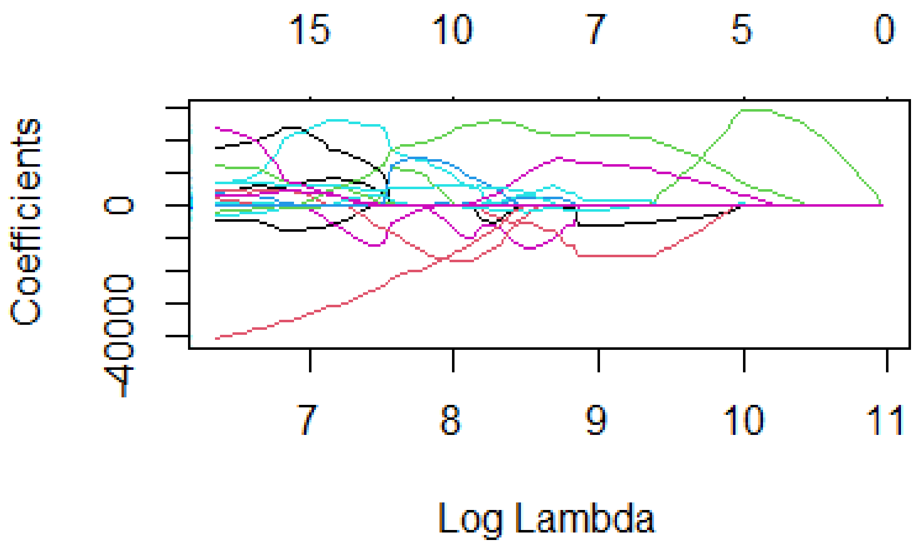

Figure 2. Regularization of GDP model predictors

At the next stage of modeling, we checked the statistical significance of the predictors obtained. For this, the OLS was used. The program code is shown below.

```
GDP_data=read.csv("("Путь к файлу/Файл.csv")")
x=model.matrix(GDP_data$ ВВП млн долларов США ~.-
1,GDP_data)
x=scale(x)
y = GDP_data$ ВВП млн долларов США
lm(y~x)
summary(lm(y~x))
```

Below are the results of checking the statistical significance of the predictors obtained (Table 3).

Table 3. Modeling GDP using LASSO and OLS.

```
Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
-57581  -4095   2788   8489  24927

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                137133       6380  21.496 2.31e-08 ***
xBrent                      16553      11706   1.414    0.195
xOther.vehicles             23999      16373   1.466    0.181
xYouth.unemployment        -14291      35128  -0.407    0.695
xUnemployed.labor           -7643      28022  -0.273    0.792
xPrice.imported             -5055      12025  -0.420    0.685
xPreschool.organizations     3040      26042   0.117    0.910
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24710 on 8 degrees of freedom
Multiple R-squared:  0.9177,    Adjusted R-squared:  0.856
F-statistic: 14.87 on 6 and 8 DF,  p-value: 0.0006008
```

The model was improved by the method of backward elimination. The results are shown below (Table 4).

Table 4. Modeling GDP using LASSO, OLS and backward elimination.

```
Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
-57202  -3069   2652   6510  34866

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           137133       5596  24.504 5.99e-11 ***
xBrent                 18280       8117   2.252   0.0457 *
xOther.vehicles        23932       9694   2.469   0.0312 *
xYouth.unemployment   -27068       9543  -2.836   0.0162 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21680 on 11 degrees of freedom
Multiple R-squared:  0.9129,    Adjusted R-squared:  0.8892
F-statistic: 38.43 on 3 and 11 DF,  p-value: 4.016e-06
```

As a result of modeling, we obtained three statistically significant predictors of GDP in US dollars: the cost of Brent crude, the production of other vehicles, and the youth unemployment rate.

Today, according to the model, despite numerous state industrialization programs, Kazakhstan has a strictly oriented raw material economy. This is indicated by the dependence of GDP on oil prices. The model did not include the production of motor vehicles, which is the focus of the business community and the state. Without going beyond our research, we can assume that the production of other vehicles has a positive effect on the foreign trade balance and the tenge exchange rate. The appearance of youth unemployment in the GDP model has far-reaching implications. And above all, to the outflow of intellectual capital from the country.

Stock Market Modeling Results

In the stock market, we obtained analytical dependencies in terms of capitalization, index, and trading volume. KASE capitalization - CAPITAL, is expressed by the following dependence:

$$CAPITAL = 10548338.2 + 4005296.6\ MonBase + 190912.7 DepPop, \quad (4)$$

Where MonBase - monetary base (reserve money), DepPop - deposits of the population. The results of applying the *plot(cv.lasso)* function according to the KASE capitalization model are presented in Fig. 3.

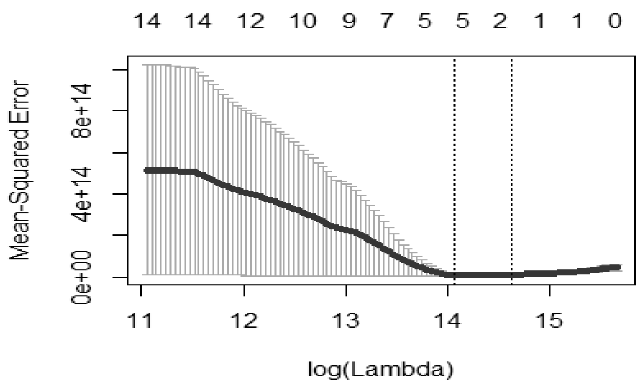

Figure 3. Mean square error of KASE capitalization model

In terms of interpretation, the resulting model is clear and transparent. But it copes very mediocre with the second purpose of modeling – predictive ability. The standard error of the capitalization model is about $6*10^{12}$. The problem may be with the accounting methodology. The National Bank of the Republic of Kazakhstan in the capitalization includes, along with the value of shares, the value of corporate bonds. The results of applying *plot(fit.lasso, xvar = 'lambda', label = TRUE, col = "black")* according to the KASE capitalization model are presented in Fig. 4.

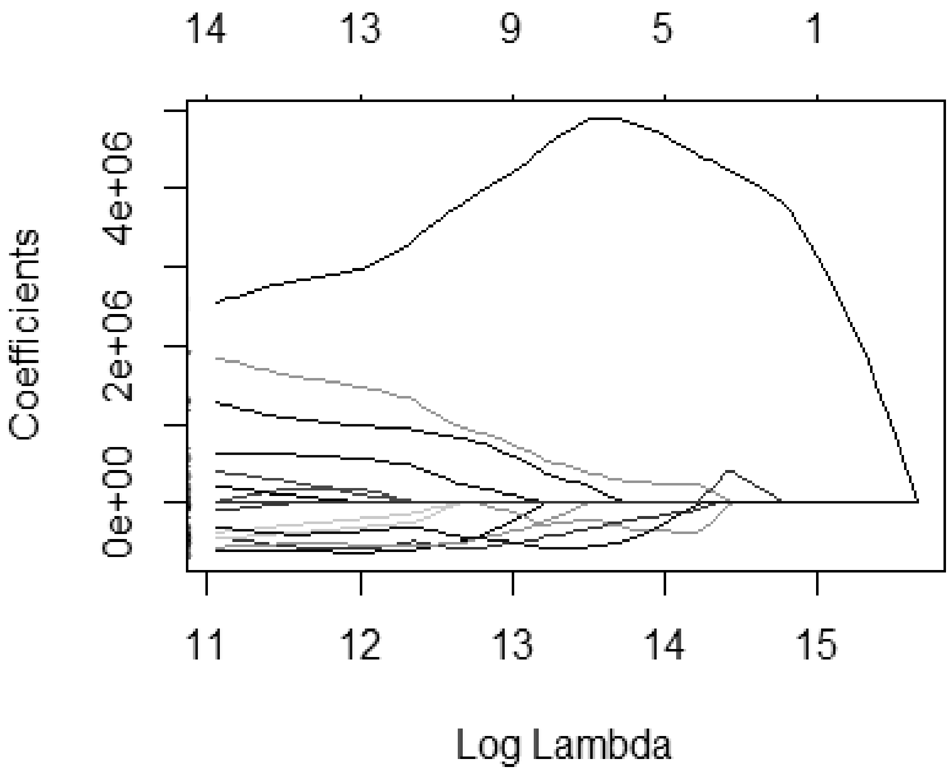

Figure 4. Regularization of KASE capitalization model predictors

Checking the predictors showed that the contributions of the population are not statistically significant: Std. Error = 1594735, t-value = 0.886, Pr (> | t |) = 0.39311. After improving the model by step-by-step elimination and inclusion of predictors, the final model looks like this (Table 5):

Table 5. KASE capitalization model

```
Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-1852611  -816277   -69085   327748  3156812 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    10548338     390091   27.04 4.03e-12 ***
xMonetary.base  7022266     433217   16.21 1.60e-09 ***
xBrent         -1356179     433217   -3.13  0.00868 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1511000 on 12 degrees of freedom
Multiple R-squared:  0.9577,	Adjusted R-squared:  0.9506 
F-statistic: 135.7 on 2 and 12 DF,  p-value: 5.763e-09
```

It should be noted that the inclusion in the model of such predictors as GDP, national currency rate, etc., led to a deterioration of the model.

The dynamics of the movement of the KASE - INDEX is determined by the following predictors and relevant coefficients:

$$\text{INDEX} = 1201.8 + 285.94 LBud + 266.88 ConPrice + 99.9 BuildPrices + 19.74 TariffServices + 249.82 FTSE + 195.43 Employee + 164.35 Invest + 129.93 Population + 101.44 Rate\,\$ + 64.33 Migr - 61.15 ProzhitMin - 38.31 unemployed + 20.91 Scott, \quad (5)$$

Where *LBud* - Local budget, budget deficit (surplus), *ConPrice* - Consumer price index, percent of the previous year, *BuildPrices* - Price index of construction, as a percentage of the previous year, *TariffServices* - Index of tariffs for services, as a percentage of the previous year, FTSE - The average annual stock index of the London Stock Exchange, *Employee* - Self-employed workers, as a percentage of the previous year, Invest - Investments in fixed assets, Population - Population, percentage of the previous year, Rate $ - Average annual exchange first the US dollar, *Migr* - migration balance of all flows, ProzhitMin - The share of the population with incomes below the subsistence minimum, the *unemployed* - the unemployed as a percentage of the previous year, *Scott* - The number of head of cattle. The KASE index MSE chart is presented in Fig. 5. Unlike previous models, the minimum MSE is achieved with the number of predictors equal to 13. With further compression, with an increase in lambda and a decrease in the number of predictors, the model error and its variability increase. The index is very sensitive along with the predictors themselves to the derivatives of the predictors. This applies to price indices, population growth rates, self-employed, unemployed. Such a structure of the model ensures its maximum predictability; the standard error of the model is 1880.92.

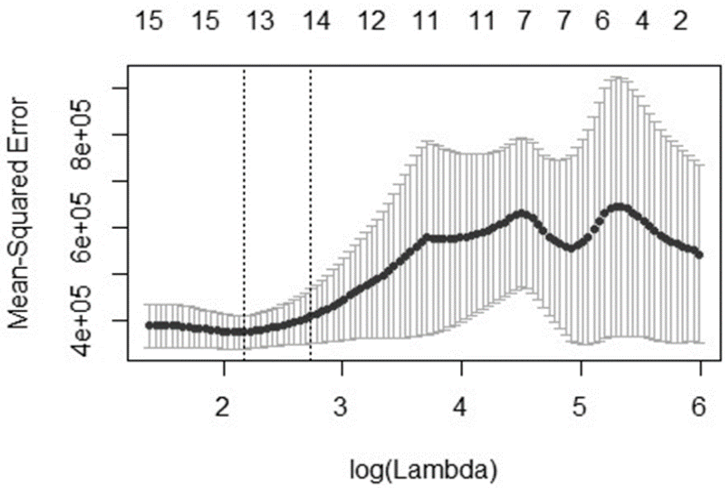

Figure 5. MSE chart of the KASE index model

After checking the predictors for statistical significance, the following were excluded: FTSE, Unemployed, Livestock and LiveMin. An interesting phenomenon: we made several attempts to include the FTSE in the final model, all of which led to a deterioration in its significance. This is despite the fact that the influence of the FTSE on the intraday movement of the INDEX is obvious. For the most part, this connection is carried out through the shares of Kazakhstani companies, which are simultaneously traded on both sites. It turns out, in the long term, the influence of developed stock markets on developing ones, compared to the influence of local predictors, is not so significant. As a result, we got the following model (Table 6):

Table 6. Model of the KASE index

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)           1201.81      26.81  44.827 1.04e-07 ***
xConsumer.price        484.16      47.09  10.282  0.00015 ***
xPrice.construction    144.64      50.58   2.860  0.03543 *  
xTariffs.services      149.96      45.33   3.308  0.02129 *  
xEmployee              563.08     115.25   4.886  0.00453 ** 
xInvestments           651.20      83.96   7.756  0.00057 ***
xPopulation            145.44      54.41   2.673  0.04418 *  
xTenge.US.dollar       335.15      62.27   5.382  0.00299 ** 
xMigration.balance     201.52      41.89   4.811  0.00484 ** 
xLocal.budget          242.44      81.58   2.972  0.03110 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 103.8 on 5 degrees of freedom
Multiple R-squared:  0.9918,	Adjusted R-squared:  0.9772 
F-statistic: 67.54 on 9 and 5 DF,  p-value: 0.0001097
```

The KASE index has shown great sensitivity to investments in fixed assets. Inflation, characterized by price indices and growth in tariffs for services, has an equally significant impact on the index. Investors timely reflect the devaluation of the national currency in the share price.

From year to year, there is a steady increase in the deficit of local budgets, which negatively affects the stock market index. The problem may lie both in inter-budgetary relations, excessive centralization, and in the unprofitableness of local budgets themselves. Solving the problems of local budgets can most beneficially influence the development of the stock exchange.

The inclusion in the index of the growth rate of the number of self-employed workers, in our opinion, speaks of the role of these indicators in the economy. Today, the self-employed population is a growth driver.

Separately, it is worth considering the effect of the migration balance. Since 2012 there is a negative balance of migration in the whole country. Every year, this trend intensifies. In 2018, 41.9 thousand people left the country (a year earlier - 37.7 thousand people), and 12.7 thousand people arrived (a year earlier - 15.6 thousand people)[4]

---

[4] Where Kazakhstanis most often go - research. [Electronic resource]. URL: https://www.zakon.kz/4981777-kuda-chashche-vsego-uezzhayut.html (Date of access: 20.08.2019)

In the first half of 2019, many more people left Kazakhstan than arrived. 20.1 thousand people left - 16.9% more than in the same period last year (17.2 thousand people). In turn, only 5.6 thousand people arrived in Kazakhstan this year - 11.3% less than the same period a year earlier (6.4 thousand people).

Most often they leave for Russia, Belarus, Germany, the USA, Israel and Canada. They come from Uzbekistan, China, Mongolia, Iran and Georgia. Since there is no information on the competencies of departing and arriving, judging by the recipient countries and their migration policies, we can assume that those who depart are specialists who were not needed by the economy of Kazakhstan. There is a depletion of the country's intellectual capital.

The only predictor affecting the volume of trading in corporate shares and bonds to some extent was the KASE index. At the same time, the standard error of the model turned out to be too large, of the order of $2.74*10^{12}$. In our opinion, this situation is explained by the fact that the bulk of the deals on KASE with corporate securities, recently, is conducted in the repo sector (Assylbekov and Assylbekova, 2018). Stock market participants are largely concerned about the increase in short-term liquidity.

Civil Aviation Modeling Results

As a result of modeling, we obtained a linear regression of passenger traffic **(revenue passenger-kilometres)** in civil aviation with 11 predictors. The standard error of the model is 170136.9. The MSE graph of the passenger turnover model in civil aviation is shown in Fig. 6. As can be seen from the figure, the number of regressors decreases with increasing lambda and, in the limit, can be brought to unity. But, in this case, the MSE and its variance begin to grow.

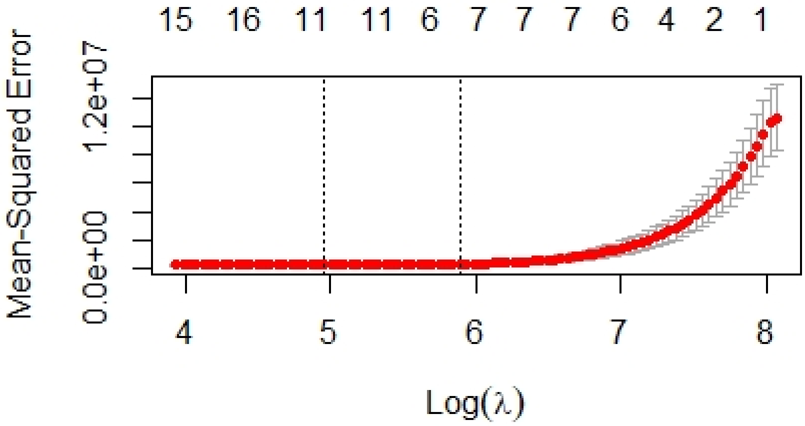

Figure 6. MSE graph of passenger traffic model

After checking the predictors for statistical significance, a model with 5 factors was obtained (Table 7)

Table 7. Model of passenger traffic in civil aviation of the Republic of Kazakhstan

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                7104.44      34.39 206.604  < 2e-16 ***
xYouth.unemployment       -1564.32     290.49  -5.385 0.000442 ***
xnon.metallic.products     1104.35     151.03   7.312 4.51e-05 ***
xM3                       -1952.37     598.92  -3.260 0.009841 ** 
xPublic.deposits           1795.07     445.40   4.030 0.002972 ** 
xGross.agricultural.output  948.20     206.48   4.592 0.001305 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 133.2 on 9 degrees of freedom
Multiple R-squared:  0.999,	Adjusted R-squared:  0.9984 
F-statistic:  1770 on 5 and 9 DF,  p-value: 3.507e-13
```

where: Youth unemployment - the youth unemployment rate of 15-28 years, non-metallic products - production of other non-metallic mineral products, M3 - Broad money, Public deposits - Public deposits at the end of the year.

As a result of modeling, we also obtained a linear regression of passenger transportation in civil aviation with 13 predictors. The standard error of the model is 75971.48. The MSE graph of the model of transported passengers in civil aviation is shown in Fig. 7. We see that the schedule of transported passengers is similar to the schedule of passenger turnover.

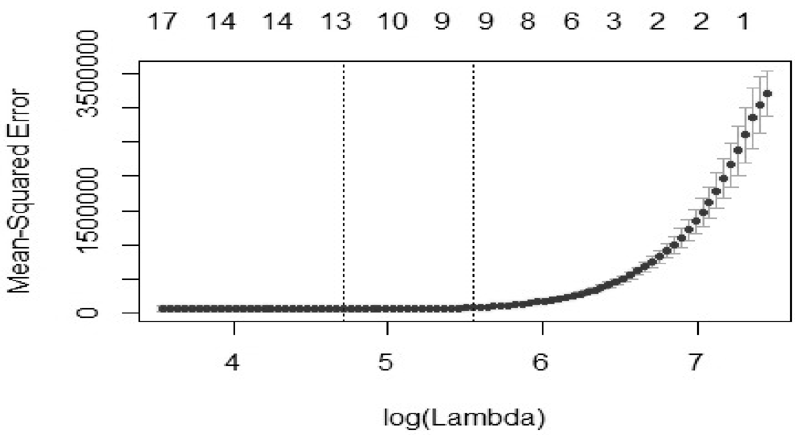

Figure 7. Graph of MSE model of transported passengers.

After checking the predictors for statistical significance, a model with 6 factors was obtained (Table 8). Unlike the previous model, it takes into account the output of the manufacturing industry as a percentage of 1991.

Table 8. Model of transportation of passengers of civil aviation of the Republic of Kazakhstan

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   3691.62      14.38 256.806  < 2e-16 ***
xYouth.unemployment           -703.74     128.46  -5.478 0.000589 ***
xnon.metallic.products         364.84      73.83   4.942 0.001132 **
xM3                           -978.60     252.39  -3.877 0.004692 **
xPublic.deposits              1078.58     190.13   5.673 0.000469 ***
xGross.agricultural.output     449.43      86.78   5.179 0.000844 ***
xManufacturing.industry        240.51      84.58   2.844 0.021693 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 55.67 on 8 degrees of freedom
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.999
F-statistic:  2418 on 6 and 8 DF,  p-value: 1.384e-12
```

The Broad money at the end of the year - M3, as in the previous model, is considered with a negative weight, and deposits of the population - with a positive weight.

Based on the results of modeling cargo turnover, we can conclude that in this case the object has no predictors, since the standard error is a significant amount, in the order of $3.38 * 10^8$ with high variance, and which do not decrease with compression. The transport of goods in Kazakhstan's civil aviation is independent of most external social, economic and financial predictors. Rather, it is subjective, due to the internal problems of the industry's development. Our additional correlation analysis confirms this hypothesis.

Conclusion

Before proceeding to the conclusions, we would like to make two notes, limitations. Firstly, the research was conducted in a developing economy and in the border stock market. Secondly, the identified predictors are external in relation to the objects of study.

The history of the development of the economy of Kazakhstan has clearly shown that statistically significant models and economically sound models are not the same thing.

Our research has led to the following conclusions. Application of the LASSO method is necessary and useful for the purpose of data mining, identifying external predictors of the development of various sectors of the economy, detecting hidden dependencies, regularizing models. In this regard, in our opinion, LASSO has advantages over OLS.

Our models, with the exception of the stock market models, include, with negative weights, youth unemployment. This predictor has a dominant economic character.

All of our models, without exception, have proven to be immune to the influence of the mining industry. Its inclusion in the model led to a deterioration in overall results, and the predictor itself turned out to be statistically insignificant. It turns out that the industry works only for itself.

An increase in the money supply negatively affects the performance of certain sectors of the economy. The increase in the money supply has a positive effect only on the capitalization of the stock market. This partially confirms the opinion that pumping money into the economy only leads to the emergence of financial bubbles. More general research is needed to draw more general conclusions. Preliminarily, it can be assumed that the constant assistance of the state to the banking sector, by increasing liquidity, is ill-founded. At the same time, household deposits are good indicators for the growth of industries. Such results require adjustments to the existing economic policy of the state. Moving on to particular conclusions, we note the following.

The KASE index model turned out to be more useful in terms of predictive value. It has the smallest standard error of all models and includes the largest number of predictors. The stock market index turned out to be sensitive to a wide range of social and macroeconomic indicators: population growth, unemployment, inflation, investment, devaluation. Our opinion: the development of the stock market does not require any specific financial measures; it is necessary to deal with the economy as a whole. The volume model of trading in corporate securities turned out to be the least represented both in predictability and in interpretability.

From this perspective, the LASSO method performs well as a trigger; the absence of positive results contributes to the generation and testing of new hypotheses of an intra-systemic nature. This suggests that the causes are subjective, internal and / or not included in our many predictors. In our study, these may be the prevalence of repo transactions in the market, real low (negative) profitability of the stock and bond sector, etc.

The analysis of modeling patterns in civil aviation showed the stability of the standard error of the model with the growth of a hyperparameter (lambda). The patterns of the stock market model, on the contrary, indicate its sensitivity to its changes. We believe that in this case, we can make a generalization, and talk about greater stability of the real sectors of the economy in relation to the financial one.

This assumption is supported by the small variability of the composition of predictors when modeling various aspects of the real sector.

LASSO modeling in civil aviation has revealed stable links between the industry and per capita income in various forms, the development of individual industries, and the money supply. Passenger and passenger transportation models are representative of both predictability and interpretability.

Transportation of goods in the civil aviation of Kazakhstan does not depend on most external social, economic and financial predictors. Rather, it is subjective in nature, due to internal problems in the development of the industry. The task of the aviation community is to establish and eliminate the causes of such an imbalance in development.

Acknowledgments

The article was prepared as part of grant projects of the Ministry of Education and Science of the Republic of Kazakhstan:

- "Positioning, genesis and optimization of the stock market of the Republic of Kazakhstan in the context of integration and globalization" IRN AP05135054
- "Competitiveness and stress resistance of the civil aviation of the Republic of Kazakhstan" IRN AP05136068

References

1. Ali MA, Amir N. Stock Market Development and Economic Growth: Evidence from India, Pakistan, China, Malaysia and Singapore. International Journal of Economics Finance and Management Sciences. 2014; 2: 220-226
2. Assylbekov A.P., Assylbekova B.S. Investment Risks of Frontier and Emerging Stock Markets // Central Asian Economic Rewiew. - 2018. -. No. 5-6 (123). - S. 117-129

3. Beck, T., & Levine, R. (2004). Stock markets, banks, and growth: Panel evidence. *Journal of Banking & Finance*, *28*(3), 423-442.
4. Chakrabarti, S., Ester, M., Fayyad, U., Gehrke, J., Han, J., Morishita, S., ... & Wang, W. (2006). Data mining curriculum: A proposal (Version 1.0). Intensive Working Group of ACM SIGKDD Curriculum Committee, 140.
5. Dobruszkes, F., Lennert, M., & Van Hamme, G. (2011). An Analysis of the Determinants of Air Traffic Volume for European Metropolitan Areas. Journal of Transport Geography,19(4),755-762. http://dx.doi.org/10.1016/j.jtrangeo.2010.09.003
6. Demirgüç-Kunt, A., & Maksimovic, V. (1996). Stock market development and financing choices of firms. *The World Bank Economic Review*, *10*(2), 341-369.
7. Ding, X., Zhang, Y., Liu, T., &Duan, J. (2015, June). Deep learning for event-driven stock prediction. In Twenty-fourth international joint conference on artificial intelligence, 2327-2333
8. Fischer T., Krauss C. Deep learning with long short-term memory networks for financial market predictions //European Journal of Operational Research. – 2018. – T. 270. – №. 2. – C. 654-669.
9. Garcia, V.F., & Liu, L. (1999). Macroeconomic determinants of stock market development. Journal of Applied Economics, 11(1), 29–59.
10. Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Zhou, Y. (2015). Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, *64*, 59-63.
11. Greenwood, J., & Smith, B. D. (1997). Financial markets in development, and the development of financial markets. *Journal of Economic dynamics and control*, *21*(1), 145-181.
12. Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer, 27(2), 83-85.
13. Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: deep portfolios. Applied Stochastic Models in Business and Industry, 33(1), 3-12.

14. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
15. Kalayci, S., & Yazici, S. (2016). The Impact of Export Volume and GDP on USA's Civil Aviation in between 1980-2012. International Journal of Economics and Finance, 8(1), 229-235.
16. Kuttner, K. N., & Mosser, P. C. (2002). The monetary transmission mechanism: some answers and further questions. Economic Policy Review, 8(1).
17. Lee, K. S., & Werner, R. A. (2018). Reconsidering monetary policy: An empirical examination of the relationship between interest rates and nominal GDP growth in the US, UK, Germany and Japan. Ecological Economics, 146, 26-34.
18. Levine, R. (1991). Stock markets, growth, and tax policy. *The Journal of Finance*, *46*(4), 1445-1465.
19. Masoud, N., & Hardaker, G. (2012). The impact of financial development on economic growth: Empirical analysis of emerging market countries. Studies in Economics and Finance, 29(3), 148–173.
20. Naceur, S.B., & Ghazouani, S. (2007). Stock markets, banks, and economic growth: Empirical evidence from the MENA region. Research in International Business and Finance, 21(2), 297–315.
21. Odhiambo, N.(2010). Stock market development and economic growth in South Africa: An ARDL-bounds testing approach. A paper presented at the Annual American Business Research Conference, Las Vegas, Nevada, USA, 8
22. Schumpeter, J. A. (1935). A theorist's comment on the current business cycle. *Journal of the American Statistical Association*, *30*(189), 167-168.
23. Singh, Dharmendra (2010), Causal Relationship Between Macro-Economic Variables and Stock Market: A Case Study for India, Pakistan Journal of Social Sciences, Vol. 3(2), pp. 263-274.
24. Tripathi, V. and Kumar, A., 2014. Relationship between Inflation and stock returns-evidence from BRICS markets using panel cointegration test, International Journal of Accounting and financial reporting, 4 (2), pp. 647-658.