Analyze Exporting Goods and Services of Bangladesh Using Data Mining Technique

Md.kamruzzaman

Pabna university of science and technology, pabna
kanakcsepust07@gmail.com

Abstract. In this paper we investigated the impact of exports goods and services on the economic growth of Bangladesh using the data of the period 1989-2012. Despite structural limitations in the Bangladesh economy, the export sector performed well throughout the 1990s. In this paper, linear regression model is used to forecast future export goods and services trends of Bangladesh. Then the linear regression model is trained on this dataset. The results show that there is a bidirectional long run relationship between exports goods and services and its reasoning factor (Age dependency ratio, Arable land, Export value index, internal debt stocks, Merchandise imports, industry value added and Official exchange rate). These results provide evidence that growth of exports goods and services in Bangladesh is increased by developing those independent reasoning factor.

1 Introduction

Export of Goods and Services is one of the most important macro-economic indicator of Bangladesh. It is an integral part of the total developmental effort and national growth of all economies including Bangladesh.It has been theoretically argued that export play a crucial role in economic development. A South Asian nation physically located near economic powerhouses India and China, the People's Republic of Bangladesh shipped $42.2 billion worth of goods around the globe in 2018[1]. Bangladesh is world's second-biggest apparel exporter after China. Garments including knit wear and hosiery account for 80% of exports revenue; others include: jute goods, home textile, footwear and frozen shrimps and fish, merchandise, Knit or crochet clothing, freight, insurance, transport, travel, royalties, license fees, and other such as business, personal, and government services. Despite many difficulties faced by the sector over the past years, it continued to show robust performance, Competitive strength. Exclusive Export Processing Zones (EPZ) are established to attract foreign direct investment and export promotion.

The export earnings also continuously increased over the years. A healthy performance of Bangladesh's export compared to the world and the SAARC countries is to be noted. The contribution of the export sector to Bangladesh's GDP has been gradually increasing over the years. According to the latest data of Bangladesh Bureau of Statistics (BBS), the contribution of merchandise exports to the country's GDP came down to 13.37% in fiscal year 2017-18 from 13.95% in FY 2016 -17.In 2014-15, the contributing share of exports to the economy was 15.98%, which slid to 15.47%.On the other hand, Bangladesh registered a 7.81% GDP growth in FY18, which was 7.28% the previous year. In FY18, Bangladesh's GDP was Tk22504793 million at current market prices, while overall export earnings from goods was Tk301045675 million [2].

The aim of our study is to find out the factor that is responsible for exporting goods and services from Bangladesh. We also try to find out how to develop those factor so that exportation from Bangladesh will be increased. To find out a smooth, hassle free and effective exportation process. Here, we tried our best to utilize every part of our knowledge gained through learning and practical works done under different subjects, relevant books, research-studies, articles, journals, and websites regarding this matter. We have tried our best to find all possible factor regarding to the exportation process and their viability, so that reader can get a clear picture of exportation relating factors of goods and services and can take decision which is best for their context.

2 Literature Review

Haydory Akbar Ahmed, Md. Gazi Salah Uddin (2015) describe Export, Imports, Remittance and Growth in Bangladesh. Annual data on Real GDP, exports, imports, implicit GDP deflator and remittance from 1976 to 2005 are used for this paper. Real GDP, export, import and implicit GDP deflator (base year 1990) data are collected from UN Statistical Division website3. The first step is to determine whether the variables used are stationary or not. If they are non-stationary in that case the issue is- to what degree they are integrated. This can be addressed by the Augmented Dickey–Fuller (ADF) tests. If all the variables in a multivariate model are integrated. Then the next step is to find out whether they are co-integrated or not using Johansen's framework [3].
Sayef Bakari, Mohamed Mabrouki investigated the Impact of exports and imports on economic growth new evidence from panama .This paper investigates the relationship between exports, imports, and economic growth in Panama. In order to achieve this purpose, annual data for the periods between 1980 and 2015 were tested using the Johansen co-integration analysis of Vector Auto Regression Model and the Granger-Causality tests The data set entails of observation for GDP (current US$), exports of goods and services (current US$), and imports of goods and services (current US$). All data set have brought from World Development Indicators 2016[4].

Afaf Abdull J. Saaed and Majeed Ali Hussain (2015) examined Impact of Exports and Imports on Economic Growth: Evidence from Tunisia. The study used Granger Causality and Johansen Co-integration approach for long run relationship Using Augmented Dickey-Fuller (ADF) and Phillip-Perron (PP) stationarity test, the variable proved to be integrated of the order one at first difference. Johansen and Juselius Co-integration test was used to determine the presence or otherwise of a co-integrating vector in the variables Growth .The data set consists of observation for GDP, exports of goods and services (current US$), and imports of goods and services (current US$). All data set are taken from World Development Indicators 2014[5].

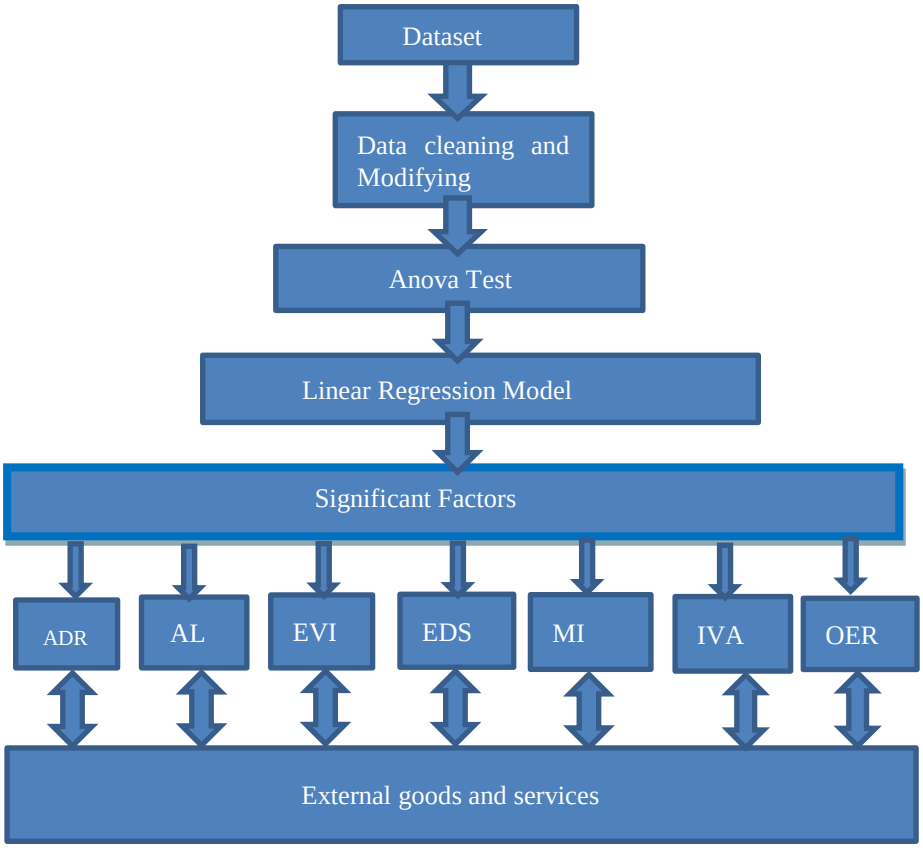
Mohammad Mafizur Rahman describes The Foreign Trade of Bangladesh: Its Composition, Performance, Trend, and Policy. This paper describes some issues which are very much related to exporting such as Exports Performance Compared to Imports, Composition and Performance of Imports of Bangladesh, Comparative Performance of

Bangladesh's Export and Import Sectors, Region-wise Exports of Bangladesh, Directions of Bangladesh's Exports and Imports, Export Policy and Reform Programme, Tariff Rationalization etc[6].

Md. Tareq Ferdous Khan and Nobinkhor Kundu (2012) describes Future Contribution of Export and Import to GDP in Bangladesh. The time series analysis by Box-Jenkins approach based on data from 1981 to 2010 indicates that the growth of the indicators will continue in the future. The autoregressive integrated moving average (ARIMA) which is very popular and also widely known as Box-Jenkins methodology (Box and Jenkins, 1976) is adopted to forecast the future exportation from Bangladesh [7].

3 Proposed system structure

This paper proposed system structure is a summary that details an outline of work. It identifies a problem and clearly states all the questions that will be researched as well as describes the resources and materials that one need. There are considered several steps to build a Regression model that analyse export goods and services and its responsible factors. First, we have collected data that pre-process and extract some features for analysing further manipulation of it. Then we used linear regression algorithms to classify them visualize predicted model with the appropriate figure. The proposed system describes exportation of goods and services from Bangladesh. The system describes how the correlating factor is significant with the outcomes.



ADR=Age dependency ratio, AL= Arable land, EVI=Export value index,

EDS=External debt stock, MI=merchandise imports, IVA=Industry added,

OER=Official exchange rate.

3.1 Dataset

The dataset use for this research is collected from the world development indicators 2015(WDI). The data set out in this paper is on annual data referred to the period 1989-2012. Data collection is a very important step of building prediction model because it is the base of all our predictions, a minor error in the data cause a blunder in the prediction.

3.2 Data cleaning and modifying

Data preprocessing is the primary task to prepare data of export goods and services for further analysis and getting good results. Data quality is explained in terms of accuracy, consistency, completeness, believability, interpretability and timeliness. These qualities are assessed by the usage of the data. In this study, we removed several tuples from which have multiple unclear, duplicate and missing values from existing data.

3.3 Anova test

Anova test is sometimes called as significance test. In order to find out the significant properties of the variables under consideration (Exports Goods and Services) we carry out Anova test. The purpose of anova test is to find out how the input factors affect the outcome factors.

3.4 Model Selection

In order to quantitatively forecast the status of export goods and services, different data mining techniques can be used. The associated task for the dataset used in this paper is linear regression. Therefore, liner regression model is used to forecast the status of export goods and services. The linear regression model equation is given by,

$$y = \varepsilon + x_1\beta_1 + x_2\beta_2 + \ldots\ldots\ldots\ldots + x_n\beta_n \qquad (1)$$

It predicts a variable y (target variable) as a liner function of another variable X (input variable/features). ε is the intercept and $\beta_1, \beta_2, \beta_3 \ldots\ldots \beta_n$ is the coefficient of $x_1, x_2, x_3, \ldots\ldots\ldots\ldots x_n$..

3.5 Model Validation

K fold cross validation model is used in this proposed model. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. This approach

involves randomly dividing the set of observations into k groups, or folds, of approximately equal size. Here a specific value of 4 k is chosen that is why the model is called as 4-fold cross-validation. The first fold is treated as a validation set, and the method is fit on the remaining k − 1 folds. We also calculate mean absolute error (MAE), root mean squared error (RMSE) and relative absolute error (RAE). The equation are given below

$$MAE = \frac{\sum(|y_a - y_p|)}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum(|y_a - y_p|)^2}{n}} \quad (3)$$

$$RAE = \frac{\sum(|y_a - y_p|)}{\sum(|y_a - \bar{a}|)} \quad (4)$$

4 Implementation

4.1 Data preparation

The world development indicators published data in the year between1960-2015.The first step of implementation was to collect data and preprocess that data. Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. The world development indicator published data on 1350 different factors. There is some repeated indicators in that data set such as same thing calculated in both $US and LCU. We deleted one of that repeated factor and after doing this there is 680 indicators remaining on that dataset. There is a lots of indicators which has missing values. we remove those missing values indicators and after deleting we get a dataset of 80 indicators. After cleaning datasets get the data of the year between 1989- 2012.

4.2 Identifying significant factors and building model

The second step of implementation is ANOVA test. The analysis of variance test is also known as the ANOVA test .An ANOVA test is a way to find out if survey or experiment results are significant individually. In other words, they help us to figure out if we need to reject the null hypothesis or accept the alternate hypothesis. Basically, we're testing groups to see if there's a significance between the output (Export goods and services) and their responsible factors. There is variability in the response variable. It

is the uncertainty that would be present if one had to predict individual responses without any other information. The best one could do is predict each observation to be equal to the sample mean. We perform ANOVA test on this 80 indicators. The ANOVA test result finds that 32 factors are individually significant with the outcome factor.
If all the indicators are significant individually then the next step is to find out whether they are co-integrated or not using linear regression model. Applying linear regression model is the third step of implementation. Linear regression is a basic and commonly used type of predictive analysis .These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. We apply linear model on this 32 factors. The result shows that 7 factors are jointly significant with the export goods and services and those factors are Age dependency ratio, Arable land, Export value index, External debt stock, Merchandise imports, Industry value added, And Official Exchange Rate. The linear regression model is simple and provides enough description of how the input affects the output.

The linear regression model function here is used:

$$EGS = -113.100 + (0.88820 \times ADR) - (0.04571 \times EVI) - (0.03147 \times EDS) + (0.0000000006187 \times MI) + (0.90520 \times IVA) + (0.44640 \times OER) \quad (5)$$

Here, EGS=Export goods and services

Table 1. Export goods and services indicators

Indicator	Coefficient	p-value
Age Dependency Ratio(ADR)	0.88820	0.00001740
Arable land(AL)	0.39030	0.00593000
Export value index(EVI)	-0.04571	0.01051000
External debt stocks(EDS)	-0.03147	0.00000168
Merchandise imports(MI)	0.0000000006187	0.00179000
Industry value added(IVA)	0.90520	0.00066000
Official exchange rate(OER)	0.44640	0.00000553

A histogram is an accurate representation of the distribution of numerical data. Histogram is used for continuous data. Histograms give a rough sense of the density of the underlying distribution of the data, Histogram of the dataset are shown below

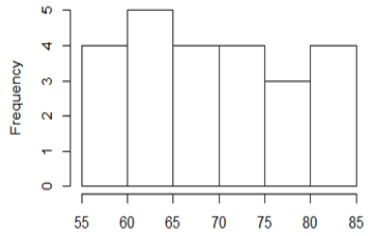


Figure 1: Age Dependency Ratio

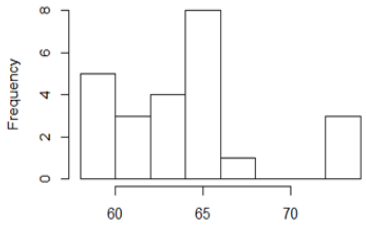


Figure 2: Arable land

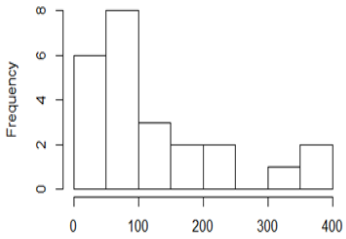


Figure 3: Export value index

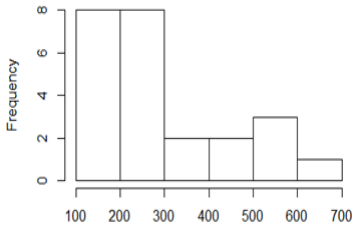


Figure 4: External debt stock

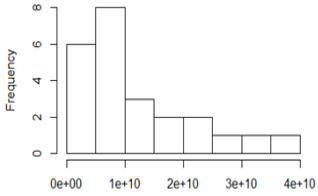


Figure 5; Merchandise imports

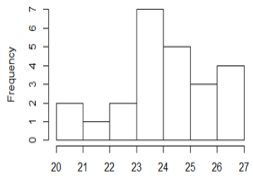


Figure 6: Industry value added

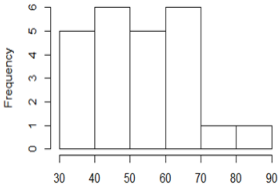


Figure 7: Official exchange rate

The final step of implementation is Cross validation. Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. Validation help us to evaluate the quality of the model. Validation help us select the model which will perform best on unseen data. Validation help us to avoid overfitting and under fitting .Here arrange the training examples in a random order and split it into 4 training sets. So 4 fold cross validation is used to validate the model.

The quantile-quantile (q-q) plot is a graphical technique for determining if two data sets come from populations with a common distribution. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the rest data set. Here the datasets are normally distributed, so the points in the QQ-normal plot lie on a straight diagonal line

.The data are from a distribution of the same type (up to scaling and location), a reasonably straight line is being observed.

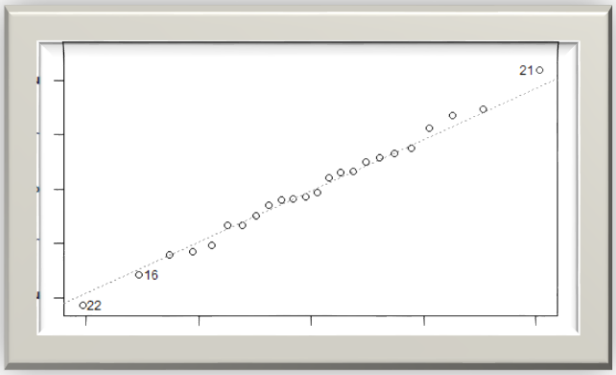


Figure 7: Normal Q-Q plot

If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data otherwise, a non-linear model is more appropriate. So from the data set diagram it is clearly said that the significant factor are linearly related to each other.

When conducting a residual analysis, a "**residuals versus fitted plot**" is the most frequently created plot. It is a scatter plot of residuals on the *y* axis and fitted values (estimated responses) on the *x* axis. The plot is used to detect non-linearity, unequal error variances, and outliers. The residual vs fitted graph below describes how the model is more linearly appropriate.

The residual vs fitted graph below describes how the model is more linearly appropriate.

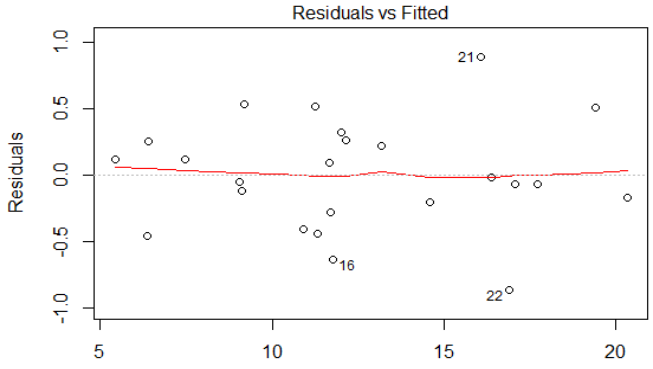


Figure 8: Residuals vs Fitted.

If we find equally spread residuals around a horizontal line without distinct patterns, that is a good indication we don't have non-linear relationships. So the Residuals vs fitted plot signifies that there is a linear relationship between the input factors and the outcome (Export goods and services).

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. Validation help us evaluate the quality of the model. Validation help us select the model which will perform best on unseen data. Validation help us to avoid overfitting and under fitting .Here arrange the training examples in a random order and split it into 4 training sets. 4-fold cross validation is used to validate the model.

Table 2: Testing and Training dataset

No.	Training set	Testing set
1	1995-2012	1989-1994
2	1989-1994,2001-2012	1995-2000
3	1989-2000,2007-2012	2001-2006
4	1989-2006	2007-2012

5 Result and Discussion

Cross-validation, a standard evaluation technique, is a systematic way of running repeated percentage splits. Divide our dataset into 4 pieces ("folds"), then hold out each piece in turn for testing and train on the remaining 3 together. This gives 4 evaluation results, which are averaged. In "k fold" cross-validation, when doing the initial division we ensure that each fold contains approximately the correct proportion of the class values. The 4 fold cross validation result are given below

Table 3: Result of error calculation of different model

No.	Model	MAE	RMSE	RAE
1	Base	0.315167	0.400965	0.094697
2	4 fold cross validation	0.842541	1.002526	0.674535
		0.760834	0.826956	0.957292
		0.62565	0.80976	0.410470
		1.370289	1.556234	0.982764
3	Best 1	0.354332	0.516433	0.105533
4	Best2	0.404121	0.516952	0.121770

From the result it is clear that the base model has less mean absolute error, root mean squared error and relative absolute error than any other cross validation model. We apply best two folds result on the base model and get a result which is given by Best 1 and Best 2. Best1 and Best2 model is slightly less accurate than the original model. So from the result we conclude that the Base model has less error compared to other model.

Now we test our model using different types of algorithms such as Random Forest, Random Tree, Gaussian processes and Decision Tree. This will help us to know the correlation coefficient within the outcome and input factors. It will also find out different types of error such as MAE, RMSE and RAE. The table shows the result below

Table 4: Model check using different algorithms

Rules/Functions/Algorithms	Correlation coefficient	MAE	RMSE	RAE
Random Forest	0.97260	0.7967	0.9756	0.232
Random Tree	0.9508	0.9991	1.2759	0.291
Gaussian Processes	0.9732	0.8111	1.0071	0.236
Decision Table	0.8670	1.7092	2.0931	0.498

We research our dataset and find some real evidences between the export goods and services and its inputs factors. Those things are discussed below. Age dependency ratio is the ratio of dependents people younger than 15 or older than 64. The data signify that a low dependency ratio in the exporting country can generate more output and hence export more. .Arable land includes defined by the FAO as land under temporary crops. A country exportation develops economically, the relative importance of agriculture declines land. Export values are the current value of exports converted to U.S. dollars. The income from exportation increased when the export value is high. External indebtedness affects a country's exportation income. To earn more from exportation the debt liabilities must be very small. The difference between the value of goods and services exported out of a country and the value of merchandise imports into the country are necessarily equal. So when merchandise imports increase from Bangladesh the export income also be increased. Industry value added is embedded as the input of a country's export. So exportation is much more significant when industry added more value to a country income. The exchange rate exerts a great influence on a country's exportation. The increment of exchange rate can increase the exportation income.

Results of Linear Regression tests suggest the existence of at least seven co-integrating relationships among the variables in the series at 5% level of significance. This implies that the series under consideration are driven by at least seven common trends. We save the Residual standard error (0.4871) which is the difference between a set of observed and predicted values. The residual standard error calculates how much the data points spread around the regression line. Which are used as the error-correction term in the subsequent tests for Linear Regression model. We also calculate multiple R-squared

(.9906) and Adjusted R-squared (0.9865) which is really fantastic. The higher the values the greater the model.

6 Conclusion

The aim of this study is to find the factor that are related to Exporting goods and service using the data of Bangladesh during the period 1989-2012.The Anova test and Linear Regression model are applied to investigate the relationship between the export goods and services and those factors that are responsible for exporting. Linear Regression test finds the co-integration relationship between the input responsible factors. The finding is clarified that export goods and services are found co-integrated with seven factor. So improving those factors earning from exports goods and services may will be increased. Cross validation model is used to measure the effectiveness of the model.

7 Referances

[1] Bangladesh's top 10 exports, http://www.worldstopexports.com/bangladeshs-top-10-exports/, 8 July 2019.

[2] Bangladesh to set export target of $39 billion for FY19, https://www.dhakatribune.com/business/2018/08/08/bangladesh-to-set-export-target-of-39-billion-for-fy19 , 8 July 2019.

[3] Haydory Akbar Ahmed1 Md. Gazi Salah Uddin (2015).Export, Imports, Remittance and Growth in Bangladesh.

[4] Sayef Bakari, Mohamed Mabrouki (2007). IMPACT OF EXPORTS AND IMPORTS ON ECONOMIC GROWTH: NEW EVIDENCE FROM PANAMA.

[5] Afaf Abdull J. Saaed and Majeed Ali Hussain (2015). Impact of Exports and Imports on Economic Growth: Evidence from Tunisia.

[6] Mohammad Mafizur Rahman (2008), the foreign trade of Bangladesh: its composition, performance, trend, and policy.

[7] Md. Tareq Ferdous Khan and Nobinkhor Kundu (2012).Future Contribution of Export and Import to GDP in Bangladesh: A Box-Jenkins Approach.