International Journal of Economics, Business and Management Research

Vol. 5, No.10; 2021

ISSN: 2456-7760

BILINEAR REGRESSION METHODS FOR MACROECONOMIC IMPACT ON CORRUPTION IN GHANA

Solomon Crusader Abo[1], Paul Hammond[2]* and Edmond Amissah[3]

[1]Department of Planning and Statistics/EMIS, Berekum Municipal Education Office, Ghana.
Tel: +233205662274

[2]University of Energy and Natural Resources, Sunyani – Ghana
Tel: +233244847415

[3]University of Energy and Natural Resources, Sunyani – Ghana
Tel: +233240476082

Abstract

This paper aimed to compare Ordinary Least Square (OLS), Principal Component Regression (PCR), and Partial Least Square Regression (PLSR) methods in handling multicollinearity among macroeconomic variables using real and simulated data. The real data on macroeconomic variables from The Global Economy was employed spinning 2001 to 2020. To assess the effect of a sample size of Root Mean Square Errors (RMSE), multivariate random normal data was simulated where there was the presence of multicollinearity among all explanatory variables with various samples, n = 20, 50, 100, 200, and 500. The simulations, as well as the analysis of the data, were implemented in R software. Comparing the performances of the three methods through Root Mean Square Error (RMSE), R2, and optimal value, the results of the real data indicated that the PLSR model performs better than the OLS and PCR models when there is the presence of multicollinearity in macroeconomic data. It was evident that only the unemployment rate and fiscal freedom influence the corruption perception index in Ghana. Again, based on the simulation results at n = 20, 50, 100, 200, and 500, PLSR still performs better than the PCR in both small and high samples. Finally, the results indicated that sample size affects the value of RMSE, as the higher the sample size used the lower the RMSE value from both PCR and PLSR. It was concluded that the PLSR method is more capable of overcoming multicollinearity problems in both real and simulated macroeconomic data. Again, we concluded that sample size affects the value of RMSE in macroeconomic data. It was recommended that PLSR is the best tool for determining macroeconomic variables that affect Corruption Perception Index (CPI). Also, the government of Ghana should formulate policies to drastically curb unemployment and again, to ensure that people pay various forms of taxes in Ghana. We also recommended that a large sample size must be employed in working with macroeconomic data, considering the PLSR method as efficient and effective in handling multicollinearity in such data.

Keywords: Bilinear Regression Methods, Macroeconomic Indicators, Corruption Perception Index, Multicollinearity, Root Mean Square Error, Principal Component Regression

1. Introduction

Macroeconomic indicators do exhibit high levels of multicollinearity. The problem of multicollinearity generally violates statistical modelling assumptions, thereby causing statistical results misinterpretations. To ensure the robustness of any analytical model for this kind of economic data, it is very crucial to eliminate this problem. This phenomenon has been dealt with severally in estimating regression parameters. To circumvent this problem, the ideal approach is to completely remove the variables that cause multicollinearity. However, the essential information on these variables will be lost. This then ushers many researchers into using statistical approaches that keep the correlated variables yet give significant results. Application of bilinear regression methods: mainly PCR and PLSR have been proven for handling this problem perfectly in several fields. However, there is no or limited comparison study of the two in handling multicollinearity among macroeconomic variables, especially where behavioural patterns are depicted. Again in the presence of multicollinearity, the effect of sample size on RMSE has not been identified for this kind of data. There is, therefore, the need to explore the impact of macroeconomic variables on corruption in Ghana. Considering this gap, this study established the influence of macroeconomic indicators on corruption by comparing the performance of PCR and PLSR for handling multicollinearity using real and simulated studies. This comparison is necessary because none of the two methods has provided a clear indication of an advantage over the other in macroeconomic studies.

There are several procedural comparison studies between PCR and PLSR. The two approaches have common features, with established comprehensive theoretical relationships in the literature. In both theoretical and simulation studies, it is established that PCR and PLSR do not show significant differences in the prediction errors, though the PLSR uses fewer latent variables than the PCR (Wentzell & Montoto 2003; Helland 1988; Bhandare et al. 1994). On the other hand, some similar studies have found counterview that PLSR seems to predict better than PCR (Dupuy et al., 1994). Jouan-Rimbaud et al. (1995) also found that the quality of the models for PCR and PLSR is similar, but the PCR model has a simpler capacity in selecting latent variables. This is an indication that the PLSR seems better for researchers in science. These studies have examined the relative merits of the two methods in predictive capacity, usually in science. However, no study has studied the two methods in economic settings where variables are always correlated. Hence this Paper illuminated the predictive performance of these techniques from a macroeconomic perspective, where real and simulation studies have been carried.

Corruption is the abuse of entrusted power for private gain (Jouan-Rimbaud et al., 1995). The categories of corruption include bribery, embezzlement, facilitation payment, fraud, collusion, extortion, patronage, clientelism, nepotism, and cronyism (Rohwer, 2009). The features of corruption are economic, religious, political, administrative, social, and cultural factors, both domestic and international. Corruption is the greatest impediment to economic growth, social development, and reduction of poverty (Rohwer, 2009). Ghana was ranked 81st out of 176 countries globally and 11th among 52 African countries on the Corruption Perception Index (CPI) with great effects of corruption on natural resources management, the judicial system, and the police service (Pring & Vrushi, 2019). On the economic cost due to corruption, mismanagement and corruption cause a yearly loss of $3 billion in Ghana (The Africa Report, 2019). Ghana alone loses about $2.3bn through corruption, with many citizens having to pay bribes before accessing basic public services (The Africa Report, 2019). It is also noted that in Africa, corruption hinders governance as its yearly loss is over $50 billion and $148 billion for illicit flows and corruption respectively (Pring & Vrushi, 2019). The alarming rise in corruption in Ghana has adversative effects on the development and growth of the country. The surge rate of unemployment in Ghana is a result of corrupt acts and actions such as the falsification of ages and documents. Pring and Vrushi (2019) state that in every four people in Africa, more than one pays a bribe for public service. It is evident from Figure A.3 that for the period 2001 to 2020, the corruption perception was remarkable between 2009 and 2016. Since 2017, the corruption perception exhibits a decreasing trend. This trend corroborates what other studies have established, however, there is a gap in identifying which sector of the macro-economy contributes to this alarming trend of corruption in Ghana.

Several studies have established that there is a relationship between corruption and macroeconomic variables. It is found that corruption and shadow economy are related and have a high negative relationship with economic growth (Borlea et al., 2017). There is the adverse effect of religion on corruption (Chang & Golden 2004; Bonaglia et al. 2001; Paldam 2001) while Shadabi (2013) argued that Islam and Christianity have no significant effect on corruption. Braun and Di Tella (2000) documented that the amount of corruption in a country is positively correlated with the levels of inflation and political right or competition. Globalization increases corruption (Das and DiRienzo 2009; Caiden et al. 2001. Alternatively, Akhter (2004) opposed that globalization reduces corruption levels. The levels of corruption are positively correlated with the country's population size (Amin & Soh, 2019), fiscal freedom or burden (Achim & Borlea 2020; Ivanyna et al. 2010; Fjeldstad 2003), unemployment (Adjor & Kebalo, 2018), and Gross Domestic Product (GDP) (De Rosa et al., 2010). In contrast, Jiang and Nie (2014) found the existence of high GDP growth amidst the prevalence of government corruption in China. Again, other studies highlight the positive impacts of corruption on sustainable development (Zaman & Goschin 2015; Sahakyan & Stiegert 2012).

The remainder of this study is arranged as follows. Section 2 defines the notation and the methods under consideration. In section 3, the results and discussions of the results are shown in connection with the real data and simulation studies. Sections 4 and 5 present the summary and conclusion as well as the recommendations for policy implications respectively.

2. Methodology

2.1. Exploratory Data Analysis

The exploratory analysis was conducted to detect the multicollinearity and Model Adequacy for the proposed methods. GG plot was used for correlation analysis, VIF analysis for multicollinearity detection as well as KMO and Bartlett test for model adequacy.

2.2 Ordinary Least Squares (OLS)

Multiple regression model studies the association between one dependent variable and two or more independent variables. Consider a dataset of n observations, where each observation has a scalar response, $y_{i,}$ and a vector of p predictors, X_{ij} for $i = 1, 2, \ldots, p$. The Multiple Linear Regression (MLR) model is given by;

$$y = \beta X + \varepsilon \tag{1}$$

Where y is the response vector
β is the regression coefficients
X is the matrix of predictor variables
ε is the residual error's vector

The estimation of β coefficients in (1) requires the use of the OLS approach. OLS gives unbiased estimates of β when the X matrix has a full rank. Thus, all the predictor variables are retained but do not exceed the number of observations. The usual OLS criterion minimizes the sum of squared distances between the observed responses and the fitted responses from the regression model. OLS assumes that all the explanatory variables are independent and randomly sampled. Thus, there is no collinearity between them. Phatak and De Jong (1997) obtained the vector of fitted values of $\hat{\beta}$ in (1) by minimizing the sum of squared residuals (SSR).

$$SSR = \sum_{i=1}^{n} (y - \beta X)^2 \tag{2}$$

Taking partial derivative of SSR with respect to β_i gives;

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'y \tag{3}$$

$$\hat{y}_{OLS} = X\hat{\beta}_{OLS} = X(X'X)^{-1}X'y = H_x y \tag{4}$$

2.3 Bilinear Regression Methods

Macroeconomic variables are highly correlated and so the estimated coefficients in OLS will be magnified, thereby resulting in wrong regression estimates. For this reason, bilinear regression methods: PCR and PLSR are compared for performance analysis.

2.3.1 Principal Component Regression

PCR eliminates the dimensions of the X-space that poses the multicollinearity problem. This helps in finding the number of principal components of maximum variation in X, which has optimum predictive power for the model.

Denote $V = [V_1, \ldots, V_p]$ and $\lambda_1, \ldots, \lambda_p$ be the matrix of size $p \times p$ of columns that are the normalized eigenvectors and the corresponding eigenvalues respectively. Again, let $Z = [Z_1, \ldots, Z_p] = XV$, then the *ith* principal component of X is $Z_i = XV_i$. The PCR model is;

$$Y = \beta X + \varepsilon = XVV'\beta + \varepsilon = Z\gamma, \qquad \gamma = V'\beta \tag{5}$$

The least square estimator of γ is;

$$\hat{\gamma} = (Z'Z)^{-1} Z'Y = \Lambda^{-1} Z'Y \tag{6}$$

This gives the principal component estimator as (Coxe, 1984; Jackson, 1991; Jolliffe, 2002);

$$\hat{\beta}_{PCR} = V_{\hat{\gamma}} = (Z'Z)^{-1} V \Lambda^{-1} Z'Y \tag{7}$$

2.3.2 Partial Least Squares Regression

Partial least squares regression is a kind of multiple regression that is used to describe the relationship between single or multiple response variables and predictors through the use of latent variables. Wold et al. (2001) posit that PLSR is most suitable for handling datasets with strong collinearity, noisy, multiple predictors, and fewer observations than the predictors. In this sense, it shows that an increase in response variable as a change in given predictor variable when all other predictors are held constant.

From Nonlinear Iterative Partial Least Squares,

$$Y = UQ^T = TBQ^T \tag{8}$$

Where $U = (u_1, \ldots, u_p)$ is the latent scores of Y

$T = (t_1, \ldots, t_p)$ is the latent variable scores of X

Q is the corresponding loadings
B is a matrix that has the ith diagonal element as b_i
The model (8) reduces to (11) if;

$$T = XW(P^T W)^{-1} \tag{9}$$

$$Q^T = B^{-1}(T^T T)^{-1} TY \tag{10}$$

$$Y = XW(P^T W)^{-1}(T^T T)^{-1} T^T Y \tag{11}$$

$$\hat{\beta}_{PLS} = W(P^T W)^{-1}(T^T T)^{-1} T^T Y \tag{12}$$

Upon obtaining the PLSR estimates, the contributions of X to Y is measured from the Sum of Squares Decomposition (SS) of Y. Consider n-vector x and $n \times k$ matrix X,

$$SS_x = x^T x = \sum_{i=1}^{n} x_i^2 \tag{13}$$

$$SS_X = \sum_{j=1}^{n} x_j^2 \tag{14}$$

The decomposition of the total sum of squares has two SS;

$$SST = SS(Y) = SSR + SSE \tag{15}$$

SSR is the summation of SS of latent variables;

$$SSR = \sum_{i=1}^{p} SS(b_i t_i q_i^T) = \sum_{i=1}^{p} b_i^2 SS(t_i) = \sum_{i=1}^{p} SSR_i \tag{16}$$

2.4 Measure of Model Performance

To assess the performance comparison of the models, the Root Mean Square Error (RMSE) of regression coefficients and their associated Coefficient of Multiple Determination (R^2) are measured. RMSE is a measure of the differences between the observed values and the predicted values of a model or an estimator. It tells how close the data points are to the

ISSN: 2456-7760

regression line. The R^2 on also measures the proportion of variation in the response variable that is predicted from the set of predictors in a given multiple regression model. R^2 lies in the interval $0 \le R^2 \le 1$. In a well-fitted model, R^2 value close to 1 indicates that the model explains a large portion of the variance in the outcome variable. The model that performs better is indicated by the lower values of RMSE and number of components as well as higher R^2 estimates.

In terms of Sum of Squares, let n, y, $\hat{y}$ $\bar{y}$ and denote the total number of observations, observed value, predicted value, and the mean value of the response variable respectively. Then,

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (17)$$

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \quad (18)$$

where SSR denotes the Sum Squares due to regression and SST is the Total Sum of Squares.

2.4.1. Variable Importance in the Projection (VIP)

VIP scores estimate the significance of each variable in the projection used in a PLSR model and are frequently used for variable selection. The VIP scores and the PLSR estimate select the most influential predictors (Chong & Jun, 2005). The threshold for VIP value is 0.8 (Eriksson et al., 2006). Thus, a variable with a VIP Score close to or greater than 1 is regarded as being significant in a given model, while variables with VIP scores less than 1 are insignificant in the model. The VIP score for ith predictor is estimated by (19)

$$VIP_i = \sqrt{k \frac{\sum_{i=1}^{n} W_{ia}^2 b_a^2 t_a^T t_i}{\sum_{i=1}^{n} b_i^2 t_i^T t_i}} \quad (19)$$

Considering the Nonlinear Iterative Partial Least Squares algorithm, W_{ia} is a weight of the ith predictor to the ath latent variable. It is evident from equation (19) that if all predictor variables contribute equally to the model, the total expected VIP score should not exceed 1, since the Sum of Squares (SS) of all VIP values must be equal to the number of predictors.

3. Results and Analysis

3.1 Description of Data and Variables

To assess the impact of macroeconomic indicators on corruption in Ghana, the secondary data on the Ghanaian economy was obtained from TheGlobalEconomy.com spinning the period 2001 to 2020. For this period, corruption perception index, unemployment rate, political rights, Gross Domestic Product, and Fiscal Freedom were up to date. The missing observations for Globalization, Inflation Rate, Population Size, Economic Growth, Shadow Economy, proportions of Christian and Muslim compositions were imputed using mice package in R. With the aid of the imputation, a complete dataset consisting of 20 observations and 12 variables was attained. Table 1 shows the descriptions of the individual variables. Both the imputation and the analysis of data were implemented in RStudio.

Table 1: Variable Description

Variable Name	**Variable Type**	**Description**
Corruption Perceptions Index (1-100)	Dependent (Y)	The Corruption Perceptions Index is an indicator of perceptions of public sector corruption, i.e. administrative and political corruption. The indicator values are determined by using information from surveys and assessments of corruption, collected by a variety of reputable institutions.
Unemployment rate	Independent (X_1)	Unemployment rate refers to the share of the labour force that is without work but available for and seeking employment.
Globalization Index (0-100)	Independent (X_2)	This is the overall index of globalization on the economic, social, and political dimensions of globalization. High values denote the greater presence of globalization.
Inflation Rate	Independent (X_3)	Inflation Rate is explained as the overall economic condition being the level of variation in prices for the rate of commodities and services.
Population Size	Independent (X_4)	Population Size is the total number of people in Ghana.
Political Rights (1-7)	Independent (X_5)	The Political Rights ratings from the Freedom House evaluate three categories: electoral process, political pluralism and participation, and the functioning of government. The index ranges from 1 (strong rights) to 7 (weak rights).
Economic Growth	Independent (X_6)	The annual percentage growth rate of GDP at market prices is based on constant local currency. It is calculated without making deductions for the depreciation of fabricated assets or depletion and degradation of natural resources.
Gross Domestic Product (GDP)	Independent (X_7)	The Gross Domestic Product (GDP) is the sum of the gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. The data are in billion local currency units using current prices.
Shadow Economy	Independent (X_8)	Shadow Economy is defined as the transactions and economic activities that are illegal and taxes are not been paid.
Fiscal Freedom (0-100)	Independent (X_9)	The Fiscal freedom index measures the tax burden imposed by the government. It is composed of three quantitative factors: the top marginal tax rate on individual income, the top marginal tax rate on corporate income, and the total tax burden as a percentage of GDP.
Christianity	Independent (X_{10})	Christians as a percentage of the total population.
Islamic	Independent (X_{11})	Muslims as a percent of the total population.

Vol. 5, No.10; 2021

ISSN: 2456-7760

3.2. Descriptive Analysis

Table 7 (Appendix) displays the summary statistics of the variables. According to Table 2, the average corruption perception index in Ghana is 39.75% with median and standard deviation values of 39% and 4.44 respectively, while the minimum value is 33 and the maximum value is 48. The average unemployment rate for Ghana during this period was 5.86 percentage points with a minimum of 4.16 points and a maximum of 9.28 points. For the globalization index, the average value of 56.31% was attained with a minimum and a maximum index of 52.69% and 61.46% respectively.

The average rate of inflation for the period is 14.80%, while the respective minimum and maximum inflation rates are 7.10% and 32.90%. The percentage point of 24.50 (with 19.76 minimum and 29.77 maximum) is an indication of the average annual population size. For the political right index, it was realized that the mean value is 1.25 with minimum and maximum values as 1 and 2 respectively. On average, economic growth of 5.99 million US dollars with a standard deviation of 2.83 is recorded annually for the period. Similarly, the sum of the gross value added by all residents within a year for the period is 33.47 with a standard deviation of 22.06, where the minimum GDP is 5.31 and the maximum is 66. Again, it can be seen from Table 7 that an average percentage point of 40.91 with a standard deviation of 2.62 was recorded for shadow economy within the said period. Table7 again depicts that as high as 80.90 (SD=4.33, Min=73, Max=29.77) percentage point of fiscal freedom was enjoyed in Ghana from 2001 to 2020. It is also evident that during the period under review, the majority (63.76%) of the Ghanaian Populace are Christians while 17.65% were Muslims.

3.3. Detection of Multicollinearity and Model Adequacy

In this study, Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett Test of Sphericity were used to test the hypothesis that the dataset is sufficient for the performance comparison of regularized multiple regression methods.

Hypothesis

H_0: Correlation matrix is not an identity matrix

H_1: Correlation matrix is an identity matrix

Table 2 indicates that the Bartlett Test is significant (Chi-sq=232.01, df=55, P-Value<0.01). Therefore, there is enough statistical evidence to reject H0 and conclude that the correlation matrix is an identity matrix. This is an indication that the assumption of variable independence was met. Hence, regularized multiple regression methods are suitable for the analysis. Again, though the Measure of Sampling Adequacy (MSA) for the political right index was below the threshold for KMO statistic (0.50), the overall MSA according to Table 3 is 0.79. Thus, 79% of the variations in the corruption perception index in Ghana are explained by the macroeconomic indicators understudy, and hence, the sample was adequately enough for the study. Table 2 shows that Population Size (X4), GDP (X7), Shadow Economy (X8), Fiscal Freedom (X9), Christianity (X10) and Islamic (X11) have VIF estimates greater than 10. This indicates that these variables are highly correlated, hence the presence of multicollinearity. This assertion is further affirmed by the findings in Figure A.2. There is an indication of high correlations (both positive and negative) between the response and predictors as well as among the predictors. This,

ISSN: 2456-7760

therefore, corroborate the application of the Bilinear Regression Methods rather than OLS regression.

Table 2: Estimates for KMO and Bartlett Test

Indicator	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
VIF	8.92	9.61	4.32	55.91	8.17	3.00	37.17	18.15	11.31	23.27	10.42
MSA	0.84	0.70	0.82	0.77	0.82	0.48	0.73	0.88	0.91	0.79	0.75
Overall MSA	0.79										
Bartlett Test	Chi-square	df	P-value								
Estimate	232.01	55	<0.01								

3.4. Performance Comparison of Bilinear Regression Methods

The validation plot indicates that PCR and PLSR plots stabilize at 5 and 3 components respectively. Thus, after this point, no more changes occur in the plots. It is clear-cut from Table 3 that the optimal value of components identified in the OLS, PCR, and PLSR models are 11, 5, and 3. This means that the PLSR explains 99% of the variations in the outcome variable (corruption perception index) with the least RMSE value of 0.45. This, therefore, shows that the PLSR model performs better than the PCR model when there is a presence of multicollinearity in macroeconomic data, hence the PLSR is better for explaining our data.

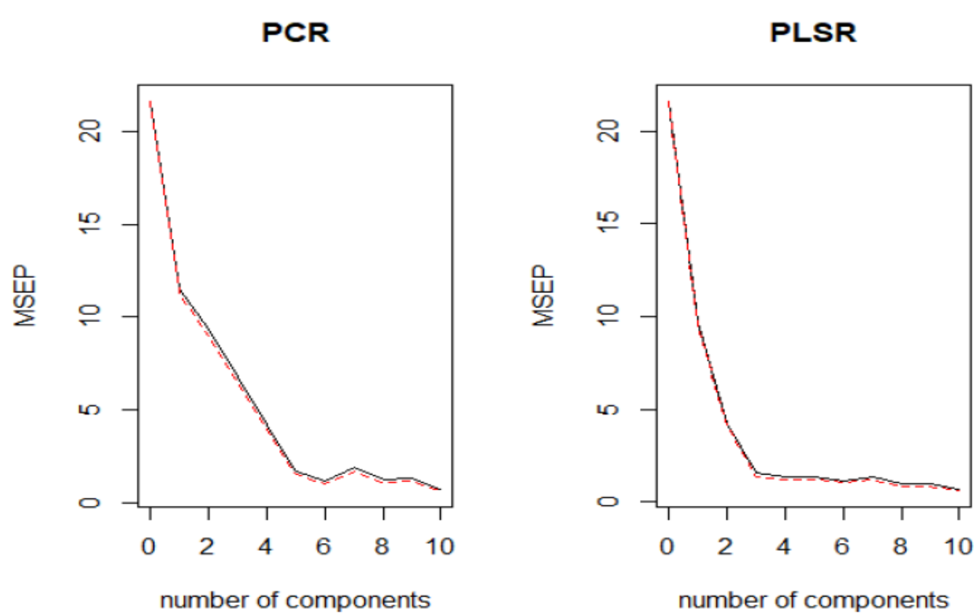

Figure 1: Real Data Validation Plot

Table 3: Model Performance Metrics

Model	RMSE	R2	No. of Components
OLS	1.44	0.89	11
PCR	0.75	0.98	5
PLSR	0.45	0.99	3

3.5. Percentage of Variance Explained

The Kaiser (or Latent Root) Criterion places the number of factors to retain to be eigenvalues greater than 1. Thus, 1 is the amount of variance accounted for by a single item ($r^2 = 1$). This means that factors that account for less variance have eigenvalues less than 1. It is evident from Table 4 that only 2 among the 11 factors cumulatively explained 94.78% and 61.56% of the variations in the independent and the dependent variables respectively. This is an indication that only 2 variables have a significant impact on the corruption perception in Ghana.

Table 4: Variance Explained

Factor	Eigenvalue	Cumulative X Variance (%)	Variance Explained for X Effects (%)
1	7.12	90.73	50.87
2	1.61	94.78	61.56
3	0.70	97.86	72.28
4	0.62	98.46	84.71
5	0.38	99.42	87.90
6	0.19	99.48	88.54
7	0.17	99.86	88.61
8	0.13	99.96	88.80
9	0.04	99.97	88.93
10	0.03	99.99	88.97
11	0.01	100.00	88.97

3.6. Variable Importance in the Projection

Table 5 indicates the VIP scores, which measure the importance of each variable in the modelling of X and Y. It is evident from Table 5 that no variable seems significantly important, however, X_1 and X_9 have a high value of VIP scores, which are closer to the cut-off point of 0.8 (Wold et al., 1993) showing that they are more relevant to predict the response variable. Therefore, the significant variables that explain the high percentage of variation in corruption in Ghana are unemployment and fiscal freedom. This affirms that the corruption levels are positively correlated with fiscal freedom (Achim & Borlea 2020; Ivanyna et al. 2010; Fjeldstad, 2003) and unemployment (Adjor & Kebalo, 2018).

Vol. 5, No.10; 2021

ISSN: 2456-7760

Table 5: Variable Importance

Variable	**X_1**	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**	**X_8**	**X_9**	**X_{10}**	**X_{11}**
Score	0.26	0.03	0.08	0.05	0.04	0.09	0.10	0.03	0.12	0.02	0.02

3.7. Simulation Study

A correlated dataset of n sample sizes (n= 20, 50, 100,200, 500) was simulated using the R package (JWileymisc) where the correlation matrix was converted to covariance to be used as the standard deviation. For the sake of convenience, we held a fixed number of predictors (p=11) as in the real data. Considering multivariate random normal function in R with varying n, the mean and the standard deviation of the original dataset were used for simulating each different sample size. The performance comparison of OLS, PCR, and PLSR methods was assessed on the values of RMSE, R2, and the optimal number of components. The optimal number of components for PCR and PLSR was obtained from the validation plots. The estimated values for the Bilinear regression methods with different sample sizes were presented in Table 6.

Table 6: Estimates for Bilinear Regression Methods with Different Sample Sizes

n	**Estimator**	**Model**		
		OLS	**PCR**	**PLSR**
20	RMSE	1.44	0.35	0.11
	R2	0.89	0.98	0.99
	Optimal Value	11	5	3
50	RMSE	1.46	0.64	0.20
	R2	0.89	0.98	0.99
	Optimal Value	11	5	3
100	RMSE	1.47	0.58	0.18
	R2	0.89	0.98	0.99
	Optimal Value	11	5	3
500	RMSE	1.47	0.53	0.18
	R2	0.89	0.99	0.99
	Optimal Value	11	6	3
1000	RMSE	1.47	0.5	0.11
	R2	0.88	0.99	0.99
	Optimal Value	11	5	3

It is evident from Table 6 that OLS estimates for all sample sizes remain the same. RMSE for PCR increases from 0.35 to 0.50 and R2 increases from 0.89 to 0.99, while optimal value sought no significant change as the sample size increases. PLSR selects 3 as the optimal value for both lower and higher samples, with constant R2 and the optimal value of 0.99 and 3 respectively. This means that PLSR gives lower RMSE and optimal values with consistent predictive accuracy (0.99) than the PCR. Hence, PLSR has better performance than the PCR. This finding corroborates other studies in the field other than economics. Thus, PLSR uses fewer latent variables than the PCR (Wentzell and Montoto 2003; Helland 1988; Bhandare et al. 1994) and also predicts better than PCR (Dupuy et al., 1994).

4. Summary and Conclusion

The results of the real data indicate that the PLSR model performs better than the PCR model when there is a presence of multicollinearity in macroeconomic data. It was also evident that only the unemployment rate and fiscal freedom influence or cause the variations in the corruption perception index in Ghana. Again, based on the simulation results at n = 20, 50, 100, 200, and 500, PLSR still performs better than the PCR in both small and high samples. Finally, the results indicated that sample size affects the value of RMSE, as the higher the sample size used the lower the RMSE value from both PCR and PLSR. It is concluded that the PLSR method is more capable of overcoming multicollinearity problems in macroeconomic data than the PCR, though PCR does better than OLS. In totality, PLSR performs better in determining macroeconomic variables that affect the corruption perception index. Again, we concluded that sample size affects the value of RMSE in macroeconomic data.

5. Recommendations

The findings of this study are of major significance to policymakers regarding how to control corrupt deeds in Ghana. Considering our findings, it is apparent that failure or insufficient handling of unemployment and fiscal freedom will result in surge rates of corruption. This in turn affects GDP and shrouded economic growth in the country. Given this, the study recommends that the government should formulate policies to drastically curb unemployment and also, to ensure that people pay various forms of taxes in Ghana. This could be in the form of public education on the need to pay tax and strict measures to punish defaulters. Additionally, we recommend that a large sample size must be employed in working with macroeconomic data. Finally, the PLSR method must be used in handling multicollinearity in macroeconomic variables, where the corruption perception index is the dependent variable.

Acknowledgment

The authors are very thankful to the editors for their immense clarifications and recommendations that momentously enriched this publication.

Declaration of Interest Statement

We wish to confirm that this paper has no known conflicts of interest and that this research did not receive any significant financial support that could have swayed its findings. To the best of our knowledge, it contains no materials previously published by another author.

References

Achim, M. V. and Borlea, S. N. (2020). Economic and financial crime: Theoretical and methodological approaches. In *Economic and Financial Crime*, pages 1–71. Springer.

Adjor, D. M. and Kebalo, L. (2018). Does corruption matter for unemployment in SADC countries? *Review of Economic and Business Studies*, 11(2):65–92.

Akhter, S. H. (2004). Is globalization what it's cracked up to be? Economic freedom, corruption, and human development. *Journal of World Business*, 39(3):283–295.

Amin, M. and Soh, Y. C. (2019). *Corruption and Country Size: Evidence Using Firm-Level Survey Data*. The World Bank.

Bhandare, P., Mendelson, Y., Stohr, E., and Peura, R. A. (1994). Glucose determination in simulated blood serum solutions by Fourier transform infrared spectroscopy: investigation of spectral interferences. *Vibrational spectroscopy*, 6(3):363–378.

Bonaglia, F., Braga de Macedo, J., and Bussolo, M. (2001). How globalization improves governance. *Available at SSRN 288354*.

Borlea, S. N., Achim, M. V., and Miron, M. G. A. (2017). Corruption, shadow economy and economic growth: An empirical survey across the European union countries. *Studia Universitatis "Vasile Goldis" Arad–Economics Series*, 27(2):19–32.

Braun, M. and Di Tella, R. (2000). *Inflation and corruption*. Citeseer.

Caiden, G. E., Dwivedi, O. P., Jabbra, J. G., et al., et al. (2001). *Where corruption lives*. Kumarian Press Bloomfield, CT.

Chang, E. and Golden, M. (2004). Does corruption pay? the survival of politician charged with malfeasance in the postwar Italian chamber of deputies. *Unpublished paper, Michigan State University and the University of California at Los Angeles*.

Chong, I. G. and Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1-2):103–112.

Das, J. and DiRienzo, C. (2009). The nonlinear impact of globalization on corruption. *The International Journal of Business and Finance Research*, 3(2):33–46.

De Rosa, D., Gooroochurn, N., Görg, H., et al. (2010). Corruption and productivity firm-level evidence from the beeps survey. Technical report, Policy Research Working Paper.

Dupuy, N., Willems, A., Pot, B., Dewettinck, D., Vandenbruaene, I., Maestrojuan, G., Dreyfus, B., Kersters, K., Collins, M. D., and Gillis, M. (1994). Phenotypic and genotypic characterization of Bradyrhizobia Modulating the leguminous tree acacia Albida. *International Journal of Systematic and Evolutionary Microbiology*, 44(3):461–473.

Fjeldstad, O.-H. (2003). Fighting fiscal corruption: lessons from the Tanzania revenue authority. *Public Administration and Development: The International Journal of Management Research and Practice*, 23(2):165–175.

Helland, I. S. (1988). On the structure of partial least squares regression. *Communications in Statistics-Simulation and Computation*, 17(2):581–607.

Ivanyna, M., Moumouras, A., and Rangazas, P. (2010). The culture of corruption, tax evasion, and optimal tax policy. *Economic Inquiry*, 54:520–542.

Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology*, 74(8):2204–2214.

Jiang, T. and Nie, H. (2014). The stained china miracle: Corruption, regulation, and firm performance. *Economics Letters*, 123(3):366–369.

Jolliffe, I. T. (1986). Principal components in regression analysis. In *Principal Component Analysis*, pages 129–155. Springer.

Jouan-Rimbaud, D., Massart, D.-L., Leardi, R., and De Noord, O. E. (1995). Genetic algorithms as a tool for wavelength selection in multivariate calibration. *Analytical Chemistry*, 67(23):4295–4301.

Paldam, M. (2001). Corruption and religion adding to the economic model. *Kyklos*, 54(2-3):383–413.

Phatak, A. and De Jong, S. (1997). The geometry of partial least squares. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 11(4):311–338.

Pring, C. and Vrushi, J. (2019). Global corruption barometer: Africa 2019. *Transparency International*.

Rohwer, A. (2009). Measuring corruption: a comparison between the transparency international's corruption perceptions index and the world bank's worldwide governance indicators. *CESifo DICE Report*, 7(3):42–52.

Sahakyan, N. and Stiegert, K. W. (2012). Corruption and firm performance. *Eastern European Economics*, 50(6):5–27.

Shadabi, L. (2013). The impact of religion on corruption. *The Journal of Business Inquiry*, 12(1):102–117.

Tchie, A. and Mayne-Flood, S. (2019). Ghana: Corruption and neglect of the north threatens security. *the africa report*.

Wentzell, P. D. and Montoto, L. V. (2003). Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and intelligent laboratory systems*, 65(2):257–279.

Wold, S., Johansson, E., Cocchi, M., et al. (1993). PLS: partial least squares projections to latent structures.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

Zaman, G. and Goschin, Z. (2015). Shadow economy and economic growth in Romania. cons and pros. *Procedia Economics and Finance*, 22:80–87.

Vol. 5, No.10; 2021

ISSN: 2456-7760

A. Appendix

A.1. Summary Statistics

Table 7: Summary Statistics

Variable	Total	Mean	Median	SD	Minimum	Maximum
Corruption Perceptions Index (1-100)	20	39.75	39.00	4.44	33.00	48.00
Unemployment rate	20	5.86	5.57	1.37	4.16	9.28
Globalization Index (0-100)	20	56.31	55.85	2.58	52.69	61.46
Inflation Rate	20	14.80	13.70	6.12	7.10	32.90
Population Size	20	24.50	24.48	3.19	19.76	29.77
Political Rights (1-7)	20	1.25	1.00	0.44	1.00	2.00
Economic Growth	20	5.99	5.75	2.83	2.18	14.05
Gross Domestic Product (GDP)	20	33.47	30.35	22.06	5.31	66.00
Shadow Economy	20	40.91	40.81	1.54	38.50	43.16
Fiscal Freedom (0-100)	20	80.90	83.00	4.33	73.00	29.77
Christianity	20	63.76	64.30	1.42	61.00	65.70
Islamic	20	17.65	18.00	0.49	17.00	18.00

A.2. Correlation Plot

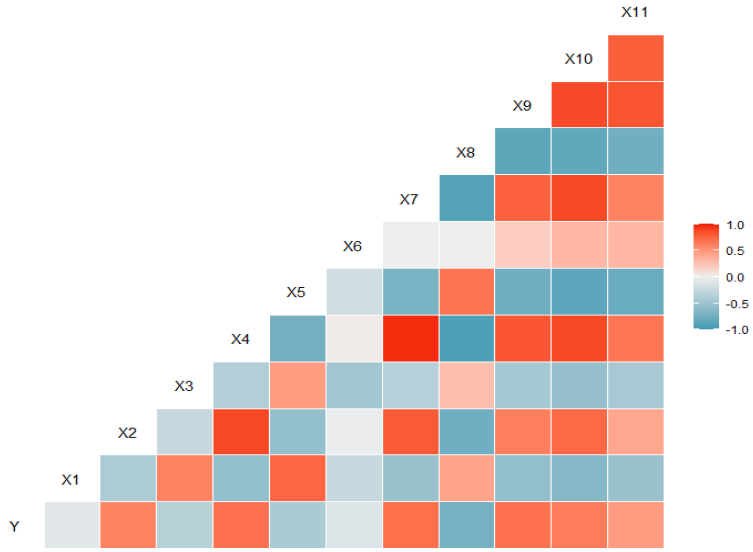

Figure 2: Correlation Plot

ISSN: 2456-7760

A.3. Trend of Corruption in Ghana

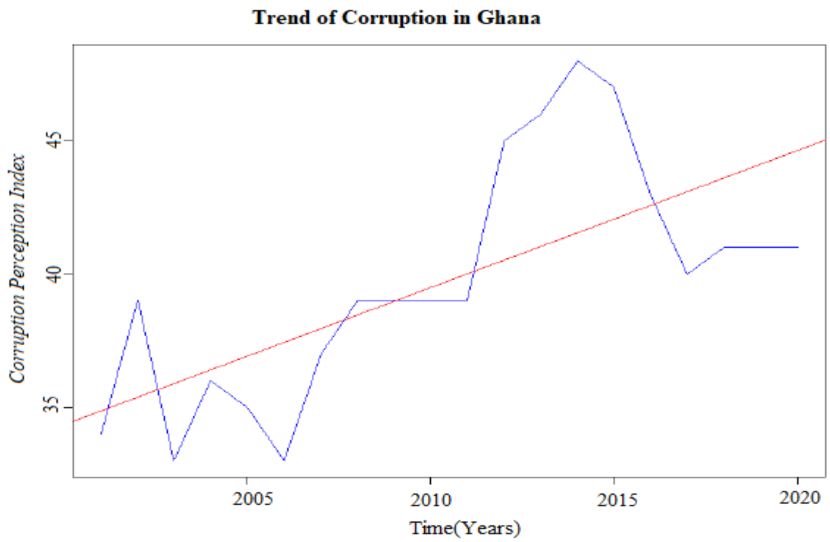

Figure 3: Plot of Corruption Trend

A.4. Validation Plots for Simulated Data

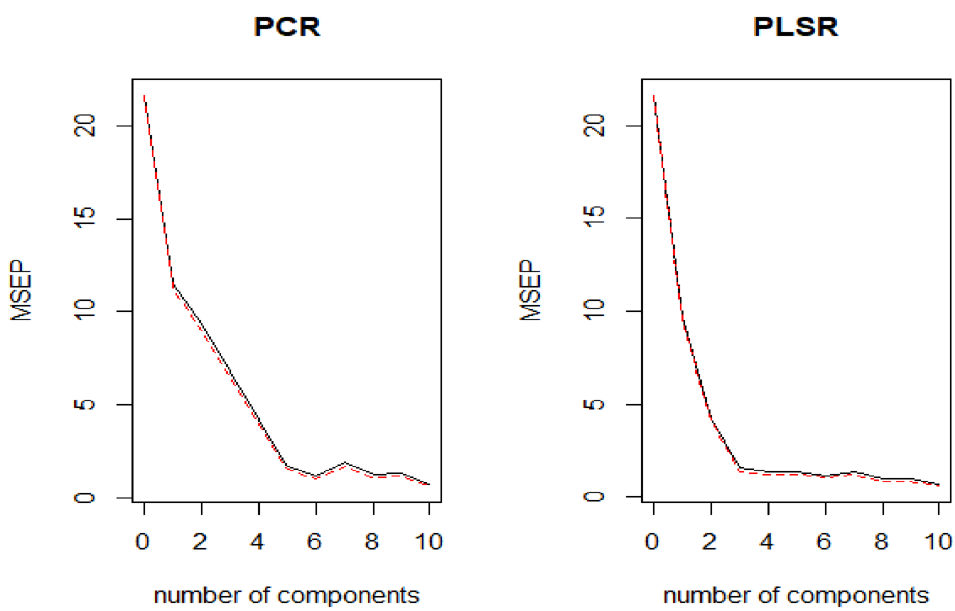

Figure 4: Validation Plots for Sample Size 20

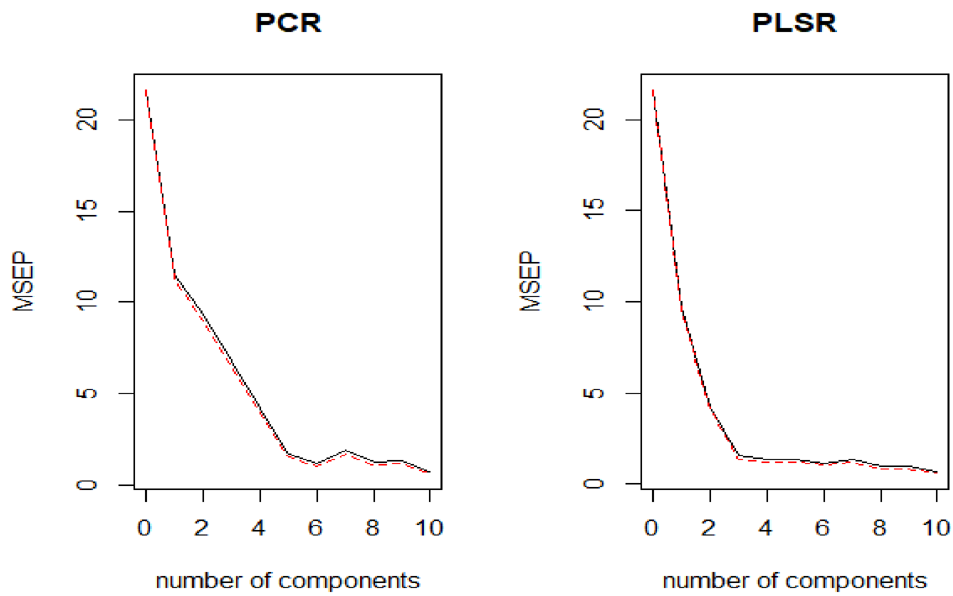

Figure 5: Validation Plots for Sample Size 50

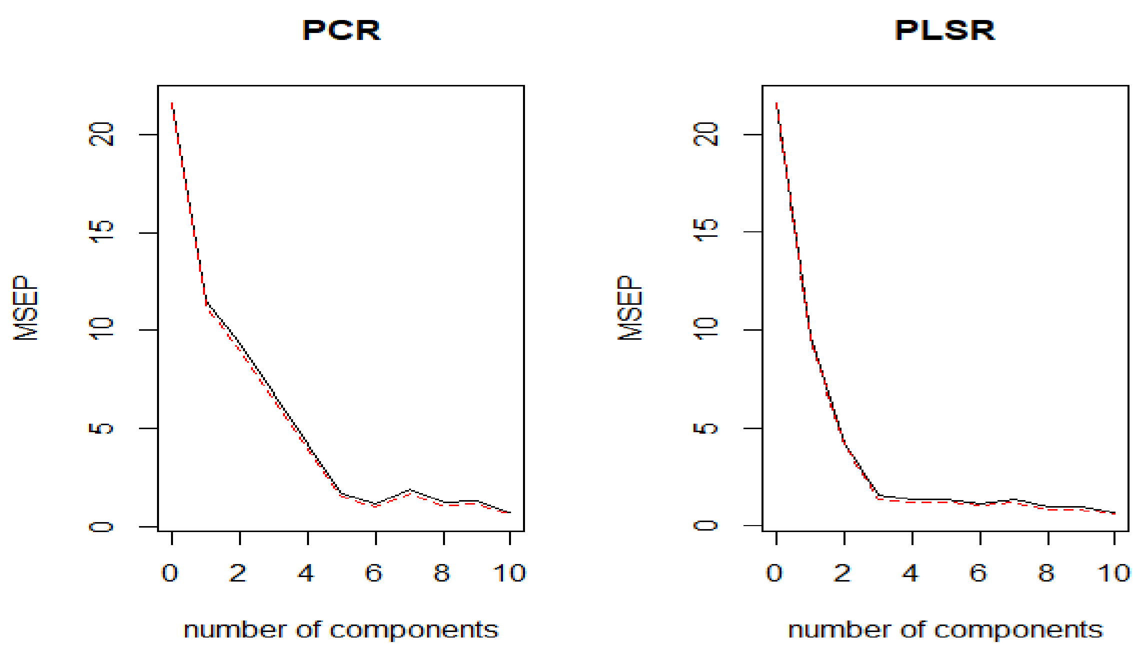

Figure 6: Validation Plots for Sample Size 100

Vol. 5, No.10; 2021
ISSN: 2456-7760

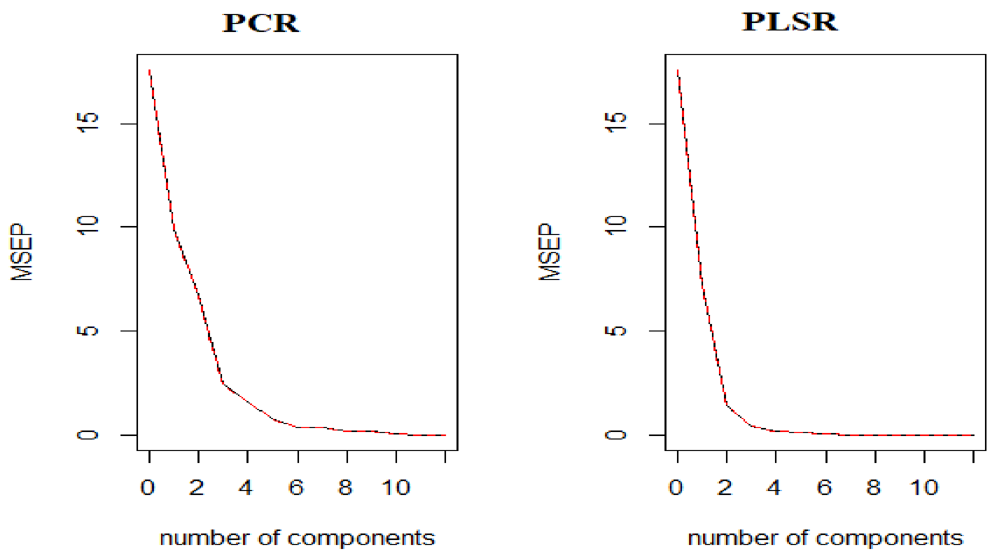

Figure 7: Validation Plots for Sample Size 500

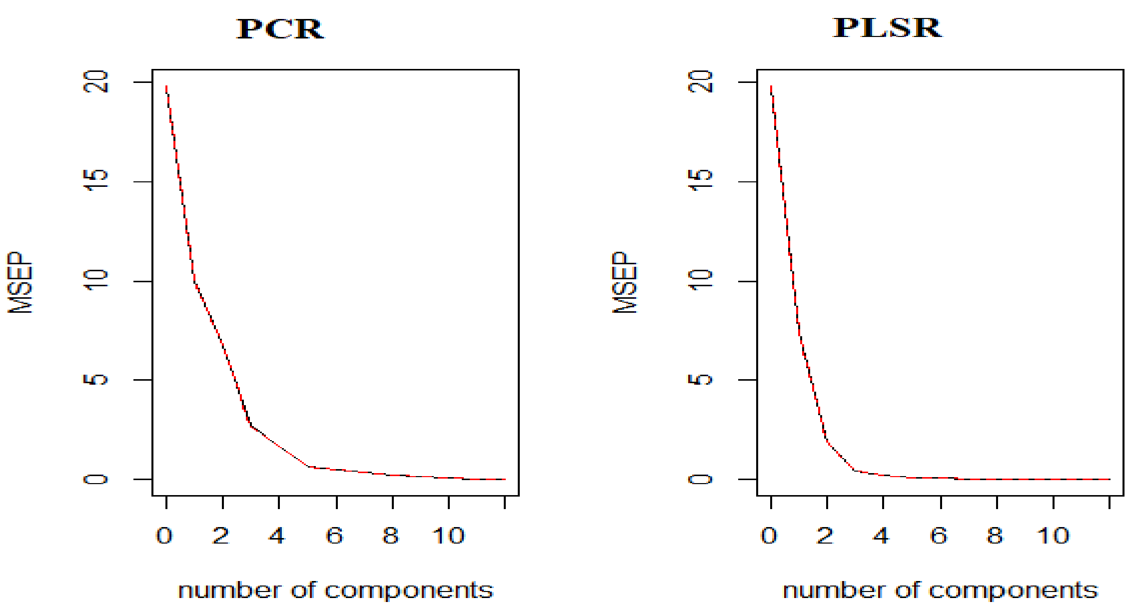

Figure 7: Valid