Stochastic Modeling and Applications
Vol. 24 No. 2 (July-December, 2020)

Received: 15th September 2020 Revised: 19th October 2020 Accepted: 24th October 2020

VECTOR ERROR CORRECTION MODELING FOR FORECASTING CARBON DIOXIDE EMISSIONS BASED ON ECONOMIC INDICATORS

A. JABARALI, A. SHEIK ABDULLAH, AND A. JAWAHAR FAROOK*

Abstract. The long-term consequences of climate change have given a significant role in the economic development of India. The purpose of this paper is to establish a time series vector error correction model for predicting the carbon dioxide emissions(CDE) based on a group of associated economic indicators (i.e. Gross Domestic Product(GDP), Energy Consumption(EC) and Trade(TR)). The CDEs are found to be cointegrated with GDP, EC and TR by employing Johansen's cointegration analysis. This paper addresses to the climate change studies by exploring the dynamic relationships between carbon dioxide emissions and the economic indicators using VEC model imposing cointegration restriction for India over the period 1971-2016. The developed Vector error correction(VEC) model is compared with auto-regressive integrated moving average(ARIMA) and vector auto-regressive(VAR) models for forecasting the long-term carbon dioxide emissions with forecast accuracy measure.

1. Introduction

The increase in greenhouse gases(GHG) emissions is considered as a major problem of global warming in the last few decades and climate change has been the important on-going concern for all societies of developing countries. Worldwide organizations have been attempting to reduce the adverse impacts of global warming and climate change on the economy. The economic growth of developing countries urges intensive use of energy. Due to future increases in atmospheric concentration of greenhouse gases, it is estimated that the annual mean warming about 3C in the decade of the 2050s and about 5C in the decade of the 2080s over regions of Asia($IPCC, 2007$). The GHG emissions are strongly associated to the socio-economics of the developing countries. Therefore, factors affecting the development of the nation are to be identified. The energy sector closely related to the climate change and energy-related emissions account for over two thirds of anthropogenic GHG emissions. Eventhough India's share of carbon dioxide emissions is very insignificant in the world total emissions, carbon dioxide emissions by the energy sector are continuously increasing in all parts of the country[17]. Regression

---

2000 *Mathematics Subject Classification.* Primary 62Mxx, 91Bxx; Secondary 62M10, 91B82, 91B84.

Key words and phrases. cointegration, vector error correction model, auto-regressive integrated moving average, vector auto-regression, economic indicators.

* Corresponding Author.

models provide meaningful predictions of time series variable only when the target variable is stationary[11]. Konea and Buke[19] examined the trends for the carbon dioxide emissions for the top 25 countries and the world total carbon dioxide emissions using regression analysis. Among the various time series models, ARIMA models are extensively used in forecasting a time series variable. Farook and Kannan[6] used Box-Jenkins methodology to build the ARIMA model to analyze and forecast the global atmospheric carbon dioxide emissions. In the univariate time series model, forecasting the value is exclusively based on the past values of the time series. The Box-Jenkins technique, introduced by Box and Jenkins[2], is used to forecast univariate time series variable. However, it predicts the values based on past values and it doesn't investigate the factors assailing behavior. The univariate time series forecasting methods are unsuitable in the absence of any explanatory proficiencies [9]. In the view of that, the advanced forecasting technique should be examined for forecasting carbon dioxide emissions. VEC models are commonly used to analyze the dynamic behavior of time series variables in empirically. The influence of economic growth, energy consumption and trade openness on the environment has become a prevalent problem in the literature. A pioneer work of studying the dynamic relationship between economic growth and environmental pollutants is found[10]. Toda and Yomamoto[27] found the causal relationship from GDP to energy consumption and from energy consumption to carbon dioxide emissions. Trade liberalization with the energy consumption will lead the economic growth but energy use will lead to higher levels of GHG emissions. Ang[1] studied the long run relationship between pollutant emissions, energy consumption and output using cointegration and vector error correction modeling. Halicioglu[12] included trade into the study of identifying the relationship among carbon dioxide emissions, income and energy consumption in the case of Turkey using Autoregressive Distributed Lag($ARDL$) model of cointegration. Fodha and Zaghdoud[7] studied that there is a long-run cointegrating relationship between two pollutant emissions (CO_2 and SO_2) and GDP. Jayanthakumaran et al., [16] used the bounds testing approach to cointegration and the ARDL methodology to obtain the long-run and short-run relationships among carbon dioxide emissions, economic growth, energy consumption and trade to compare China and India. Based on the literature, the specification of carbon dioxide emissions is defined as:

$$CDE = f(GDP, EC, TR) \tag{1.1}$$

where CDE is carbon dioxide emissions; GDP denotes the gross domestic product; EC is the energy consumption; and TR is the trade openness which is the sum of exports and imports of goods and services measured as a share of gross domestic product. For the purpose of model development, the annual data for carbon dioxide emissions (metric tons per capita), energy consumption (kg of oil equivalent per capita), GDP per capita (current US$) and trade openness (% of GDP) of India are collected from the World Development Indicators, World Bank for the period from 1971 to 2016. The statistical characteristics of these variables are tabulated in table 1. The Purpose of this paper is to develop an empirical model for forecasting the carbon dioxide emissions in India by means of the cointegration analysis and VEC modeling. The methodology used here for construction VEC model is explained in some detail in Section 2. The empirical results of the estimation of VEC model using time series data are discussed in Section 3 and the conclusions are drawn in the last section.

2. Methodology

The general VEC model is to forecast the carbon dioxide emissions in the presence of economic indicators. Engle and Granger[4] have initiated the error correction models (ECM). The vector autoregressive model and cointegration restrictions constitute the vector error correction model.

2.1. Cointegration. The long-run relationships between non-stationary variables are identified by the cointegration, which is an important econometric property of time series indicators. If the data is not in the form of stationary but based on some initial difference of linear combination of variables is stationary then the respective time series are said to be cointegrated to the order one or $I(1)$. Eventhough the time series have deviation from each other in the short run, they will come back to the trend in the long run.

Let $Y_t = (y_{1t}, y_{2t}, ..., y_{nt})^{'}$ denote $(n \times 1)$ vector of $I(1)$ time series Y_t is cointegrated if there exists an $(n \times 1)$ vector $\beta = (\beta_1, ..., \beta_n)^{'}$ such that

$$\beta^{'} Y_t = \beta_1 y_{1t} + ... + \beta_n y_{nt} \sim I(0)$$

In other words, the non-stationary time series in Y_t are cointegrated if there is a linear combination of them that is stationary or $I(0)$.

Testing the cointegration among the variables is preceded by unit root test which is conducted to identify whether the times series variables are stationary. Augmented Dickey Fuller test is most commonly used unit root test which is developed[3].

2.2. Vector Error Correction Model. VAR and VEC models are the most recently used advanced multivariate time series, which will give the prediction values of each variable based on its own lags of differences and the lags of all other variables. Since VEC model can provide the better understanding of any non-stationarity nature among the different time series and can also improve the long-run predictions over an unconstrained model. The advanced multivariate model of the VEC model is employed to predict the target variable in the system. Based on prediction accuracy measures, the VEC model and other time series model are suggested that both acceptable for forecasting the future values. However, the VEC model achieves the higher prediction accuracy than the general time series model.

A VAR with $'p'$ order is a function of all lagged endogenous variables in the system. According to Johansen[18], an unrestricted VAR(p) is defined by

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + u_t \qquad t = 1, 2, ..., T \tag{2.1}$$

where Y_t represents the independent variables; $u_t \sim N(0, \Sigma)$ is a vector of impulses which represents the unanticipated movements in Y_t. A's are estimable parameters.

However, this model is only appropriate if every series of Y_t is integrated to the order of zero, $I(0)$, which means every series is stationary[24].

A VEC model is a restricted vector autoregression that has involved the cointegration analysis[22]. According to Wong et al.,[28], the general VEC model is defined as:

$$\Delta Y_t = c + \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + u_t \tag{2.2}$$

where Y_t are the independent $I(1)$ variables being integrated to an $I(0)$ vector; c is the intercept; the matrix Γ reflects the short-run dynamic relationships between the elements of Y_t ; and u_t are residuals; Π matrix contains the long-run equilibrium information; p is the number of lags and $\Delta = (I - L)$, L is the lag operator. Engle and Granger[4] identified that the elements of Y_t are $I(1)$ variables and cointegrated with rank(Π) $= r$. If there are k endogenous variables and the coefficient matrix Π has reduced rank $r < k$, then there exists $(k \times r)$ matrices α and β, each with rank r such that $\Pi = \alpha\beta'$, is the error correction term and $\beta' Y_t$ is stationary. β denotes vector of cointegration relationships and α denotes matrix defining the long-term equilibrium by the cointegrating relationships. There are two cases to consider as follows:

- The error correction could not be derived to the models with a full rank of Π, (i.e $r = k - 1$) in such a situation Y_t is stationary and has no unit root test.
- No stationary long-run equilibrium relationship exists with zero rank of Π, which denotes that the elements of Y_t are not cointegrated.

Johansen[18] has derived two different likelihood ratio tests to determine the number of cointegration vectors: the trace test and the maximum eigenvalue test. The test statistics are defined in the equations (2.3) and (2.4) respectively.

$$LR_{tr} = -T \sum_{i=r+1}^{k} log(1 - \hat{\lambda}_i) \tag{2.3}$$

$$LR_{max} = -T log(1 - \hat{\lambda}_{r+1}) \tag{2.4}$$

where T is the sample size and λ_i is the i^{th} largest canonical correlation. The trace test is computed with H_0 : Rank(Π) $= r$ against H_A : Rank(Π) $> r$, while H_0 : Rank(Π) $= r$ against H_A : Rank(Π) $= r + 1$ is considered for the maximum eigenvalue test. It has been found that the results of the trace test is preferred first in the situation that both the test statistics yield different results.

The choice of lag length of VEC model is determined using Akaike information criterion(AIC), Schwarz Bayesian Information Criterion(BIC) and Hannan-Quinn information criterion(HQC). VEC model is more capable of forecasting time series variables, since it can establish long-run equilibrium relationships between dependent and independent variables while the past equilibrium is used as explanatory variable to describe the dynamic behavior of current variables[5]. Finally, the validity and the adequacy of the developed VEC model is assessed through the forecasting performance measure. The test results of the lag length selection are used in the Johansen cointegration test to develop the VEC models with different combinations between carbon dioxide emissions and each economic indicator. After identifying all variables are stationary and cointegrated, VEC model can be defined for carbon dioxide emissions with the economic indicator as follows:

$$\begin{aligned}\Delta CDE_t = c + \alpha(\beta^{'} Y_{t-1} + \rho_0) + \sum_{i=1}^{p} \gamma_{1,i} \Delta GDP_{t-i} \\ + \sum_{i=1}^{p} \gamma_{2,i} \Delta EC_{t-i} + \sum_{i=1}^{p} \gamma_{3,i} \Delta TR_{t-i} + u_t\end{aligned} \tag{2.5}$$

where α is the adjustment coefficient; β denotes the long-run parameters of the VEC function; $\gamma_{j,i}$ reflects the short-run aspects of the relationships between the explanatory variables and the target variable; ρ_0 is the intercept of cointegrating equations.

3. Interpretation of Computational Results

The distributions of the all variables are moderately positively skewed. The standard deviation of GDP is 296.8949, which shows the large variation compared to other variables.

TABLE 1. Descriptive Statistics for all variables during 1971-2019

Variables	Mean	Std.Dev	Minimum	Maximum	Skewness	Kurtosis
CDE	0.8533	0.3748	0.3626	1.6662	0.495	-0.659
GDP	431.1541	296.8949	120.6968	1417.0736	1.773	2.916
EC	380.1783	89.2148	275.5608	600.3056	0.766	-0.137
TR	21.5797	12.3276	7.5374	52.2695	1.182	0.31

Before constructing VEC model, all the variables should be integrated at the same order for obtaining the stationarity so that the variables CDE, GDP, EC and TR are verified whether they have the same integrated order.

The ADF unit root test is employed here to test the stationary for all the variables. The results of unit root test are presented in Table 2 and it suggests that all the variables are stationary after the first difference at 5% level of significance. From Table 3, the smallest values of AIC, BIC and HQC indicate the lag length for the VEC models is two based on the VAR lag length selection procedure.

TABLE 2. Results of ADF unit root test

Indicators	Level		First Difference	
	t-statistics	p-value	t-statistics	p-value
CDE	-0.16	0.99	-4.36	0.00
GDP	3.05	0.99	-4.34	0.00
EC	1.55	0.99	-5.6	0.00
TR	-0.91	0.94	-8.03	0.00

TABLE 3. VAR lag order selection criteria

Lag	Log L	p (LR)	AIC	BIC	HQC
1	92.5815	NA	-4.8101	-4.5461	-4.718
2	94.4498	0.0532	-4.8583*	-4.5504*	-4.7509*
3	95.4283	0.1618	-4.8571	-4.5052	-4.7343
4	96.1912	0.2168	-4.8439	-4.4481	-4.7058

Table 4 illustrates the results of Johansen cointegration test. The trace statistics along with p-value indicate that there is no more than one cointegrating relationship, while the trace and Lmax test rejects $r = 0$ at 5% level of significance. Therefore, it is found that there is one cointegration relationship among the indicators (CDE, GDP, EC and TR).

TABLE 4. Johansen cointegration test results

Hypothesis		Trace test	p-value	Lmax test	p-value
H_0	H_A				
$r = 0$	$r = 1$	63.877	0.0006	30.934	0.0148
$r \leq 1$	$r \geq 2$	32.943	0.0202	18.84	0.1031
$r \leq 2$	$r \geq 3$	14.103	0.0793	11.748	0.1210
$r \leq 3$	$r \geq 4$	2.3545	0.1249	2.3545	0.1249

Based on the developed VEC model as described in methodology, Table 5 shows the error correction model estimates for carbon dioxide emissions. The specification of the VEC model suggests that the carbon dioxide emissions are affected by gross domestic product, energy consumption and trade openness. It is examined for their model fit based on the values of R-squared and Durbin-Watson statistic. It has a higher R-square value with 0.71 which means approximately 71% of the variations in carbon dioxide could be achieved by the other indicators. Therefore, the lagged independent variables have a significant role to capture the movements in carbon dioxide emissions.

Table 5. Estimates of vector error correction model of carbon dioxide emissions Note: values in parenthesis are t-statistics at 5% level of significance

Variables	ΔCDE_t	
c	0.1104	
α	0.0625 (2.041)	
CDE_t	1	
GDP_{t-1}	0.0016 (2.452)	
EC_{t-1}	-0.0083 (-1.668)	
TR_{t-1}	0.0125 (0.3483)	
ρ_0	0.0328	
Error correction	***t*-1**	***t*-2**
ΔCDE	0.09214 (3.5090)	-0.0618 (-0.2608)
ΔGDP	0.0004 (2.627)	-0.0001 (2.627)
ΔEC	-0.0006 (-0.6770)	0.0012 (1.038)
ΔTR	0.0046 (2.228)	-0.0076 (-2.841)
R-squared	0.7104	
Sum squared residual	0.0172	
S.E of equation	0.01243	
Durbin-Watson	1.8626	

The forecasting accuracy of the VEC model is achieved by comparing the predicted values with the actual values. The predicted values are also compared with the projections found by the Box-Jenkins and VAR models. According to Box-Jenkins methodology[2, 6], the ARIMA $(0, 2, 1)$ is identified to be the best fitted model for forecasting carbon dioxide emissions and the parameters of the model are estimated

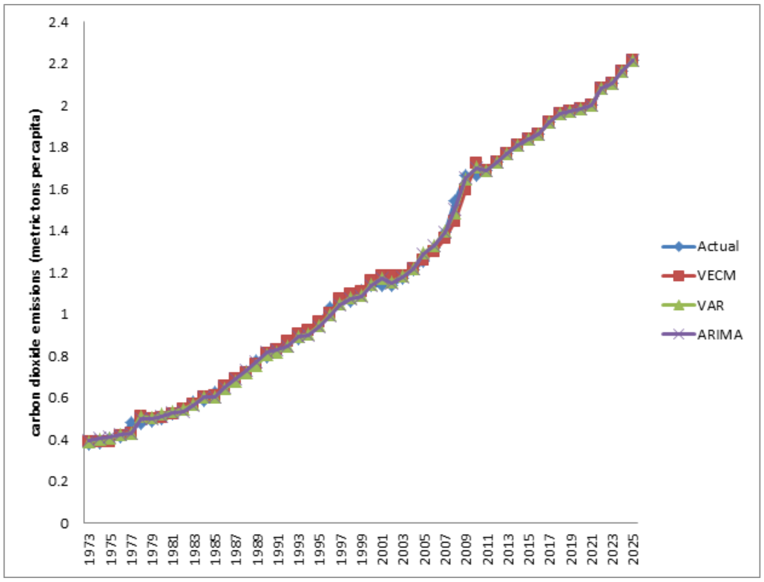

Figure 1 Graphical Representation of Three Different Forecasting Models

using maximum likelihood method. On the other hand, vector autoregressive model with lag order two (i.e. VAR(2)) is developed for forecasting carbon dioxide emissions[8]. Each variable depends linearly on its own lagged values and the lagged values of the other variables in VAR approach.

Table 6. Forecasting accuracy measure of VEC, ARIMA and VAR for predicting carbon dioxide emissions

Measures	VEC	ARIMA	VAR
Mean Square Error(MSE)	0.000405	0.0008	0.000452
Mean Absoute Percentage Error(MAPE)	0.10%	0.18%	0.12%

Table 6 shows the forecasting performances of the three different forecasting models and the graphical representation in Figure 1. It is found that the VEC model has low mean square error and mean absolute percentage error values compared to other models. From the results, the developed VEC model is tolerably efficient to forecast the carbon dioxide emissions.

4. Conclusion

An advanced multivariate model, namely VEC is presented and forecast the carbon dioxide emissions in India. A dynamic specification of the carbon dioxide emissions has been developed with the use of VEC modeling. It is found that GDP, EC and

TR have a long-run equilibrium effect in determining the carbon dioxide emissions in India by applying Johansen's cointegration analysis. The prediction accuracy of the VEC model has been validated against ARIMA and VAR models. It provides reliable forecasts compared to other aforementioned time series models fitted to carbon dioxide emissions. The higher level of the accuracy of predictive has showed that the VEC model is better to predicting the carbon dioxide emissions based on the economic indicators.

Acknowledgement

The authors are thankful to the anonymous referee for checking of the model and for supportive comments that improved this paper.

References

1. Ang, J. B.: CO_2 Emissions, Energy Consumption and Output in France, *Energy Policy*, **35** (2007), 4772–4778. The authors are thankful to the anonymous referee for checking of the model and for helpful comments that improved this paper.
2. Box, G.E.P., and G.M. Jenkins.: *Time Series Analysis: Forecasting and Control*, Oakland Calif: Holden-Day, 1976.
3. Dickey, D.A., and W.A. Fuller.: Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, **74** (1979), 427-3.
4. Engle, R.F., and C.W.J. Granger.: Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, **55**(1987), 251-76.
5. Fan, R.Y.C., Ng, S.T., and J.M.W. Wong.: Reliability of the Box-Jenkins Model for Forecasting Construction Demand Covering Times of Economic Austerity, *Construction Management and Economics*, **28**(2010), 241-54.
6. Farook, A. J., and K. S. Kannan.: Stochastic Modeling for Carbon Dioxide Emissions, *Journal of Statistics and Management Systems*, **17**(2014), 97-112.
7. Fodha, M., and O. Zaghdoud.: Economic Growth and Pollutant Emissions in Tunisia: An Empirical Analysis of the Environmental Kuznets Curve, *Energy Policy*, **38**(2010), 1150-1156.
8. Ghatak, A.: Vector Autoregression Modelling and Forecasting Growth of South Korea, *Journal of Applied Statistics*, **25**(1998), 579-592.
9. Goh, B.H., and H.P. Teo.: Forecasting Construction Industry Demand, Price and Productivity in Singapore: the Box-Jenkins Approach, *Construction Management and Economics*, **18**(2000), 607-618.
10. Grossman, G., and A. Krueger: Economic Growth and the Environment, *The Quarterly Journal of Economics*, **110** (1995), 353-377.
11. Gujarati, D. N., Porter, D.C., and S. Gunasekar.: *Basic Econometrics*, Fifth Edition, Tata McGraw-Hill Education Private Limited, New Delhi, 2009.
12. Halicioglu, F.: An Econometric Study of CO_2 Emissions, Energy Consumption, Income and Foreign Trade in Turkey, *Energy Policy*, **37**(2009), 1156-1164.
13. Heng Jiang and Chunlu Liu: Forecasting construction demand: a vector error correction model with dummy variables, *Journal of Construction Engineering and Management*, *29(9)*(2011), 969-979.
14. Heng Jiang, Youquan Xu and Chunlu Liu: Construction Price Prediction Using Vector Error Correction Models, *Journal of Construction Engineering and Management*, *139(11)*(2013), 1205-1221.
15. James M.W. Wong and S. Thomas Ng: Forecasting construction tender price index in Hong Kong using vector error correction model, *Journal of Construction Management and Economics*, *28*(2010), 1255-1268
16. Jayanthakumaran, K., Verma, R., and Y. Liu.: CO_2 Emissions, Energy Consumption, Trade and Income: A Comparative Analysis of China and India, *Energy Policy*, **42** (2012), 450-460.

17. Jeyalakshmi, S., Parameswaran, V., Mathew, J., and A. Kaur.: Statistics Related to Climate Change-India, *Ministry of Statistics and Programme Implementation, Government of India*, Central Statistical Office, Social Statistics Division, New Delhi, 2013.
18. Johansen, S. Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, **12** (1988), 231-254.
19. Kone, A.C., and T. Buke.: Forecasting of CO_2 Emissions from Fuel Combustion using Trend Analysis, *Renewable and Sustainable Energy Reviews*, **14**(2010), 2906-2915.
20. Nicholas Tsounis and Aspasia Vlachvei: *Advances in Applied Economic Research, Proceedings of the 2016 International Conference on Applied Economics*, Springer Proceedings in Business and Economics, 2017.
21. Luo, Z., Liu, C. and D. Picken.: Granger Causality among House Price and Macroeconomic Variables in Victoria, *Pacific Rim Property Research Journal*, **13**(2007), 234-256.
22. Lutkepohl, H.: *Vector Autoregressive and Vector Error Correction Models*, in Lutkepohl, H. and Kratzig, M. (eds), Applied Time Series Econometrics, Cambridge University Press, 86–158, 2004.
23. Phong B Dao and Wieslaw J Staszewski: Data normalisation for Lamb wave–based damage detection using cointegration: A case study with single- and multiple-temperature trends, *Journal of Intelligent Material Systems and Structures*, *25(7)*(2014), 845-857.
24. Price, S.: *Cointegration and Modelling the Long Run, in Scarbrough, E. and Tanenbaum, E. (eds), Research Strategies in the Social Sciences: A Guide to New Approaches*, Oxford University Press, Oxford, 156-191, 1998.
25. Ryan Y.C.Fan, S. Thomas Ng and James M.W.Wong: Predicting construction market growth for urban metropolis: An econometric analysis, *Habitat International*, *35(2)*(2011), 167-174.
26. Sahbi Farhani, Anissa Chaibi and Christophe Rault:CO2 emissions, output, energy consumption, and trade in Tunisia, *Economic Modelling*,*38*(2014), 426-434
27. Toda, H.Y., and T. Yamamoto.: Statistical Inference in Vector Autoregression with Possibly Integrated Processes, *Journal of Econometrics*, **66**(1995), 225-250.
28. Wong, J. M. W., Chan, A. P. C., and Y. H. Chiang.: Forecasting Construction Manpower Demand: a Vector Error Correction Model, *Building and Environment*, **42**(2007), 3030-3041.

A. Jabarali: Department of Statistics, Madras Christian College, East Tambaram, Chengalpattu - 600 059.

Email address: jabarali@mcc.edu.in

A. Sheik Abdullah: PG and Research Department of Statistics, Salem Sowdeswari College, Salem - 636 010

Email address: sheik.stat@gmail.com

A. Jawahar Farook: Department of Economics and Statistics, Government of Tamil Nadu

Email address: farookstat@gmail.com