Journal of Statistical and Econometric Methods, vol.4, no.1, 2015, 1-38
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2015

Cointegration VAR and VECM and ARIMAX Econometric Approaches for Water Quality Variates

Ghulam Ali*[1]

Abstract

We use vector autoregressive, cointegration, error correction models for water quality to establish relationships between dissolved oxygen (DO), biological oxygen demand (BOD) turbidity and pH, taking into account interdependence between them. Since environmental literature did not use these models, this study fills this gap. Johansen's method which is based on eigenvalues and eigenvectors because of the matrix representation of the coefficients due to multiple equations is used for estimating functional relationship vectors. Interestingly, it makes no distinction between exogenous and endogenous variables. Johansen's approach indicates two plausible DO and BOD functions or relationship vectors. These models provide estimates of the errors if the variable deviates from the long run

* [1] United States Environmental Protection Agency, 1200 Pennsylvania Avenue NW, Washington DC, 20460, USA. E-mail: ali.ghulam@epa.gov

The views expressed herein are those of the author and may not be attributed to the United States Environmental Protection Agency or anyone else associated with it, in any way.

Article Info: *Received* : December 22, 2014. *Revised* : February 1, 2015.
Published online : March 31, 2015.

path. In the differenced DO equation, coefficients of differenced DO_{t-1}, temperature, time and error correction are significant, whereas coefficient of differenced BOD_{t-1} is insignificant. However, in differenced BOD and turbidity equations coefficients are insignificant. The error correction model also provides a reasonable explanation of the functional relationship of DO. Alternatively, we also use the ARIMAX model for explaining the dynamic behavior of these variables to contrast with the error correction version of the model. ARIMAX presented functional relationships for BOD and turbidity, although none was found with error correction model.

Mathematics Subject Classification: 62H12; 62M10; 62P12

Keywords: Cointegration; VAR; VECM; Johansen approach; Water Quality; ARIMAX

1 Introduction

For environment analysis, the Box and Jenkins methodology is used to establish functional relationships between various time series variables. An autoregressive, moving average, and autoregressive integrated moving average class of models are usually used for stochastic processes. The Box and Jenkins approach suggests differencing of the variables in order to deal with the nonstationarity of the time series. However, differencing of the series is not always desirable because it restricts the model to only short-run variations, eliminating the role of long-run variations which could be a salient feature of the stochastic process. Moreover if the series behave in an equilibrium fashion it may be useful to estimate the magnitude of the deviations from the equilibrium path. In addition, if nonstationarity is ignored, a relationship could be established when in fact none existed; spurious conclusion can be drawn from the regression of nonstationary series (Granger and Newbold [1]). Therefore, one might need to incorporate not only the long-run but also the short-run variations in the model. Although, analysis can be based on the short-run variations only but it might not be very meaningful since long run variations, if relevant, are ignored in establishing a relationship. Alternatively, de-trended variables could be used to bypass nonstationarity, but the dynamic model yet again would characterize the long-run phenomenon only. In addition, arbitrarily selecting some variables as dependent and others exogenous is not a preferred approach for analyzing interrelated time series. Hence, an alternative methodology is needed.

The cointegration approach is presented by Granger to address these kinds of issues. To deal with nonstationarity, there are two versions of the cointegration models, vector autoregressive (VAR) and vector error correction model (VECM). If the coefficient matrix associated with the stochastic equations representing relationships for variables is less than full rank, then either (1) cointegration restrictions on the coefficients are applied or (2) the error correction representation is used or (3) the random walk component is separated out leaving the estimation to proceed in the VAR representation. The error correction model not only includes differenced but also level variates. Therefore, it is useful to adopt this version of the vector autoregressive model because it provides both the short run and long run parameters.

Usually, an invariance assumption regarding stochastic process has been made to establish functional relationships between time series. Technological progress, changes in people's preferences and behavior, policy or regime and institutional developments affect environments. As a result, environmental processes may not be stationary. Dissolved oxygen (DO), biological oxygen demand (BOD), turbidity, and potential hydrogen (pH) are important variables for water quality. These variables could be interrelated and cointegrated. However, one may not know the exact nature of the association between these variables.

These might deviate from each other from time to time and still generally move together. They could deviate from each other in the short run but move together in the long run. They could be simultaneously endogenous in nature, therefore, assuming arbitrarily order of variables and use of univariate models may not be appropriate for such time series. Instead, a systems approach or a simultaneous equations approach could be more appropriate at least as a starting point of the investigation. Interestingly, a combination of the stochastic equations under cointegration approach would create a stationary process even when the variables are nonstationary.

Cointegration is extensively used in macroeconomics for macro variables, such as the money supply, interest rate, inflation and exchange rates. Although it may seem unlikely, cointegration is even used in development economics. This is because a frequently made assumption about the invariance often leads to erroneous forecasts if macro variables are indeed nonstationary (see Hendry and Juselius [2]). The environmental variable and gross domestic product could move together and thus could be cointegrated. Therefore, it is used in development economics. In economic development literature, environmental variables are used to test Kuznets curve hypothesis that maintains that relationship between economic growth and environmental pollution is an inverted U shape in developing countries. Per capita income and the environmental indicators such as sulphur dioxide (SO_2), nitrogen oxide (NO_x), carbon dioxide (CO_2) in the atmosphere, dissolved oxygen (DO) depletion, and biological oxygen demand (BOD) in water bodies are inversely related to economic growth beyond certain threshold of income. In the initial stages of development, environmental degradation increases but when a specific threshold of economic growth is reached, it begins to decrease according to the curve hypothesis. Grossman and Krueger [3] used cubic function to establish an inverted U shaped relationship. They state that although the turning point for individual pollutant varies, for most pollutants improvement begins to occur when the income reaches the neighborhood of $8,000 per capita.

Panayotou [4] showed an inverted U between deforestation and per capita income. Seldon and Song [5] show a similar relationship for carbon monooxide and NO_x. Giles and Mosk [6] had had an inverted U relationship for the emission of methane (CH_4). Nasir and Rehman [7] show Kuznets curve in the long run for carbon but not in the short run in Pakistan. However, some of the studies contradict the validity of the environmental Kuznets curve. Holtz-Eaken and Selden [8] show no Kuznets curve for CO_2. Hettige, *et al.* [9] find no improvement in water pollution with the income growth. De Bruyn, *et al*. [10] show no decline in CO_2, NO_x, and SO_2 with income. Neither Roca *et al*. [11] support the inverted U relation for CO_2, N_2O (nitrogen oxide), CH_4, and NOx and volatile non methanic compounds, except SO_2, nor Akpan and Agabi [12] for CO_2. Rather, Akpan and Agabi seem to support the pollution haven hypothesis.

On the other hand, in environmental literature, not including economic development literature, there are a number of studies that have used an autoregressive moving average (ARMA) type modeling for environmental variables. For instance, Sun and Koch [13] use a univariate autoregressive integrated moving average model (ARIMA) for salinity in Apalachicola Bay in Florida. Govindsamy [14], and Salas [15] use the Box and Jenkins methodology for surface water process such as stream flow and precipitation. Durdu [16] used an autoregressive integrated moving average or multiplicative seasonal autoregressive integrative moving average model to explain Boron concentration in a water body. The author applied this methodology to determine trends in the Boron time series. He suggested use of an ARIMA model for predicting the mean of the time series because it was considered to fit the time series reasonably well. Similarly, Ahmad, *et al*. [17] used an ARMA model to predict river water quality. Kurunc and Cevic [18] applied an ARIMA model for some water quality variables and stream flow. Ragvan and Fernandez [19] used seasonally adjusted ARIMA model for long-term trend of water quality. Benyaha, *et al*. [20] applied autoregressive and periodic autoregressive models for weekly maximum temperature time series. Hassanzadeh [21] used ARIMA for SO_2 levels. Ali [22, 23] used several time series models for pathogen indicators time series. This seems to highlight that there is an increasing trend in the time series studies to use approaches based on the Box-Jenkins univariate framework for the analysis of the water quality time series. What is not used at all is the cointegration methodology; we are unaware of any study that has used this methodology. Literature review shows no attempt to go beyond the Box-Jenkins core methodology. This study aimed to fill the gap in the literature. Although Box-Jenkins core methodology has been extensively used in hydrology, Box-Jenkins differencing of variables is not preferred for creating stationary series because it eliminates the long term dependency in the data. Differencing essentially deals with the short-term dependency. Variables can deviate from each other in the short-term but still can display long term relationship. Therefore, differencing is not always a reasonable transformation methodology for establishing long term relationships. Differencing can be used for the short term analysis.

This study differs from the previously mentioned economic development studies because it focuses on the interaction between the environmental variables rather than the per capita income and environmental variables in reference to Kuznets curve hypothesis. It tests the equilibrium between DO, BOD, turbidity and pH rather than an economic equilibrium, while cointegrating the time series in simultaneous equations fashion. Similarly, it differs from the above-mentioned ecological studies (not economic studies) because it attempts to estimate the error correction elements in relation to interaction of environmental variables. The error correction modeling approach not only determines the short run but also long-run effects.

This study introduces the cointegration methodology in hydrology. To the best of our knowledge, there is no study in the environmental literature that has used cointegration to deal with the water quality time series. This study also differs from other studies in the water quality arena, because it uses Johansen's eigenvector approach for cointegration. Having said that, we back track a little that this study might not be unique in the entire environmental literature because it might be mixed with other type of literature.

In the remainder of this study, we present the vector autoregressive (VAR) model, vector error correction model (VECM), parametric estimation and hypothesis testing, in section 2, along with mathematical details. We present the study site and data sources in section 3, and discuss the vector autoregressive models' output in section 4. We explain the output of Johansen's approach based on eigenvalue and eigenvectors for cointegration in section 5, VECM output in section 6, pairwise cointegration analysis in section 7, and autoregressive integrated moving average output in section 8. And finally, we present conclusions in section 9.

2 Econometric Approaches

2.1 Models

The models specifications proceeds with the use of the vector autoregressive framework. The endogenous variables and the coefficients are presented respectively in a vector and matrix form due to multi equation system. Lagged variables serve as explanatory variables and each lag results in its coefficient matrix, while error terms are stacked in a vector form (see Lutkepohl [24] and Juselius [25] for stacking the observations and equations).

2.2 VAR

The VAR (1) model is specified as

$$x_t = \Pi x_{t-1} + \Phi Z_t + \varepsilon_t \qquad \text{and} \qquad \Pi = \alpha\beta', \tag{1}$$

representing independent or exogenous variables by Z and error vector by $\varepsilon_t \sim IN(0, \Omega)$, $E(\varepsilon_t \varepsilon_t') = \Omega$ and $E(\varepsilon_t) = 0$. Π is an impact matrix which is expressed as a linear combinations of α and β. We can linearly combine the basis vectors (β) to get Π (similar to use of the standard basis ($such\ as$ 1, 0, 0), (0, 1, 0), (0, 0, 1) to create other vectors). However, we would need to know the basis vectors. The Johansen methodology which employs maximum likelihood function provides the basis vectors using normal distribution.

2.3 Vector Error Correction Model (VECM)

The vector error correction form of the model is

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + \Phi Z_t + \varepsilon_t \text{ , and}$$

$$\Pi = \alpha\beta'. \tag{2}$$

Here Δ is a difference operator. Δx_t and x_{t-1} are separately regressed on lagged Δx_t and independent variables in order to exclude the effect of lagged Δx_t and independent variables from these variables. This eliminates the short-run coefficients, $\Gamma_{1,}\Gamma_{2.} \ldots \Gamma_{p-1}$, and coefficients of independent variables, leaving the long-run coefficient to be estimated. The respective residuals R_0 and R_1 without the effect of lagged Δx_t and independent variables are used to specify the model

$$R_{0t} = \alpha\beta' R_{1t} + \varepsilon_t. \tag{3}$$

Post multiplying $R_0 = \alpha\beta'^{R_1} + \varepsilon$ with $R'_1\beta$, denoting $S_{ij} = \sum \frac{R_i R_j'}{T},\ and$

$i, j = 0{,}1.$ we get

$$\alpha = S_{01}\beta(\beta' S_{11}\beta)^{-1}. \quad (4)$$

This value is substituted in the concentrated likelihood function. The log maximum likelihood function for this model is specified as

$$\log L = -\frac{PT}{2} log 2\pi - \frac{T}{2} log|\Omega| - \frac{1}{2} tr[(R_0 - \alpha\beta' R_1)'\Omega^{-1} (R_0 - \alpha\beta' R_1)] \quad (5)$$

with Ω representing the covariance matrix. This is equivalent to

$$L_{max}^{-\frac{2}{T}} = |\widehat{\Omega}| + contant\ terms = |T^{-1} \sum(R_0 - \alpha\beta' R_1)\ (R_0 - \alpha\beta' R_1)'| + const.\ terms,$$

$$|\widehat{\Omega}| = |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta S_{10}|. \quad (6)$$

The only term that is optimized with respect to β is omega because other terms are constant. The min-max theorem is used to minimize the covariance expression in order to maximize the likelihood function. Rayleigh Quotient, $\frac{X'MX}{X'NX}$, is applied because block determinant of covariance is expressed in this form to determine eigenvectors that span the beta space. This approach finally provides

$$\tilde{\beta} = [v_1, \dots, v_p] S_{11}^{-1/2} \text{ and } \alpha = S_{01}\tilde{\beta}(\tilde{\beta}' S_{11}\tilde{\beta})^{-1}, \quad (7)$$

where $v_1, \dots, v_j$ are the eigenvectors corresponding to the eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0$ in a decreeing order (greater the value of λ, the greater the stationarity of the process). The advantage of this approach is that given eigenvalues (λ) are also used to test cointegration hypothesis. For further explanation of the derivation, see Johansen [26], Juselius [25] and Lutkepohl [24].

2.4 Transitory VECM

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{p-1} \Delta x_{t-p+1} + \Phi Z_t + \varepsilon_t. \quad (8)$$

Here only Π changes, rest of the coefficients remain the same.

2.5 Short Term ARIMAX

$$\Phi(L)\Delta x_t = \psi(\Delta z_t) + \Theta(l)\epsilon_t. \tag{9}$$

2.6 Cointegration Test

The likelihood ratio and $\lambda - max$ test are usually used in testing cointegrating hypothesis. The above estimated eigenvalues are not only used in likelihood ratio test for null hypothesis $H_0 = -T\sum ln(1-\lambda_r)$ versus alternative $H_1 = -T\sum ln(1-\lambda_p)$ but also in $\lambda - max$ test to test null $-Tln(1-\lambda_r)$ versus $-T(1-\lambda_{r+1})$. The $\lambda - max$ method is based on the logic that if (r+1)th lambda is zero then all the lambdas below it are also zero. For further explanation see Johansen [25].

3 Data

3.1 Site and Data

The data we used to test these models were the monthly time series of environmental measurements associated with a lagoon which is located between 41° 00’ N and 28° 45’E in the southwest of Istanbul, Turkey. The lagoon covers an area of 15.22 square kilometers with the drainage area 340 km^2. The maximum depth of the lagoon reaches 20 meters. Three streams feed the lagoon. However, the water supply has been reduced because of a dam that was built in the watershed in order to supply potable water to Istanbul. The lagoon is connected to the Sea of Marmara by a channel. The data was collected between 2005 and 2008. The data has large gaps which we did not fill, because it would not make any significant difference in our objective of presentation and demonstration of the application of the cointegration methodology. This data was presented in another study (Taner, et al [27]), therefore we did not want to change it; it could introduce distortions. The overall time series consists of the 35 observations. During monitoring, the study measured DO, pH, electrical conductivity, salinity, chemical oxygen demand, orthophosphate, nitrate and chlorophyll-a. Because the time series is small, we did not attempt to consider the possibility of the functional relationships between all these variables. We instead focus on DO and BOD, turbidity and pH which are more important variables.

4 VAR Model

To set the stage of the investigation of the functional relationship, we use the ordinary least squares methodology for DO, given the exogenous variables such as BOD, turbidity, and pH. For the time being, we ignore the cointegration of the time series. For DO functional relationship, other than the temperature and time, none of the variables are significant. This shows that the ordinary least squared (OLS) model did not fit the data. This was expected because it is not an appropriate methodology for our data. We do not present the outputs of regressions because it is not the focus of this paper.

4.1 VAR Output

To establish a functional relationship of the time series, we used the VAR model with the number of equations $p = 4$. VAR does not make assumptions about the dependency order of the variables in a system. An endogenous variable usually depends on a few other endogenous variables. It is an alternative to a big Keynesian macro econometrics model with many restrictions. Sims [28] argued that economic theory is not rich enough to suggest proper identification restrictions on the structural VAR. VAR estimates long-run relationship. To determine the order of the VAR model, we used the information criteria of Akaike (AIC), Schwarz (BIC) and Hannan and Quinn (HQC). The criteria values for two of the criteria are minimized (AIC = 13.88, BIC = 15.15, HQC = 14.31) at lag one rather than at lag two and three. Therefore, in functional estimation we limit the VAR order to one. In the DO equation, only the constant, temperature and trend are significant variables. The remaining coefficients are all insignificant, no different from zero. To make matters worse, BOD, turbidity and pH individual functional relationships have no significant variables with the exception of only a few. This briefly shows that the VAR did not fit the data. This is not surprising, because simple VAR is not suitable for nonstationary time series. As an example of the functional relation by VAR, the model output related to the DO equation is presented in Table 1. We do not present the rest of the model output to economize space for this article.

If the full rank is a realistic assumption, we run the error correction model with $r = 4$. If this specification is true then error correction term in VECM should be zero. EC1 is found to be significant in the DO equation. Other error correction terms are statistically insignificant. In the pH equation, error correction terms are insignificant. Similarly, all the error correction terms are statistically no different from zero in the BOD equation. By contrast, all the error correction terms are significant in turbidity equation. These results of the model roughly confirm that the impact matrix (Π) is not a full ranked matrix. The model output is not presented because it is not crucial to the rest of the paper.

Table 1: DO VAR

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	Indicator of low p-value
Constant	14.2613	2.9839	4.779	6.02e-05	***
DO_1	0.1839	0.1312	1.402	0.1728	
BOD_1	-0.0921	0.1581	-0.5829	0.5650	
turbidity_1	-0.0052	0.0119	-0.4366	0.6660	
pH_1	-0.5258	0.4164	-1.263	0.2179	
Temperature	-0.1722	0.0179	-7.576	4.83e-08	***
Time	0.0426	0.0179	2.377	0.0251	**
R^2	0.7838		Adjusted R^2	0.7339	
Rho	-0.0533		Durbin-Watson	2.0659	

In the table, DO_1, BOD_1, turbidity_1 and pH_1 represent lagged variables. In the last column, the greater the number of asterisks (*), the smaller the p value points out the level of significance of the variable.

5 Testing and Johansen Method Output

DO, BOD, turbidity and pH are hypothesized to be interdependent variables. It is assumed that there might be equilibrium between this set of variables, as a starting point of investigation of the dynamic relationship. These time series could move together in the long run while deviating in the short run from a long run or an equilibrium path due to shocks to the system caused by the short-run disturbances.

To determine the lag structure of the system, we used the Akaike, Schwarz and Hannan and Quinn information criteria. Criteria values were greater at a lag order of three than two. On the other hand, values were lower at a lag order of two than one. This shows that the true lag order lies between one and three bounds, suggesting a lag order of two for the system of equations.

If Π is in fact of full rank (i.e. $|\Pi| \neq 0$) then the system is stationary. It is not only without unit roots but also equivalently without stochastic trends. Simply, at full rank r ($p - r = 0$), the VECM does not include the short-term variables and converges to the stationary VAR model. However, for an evolving system of variables it is not a reasonable approach (see Hendry and Juselius [2]).

On the other hand, if the rank is zero, then $\Pi = 0$ is a null matrix. In this case, Π has zero number of eigenvalues that are different from zero. This means variables are not cointegrated and the long-term relationship does not exist between the series because Δx_t does not depend on x_{t-i}. In other words, the error correction term, Πx_{t-1} equals zero and drops out of the error correction model, leaving the differenced series in the model. Differenced series is usually stationary anyway. Rank determines the largest possible order of a matrix that has a nonzero determinant, derived from Π with eigenvalues. Or simply, rank is the number of nonzero eigenvalues (or the number of relationships). To apply the cointegration model, therefore, it is critical to determine the rank of the coefficient matrix. In order to determine the rank between the above-mentioned two extreme limits (zero and full rank), we use the trace test. However, to begin, one has to also test for an existence of a unit root in an individual series before applying the trace test.

5.1 Augmented Dicky Fuller Test

To test the existence of unit root, we used the Augmented Dicky Fuller (ADF) test without drift. According to this test, under the null, coefficient $\gamma = 0$, whereas under alternative $\gamma < 0$. The hypothesis testing with ADF is opposite to the usual hypothesis testing because the test statistic below the critical value is deemed to reject the null hypothesis. For DO, the test statistic, τ (-0.32), is greater than the critical value (-1.95), falling in the acceptance region, suggesting an integration or a unit root. Similarly, test statistic (-0.29) for BOD exceeds the critical value suggesting integration. Same is true for turbidity ($\tau =$ -0.84), even for pH ($\tau =$ -0.14), suggesting a unit root, because τ falls in the acceptance region of the test, at five percent significance level.

In order to test the unit root, as an alternative, we used the ADF test with drift. The test statistics (τ), -2.17,-1.70,-2.64 for BOD, turbidity and pH, respectively, exceed the critical value (-3.0) at a five percent significant level. Hence, the null hypothesis of unit root is not rejected, underscoring integration. However, for DO, τ is -3.58. Because of close association of the DO we include it for cointegration set. Moreover, if any of the variables is integrated of order 1, then the linear combination is also integrated of order 1, therefore we include it in the cointegration model.

5.2 Trace Test and $\boldsymbol{\lambda} - \mathbf{max}$ test

To run the error correction model, we have to determine the rank, r, of the system. We test for the cointegration by using the system level approach i.e., the Johansen trace test or the $\lambda - max$ test. The test statistic, τ (104.3) exceeds the critical value (55.1) at zero cointegration rank, rejecting the null of utmost zero value of trace. The p-value falls below 0.05 indicating cointegration rank greater than zero. Similarly, test statistic (42.7) exceeds the critical value (35.0), rejecting the null of trace one. On the contrary, τ (9.5) falls below the critical value (18.3), at cointegration rank of two, failing to reject the null hypothesis. The p-value exceeds the critical value supporting the null hypothesis. This appears to suggest cointegration rank is two of the matrix. $\lambda - max$ test did not pose any contradiction to this cointegration rank.

5.3 Johansen Approach Output

We used the system's based Johansen's approach with temperature as an exogenous variable and the unrestricted constant and the unrestricted trend. This test gives the Π matrix and its decomposition into α, the adjustment vector, and cointegration vector β when the system deviates in the short run from the long run functional relationship. This matrix represents the long run relationship.

The output of Johansen's approach or Johansen test, presented in Table 2, shows that two (0.85 and 0.65) of the eigenvalues are large compared to rest of the eigenvalues. This implies that there are at least two cointegrating long term relationships. The corresponding column vectors to these eigenvalues thus form the functional relationships.

Johansen model output indicates that the coefficient of the BOD, in first column of beta matrix, is close to zero. This indicates this variable may come first in the order of influencing other variable within the set of DO, BOD, turbidity and pH. Hence this variable should not be normalized. In other words it may not be a dependent variable in the functional relationship. Likewise, the turbidity coefficient is not very different from zero. This demonstrates that turbidity is not a very strong candidate for a dependent variable either. Similarly, the coefficient of pH is close to zero in the most significant eigenvector. Hence, it may not be a strong candidate for a dependent variable in the first relationship. Strictly speaking, none of three variables seems to be dependent variable in the most significant cointegration vector. On the other hand, coefficient of DO (0.99) is close to one in the eigenvector corresponding to the largest eigenvalue. This implies that this variable must be last in the order of influence in that vector. It means that DO is a dependent variable. As we stated, in the first cointegrating vector, the coefficient of DO is close to 1, hence one can divide the coefficients in the first column by $0.99 \approx 1$ to normalize the coefficient to prove the functional form for DO. If the coefficient is truly zero then the coefficients cannot be identified because zero cannot be used for normalization (division by zero is not possible). In that case, that variable is dropped. Normalization is essential for the uniqueness of the long term relationship. Thus, cointegration model highlights that the DO equation represents an important long term functional relationship.

To find the other functional form, we look into the diagonal elements of the beta matrix. The coefficient of pH in the second vector is not close to 1. Moreover, turbidity coefficient is not different from zero. In addition, the DO coefficient is very small. Thus, none of these variables can be a candidate for a dependent variable for the second functional relationship. On the other hand, BOD coefficient is roughly close to 1 (1.32). Hence this variable can be normalized. This may indicate that the BOD is the last in the order of influence among the variables. Therefore, BOD could represent the second long term functional relationship although not very strong, according to Johansen's approach. Because of vector form specification of the model, p values are not estimated by the Gretl software, which was used for estimation, therefore we did not present them here; p-values are estimated usually for scalers not for vectors. However, p-values are presented in the VECM framework, which would serve roughly the same purpose.

In the model output, beta represents the long-term impact, while alpha represents the magnitude of the adjustment of the cointegration vector when the relationship is deviating from the long-term relation of the process. The ith column of the beta (β) matrix shows the coefficients of each of the variables in the ith variable equation (ith cointegrating relationship), while the ith row indicates the contribution of the ith variable (same) to each of the relationships. Similarly, ith vector of alpha (α) matrix shows the speed of adjustment of the each relationship to the disequilibrium in ith relation, while the row vector shows the adjustment speeds of the ith relation to each of the disequilibrium in relationships.

Table 2: Cointegrating Beta and Adjustment Alpha Vectors

	Coefficient β $Matrix$				α $Matrix$			
	1st equation	2n equation	3rd equation	4th equation				
DO	0.9856	-0.4632	0.0238	0.9810	-1.1514	0.1723	0.0162	0.0311
BOD	0.1222	1.3227	-0.7641	-0.8684	-0.2138	-0.1685	0.4145	0.0737
Turbidity	-0.0162	-0.1145	-0.0491	-0.0195	-1.6058	8.0170	2.8220	-0.1886
pH	-0.1640	1.7579	1.6842	-4.6206	-0.0830	0.0579	-0.0934	0.0512
	Renormalized β Matrix				Renormalized α Matrix			
DO	1.0000	-0.3502	-0.4853	-0.2123	-0.135	0.2278	-0.0008	-0.1437
BOD	0.1240	1.0000	15.576	0.1879	-0.211	-0.2229	-0.0203	-0.3404
Turbidity	-0.0165	-0.0866	1.0000	0.0042	1.5827	10.604	-0.1384	0.8716
pH	-0.1664	1.3290	-34.332	1.0000	-0.082	0.0765	0.0046	-0.2365
	Long Run Π $(\alpha\beta')$ Matrix							
Equation	DO	BOD	Turbidity	pH				
DO	-1.1837	0.0477	-0.0024	0.3753				
BOD	-0.0505	-0.6298	0.00102	0.0966				
Turbidity	-5.4139	8.4149	-1.0269	19.981				
pH	-0.0606	0.0933	-0.0017	-0.2785				
Eigenvalue	0.8537	0.6463	0.2244	0.0413				

The DO relationship shows very low speed of adjustment (close to zero) to the disequilibrium in individual turbidity and pH relationships. The adjustment speeds of other relations are low because of pH disequilibrium. Similarly, pH relationship shows very little adjustment speed to disequilibrium in other relationships. Generally speaking, if alpha is a null vector (row) there is no adjustment of the relation presented by that row as a result of the disequilibrium in relations. It means that, that specific variable is affecting but is not being affected by other variables. This implies that variable is most likely a weakly exogenous variable. On the other hand, if alpha is a unit vector then the relationship is adjusting to disequilibrium in a particular relationship. This means that variable is most likely a dependent variable. Alpha seems to play almost the same role which an error correction term plays in VECM. If the dependent variables are normalized, the adjustment coefficients also change because the beta vectors change. The estimated beta, alpha and impact matrixes are presented in Table 2.

6 VECM

6.1 VECM Output

VECM includes the short run variations (changes from one period to another) also in the evolution of the functional form. For instance, suppose there is a long run relationship between two time series variables as $y = bx$ then the short run deviation $(y - bx)$ in period t-1 could affect the relationship in period t. Thus, the model determines the magnitude of the error correction term (coefficient α) as $\alpha(y - bx)$. Suppose the error correction term (EC) is -0.1. This means that the 10 percent of the error will be corrected in the next period in converging to the long run relationship. Thus, the model analyzes the shocks to the system, while estimating the functional form of the time series variables. To proceed for the estimation with VECM for long term relationships with the short term dynamics, we need to make the assumption about the rank of the impact matrix. Johansen test indicated the rank order 2 of the impact matrix. We use this order for VECM estimation. For the BOD equation, the relationship seems to be very poor. Not only the coefficients of lagged differenced BOD, DO, turbidity, and time and temperature, but also the constant, EC1, and EC2 all are no different from zero. In the turbidity equation, the intercept and coefficients of lagged differenced BOD and DO, and temperature are found to be insignificant. In the pH equation, only lagged differenced pH and DO are significant variables, while coefficients of lagged differenced BOD and turbidity, and temperature, and the parameters of trend, EC1 and EC2 are no different from zero. This shows these relationships are not strong. BOD relationship is the worst among the relationships. However, DO seems to display a very strong relationship. Lagged differenced DO, turbidity and pH are found to be significant variables. Similarly, trend, EC1 are significant parameters inversely related to differenced DO. This means that the relationship is corrected when it deviates from the long-run relationship. However, EC2 and BOD are not found to be relevant to differenced DO equation. R^2 is 0.86 with adjusted R^2, 0.83, for DO equation. VECM framework demonstrates that DO has a strong relationship which was confirmed by Johansen's framework as well.

Interestingly, the coefficient of the DO is close to one. In the functional cointegration relationship, it comes last in the order of influence. This proves or shows that in the cointegration relationship, DO is a dependent variable. Hence this variable is a normalized variable. This is an important finding. Not only are the temperature, trend and EC1 significant but also the differenced DO, turbidity and the pH at t-1 are significant explanatory variables. Moreover, not only the temperature, lagged differenced turbidity and BOD but also pH are inversely related to differenced DO. However, the coefficient of lagged differenced BOD is not different from zero. The sign of EC1 is negative, which is logical, while EC2 is statistically zero. It means when the system overshoots it is adjusted downward and when the system undershoots it is corrected upward.

Table 3: Equation: d_DO

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	Indicator of low p-value
Constant	8.9235	1.4106	6.3262	<0.00001	***
d_DO_1	0.4394	0.1190	3.6930	0.0013	***
d_BOD_1	-0.0427	0.1497	-0.2855	0.7779	
d_turbidity_1	-0.0207	0.0100	-2.0692	0.0505	*
d_pH_1	-0.7751	0.3705	-2.0922	0.0482	**
Temperature	-0.1699	0.0191	-8.9179	<0.00001	***
Time	0.0746	0.0154	4.8456	0.0001	***
EC1	-1.2146	0.1202	-10.1068	<0.00001	***
EC2	0.0871	0.1466	0.5940	0.5586	
R^2	0.8780		Adjusted R^2	0.8280	
Rho	-0.2736		Durbin-Watson	2.4978	
DO and BOD Vectors					
	β			α	
DO_1	1.0000	0.00000		-1.2146	0.0871
BOD_1	0.0000	1.0000		-0.1326	-0.2490
Turbidity_1	-0.0055	-0.0885		-5.2960	10.408
pH_1	-0.3175	1.2179		-0.10858	0.0664

In the table, d is used to indicate a differenced variable, such as d_DO, while d_DO_1, d_BOD_1, d_turbidity_1 and d_pH_1 represent lagged differenced variables. Also the greater the number of asterisks (*), the smaller the p value to point out the level of significance of the variable (last column).

Table 4: Equation: d_turbidity

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	Indicator of low p-value
Constant	-58.7311	23.1197	-2.5403	0.0187	**
d_DO_1	3.3252	1.9500	1.7053	0.1022	
d_BOD_1	-4.9821	2.4540	-2.0303	0.0546	*
d_turbidity_1	0.1833	0.1638	1.1189	0.2752	
d_pH_1	-5.1424	6.0724	-0.8469	0.4062	
Temperature	0.0178	0.3123	0.0569	0.9551	
Time	-0.5236	0.2524	-2.0745	0.0499	**
EC1	-5.2960	1.9698	-2.6887	0.0134	**
EC2	10.4076	2.4026	4.3317	0.0003	***
R^2	0.5212		Adjusted R^2	0.3253	
Rho	-0.1801		Durbin-Watson	2.3541	

In the table, d is used to indicate a differenced variable, such as d_DO, while d_DO_1, d_BOD_1, d_turbidity_1 and d_pH_1 represent lagged differenced variables for the VECM framework. Also the greater the number of asterisks (*), the smaller the p-value to point out the level of significance of the variable (last column).

We present the VECM output in Tables 3, 4 and 5. We assumed a rank of two according to the rank test. It seems that the DO indeed has a genuine relationship in the VECM framework of cointegration. We do not present differenced BOD equation to economize space for this article. We also used the transitory version of the VECM model, which presents the short term parameters only[†]. The results were not very different from that of other error correction model. The output of this approach is not shown to conserve space for this article.

[†] We used R software for this purpose (see Pfaff, B. [29]).

Table 5: Equation: d_pH

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	Indicator of low p-value
Constant	0.2904	0.8843	0.3284	0.7457	
d_DO_1	0.1800	0.0746	2.4130	0.0246	**
d_BOD_1	-0.0576	0.0939	-0.6138	0.5456	
d_turbidity_1	0.0010	0.0063	0.1520	0.8806	
d_pH_1	-0.6396	0.2323	-2.7539	0.0116	**
Temperature	-0.0230	0.0119	-1.9297	0.0666	*
Time	0.0082	0.0097	0.8483	0.4054	
EC1	-0.1086	0.0753	-1.4412	0.1636	
EC2	0.0664	0.0919	0.7223	0.4777	
R^2	0.4307		Adjusted R^2	0.1978	
Rho	-0.0930		Durbin-Watson	2.1058	

In the table, d is used to indicate a difference variable, such as d_DO, while d_DO_1, d_BOD_1, d_turbidity_1 and d_pH_1 represent lagged differenced variables for the VECM framework. Also the greater the number of asterisks (*), the smaller the p-value to point out the level of significance of the variable (last column).

6.2 Ljung-Box Test

For testing autocorrelation of residuals we use Ljung-Box test. This test like most of the other tests is based on the Chi-square distribution, because it involves the sum of square of the normalized correlations (sum of square of normal variates has Chi-square distribution). As was expected, the test fails to reject the null hypothesis of no autocorrelation of residuals at lag one for each equation supporting the model specification or the model fit, implying that residuals are independently distributed. Ljung Box test statistic for each equation is presented in Table 6.

Table 6: Ljung-Box Test for Normality of Residuals

	DO	BOD	Turbidity	pH
Chi-square value	2.627	0.042	1.138	0.303
P-value	0.105	0.837	0.286	0.582

6.3 Doornik-Hansen Test

Usually Doornik-Hansen test for the multivariate normality of residuals or errors is used to check the model specification. This test is based on the transformed square of skewness and kurtosis with the adjustment for small samples. Since it involves the sum of square of normal or standardized variates, it is a Chi-square test. Doornik-Hansen test statistic shows Chi-square 6.872, with p-value 0.550, supporting the null hypothesis of the multivariate normality of residuals.

Ljung Box and Doornik-Hansen tests both together imply that the residuals are independently and normally distributed to support the model specification.

6.4 Lagrange Multiplier Test

To further check if autoregressive conditional heteroscedasticity (ARCH) effects are present in the model because of conditional heteroscedasticity, we use the Lagrange Multiplier (LM) test. The LM test uses the lagged variables among the explanatory variables with the residuals as the dependent variable. If the coefficients of the lagged variables are zero it implies no ARCH effects in the data. At a lag order of one, the Chi-square (0.0006) is far below the critical value, with the p-value 0.980 for DO equation. Similarly, Chi- square (1.536) falls below the critical limit with the p-value 0.215 for the BOD equation. Thus, the LM test fails to reject not only the null hypothesis for DO but also for BOD, demonstrating absence of conditional heteroscedasticity of residuals. This suggests that residuals are homoscedastic. Similarly, the estimated Chi-square test statistics is 0.187 and 0.082 with p-values of 0.665 and 0.773 respectively for turbidity and pH equations. This shows the null hypothesis is not rejected implying no ARCH effects in residuals. Similarly, LM finds no ARCH effects at lag 2 for any of the equations. This supports the model specification as far are error terms are concerned in modeling.

7 Pairwise Cointegrasion Analysis

7.1 BOD and Turbidity

If we used the Johansen test for the determination of the order of the cointegration, the null hypothesis of not only zero but also one is rejected. This may imply that we cannot reach a conclusion with respect to the number of cointegration vectors. However, eigenvalue 0.49 seems significant enough to suggest one cointegration vector or one functional relationship. We use the VECM for this set of the time series. For the turbidity equation, constant, time trend and error correction terms are significant. The t value (3.86) exceeds the critical value (1.70) for the cointegration coefficient of turbidity. Moreover, the coefficient is not close to 1. However, none of the variables is significant in BOD equation. This may show that the BOD does not depend on the turbidity. On the other hand, turbidity is influenced by BOD. We may conclude this is not a very reasonable relationship.

7.2 BOD and pH

We used the Johansen test to determine the rank of the impact matrix (Π). The hypothesis of not only zero but also one was rejected. Thus, we could not reach the conclusion about the precise order of the cointegration rank. The eigenvalues are rather low in magnitude. This may indicate that we do not have a strong relationship as the VECM implied.

According to VECM, the relationship is not well established between these variables. For BOD only the constant and error correction term are significant. For pH, the situation is almost the same. The integration coefficient is almost zero for pH which indicates that pH does not depend on BOD. It might support a proposition that none of the variables is a normalizeable variable.

7.3 Turbidity and pH

We applied the Johansen trace test to determine the rank. The test results rejected the null of zero and 1 both. So we cannot reach the conclusion for the rank order. The eigenvalues (0.38, 0.23) are also not very significant. Thus, one may expect a weak relationship if it does exist. Assuming rank order of 1 for VECM, only lagged differenced pH and the temperature are statistically significant variables in the turbidity equation. Temperature is inversely related to turbidity. For the pH equation, constant, trend and EC1 are significant. In conclusion, there is some uncertainly about the above relationships. It is possible that the Johansen approach is not suitable in determining the rank order of the set of two variables.

8 ARIMAX Model

The long run relationship is determined by the VECM in the presence of the level variables, given the short-run dynamic behavior of the variables. However, VECM did not show any relationship for BOD and turbidity. Above all neither Johansen's approach nor VECM takes into account the moving average process. To determine whether there is indeed no relationship for BOD, in view of the short-term dynamic behavior without the presence of level variables, we use the differenced ARMAX model which takes into account not only autoregressive but also moving average counterpart for individual time series.

8.1 Short-Run ARIMAX for DO

To trace the dynamic behavior of DO, we used the differenced ARMA model with exogenous variables (ARIMAX). ARIMAX is not a substitute for VECM because it does not deal with the endogeneity of the variables in the multi equations framework. Moreover, it is a univariate approach. However, it reflects the relationship on the basis of the short-run dynamic behavior independent of the long-run variations. Differenced turbidity (Δturbidity) is not found to be relevant exogenous variables, in the short run. Nor is the time found to be a relevant variable in the short run. These variables are dropped from the model specification. Differenced BOD and pH (Δ BOD and ΔpH) are significant variables. Not only ΔBOD but also ΔpH has inverse relationship with ΔDO. Temperature is also a significant variable, inversely related to ΔDO. The ARIMAX output is shown in Table 7.

Table 7: ARIMAX Model for DO

$$\Delta x_t = \alpha + \phi_1 \Delta x_{t-1} + \psi(\Delta z_t) + \theta(L)\epsilon_t$$

	Coefficient	std. error	t-ratio	p-value	Indicator of low p-value
constant	0.7085	0.3715	1.907	0.0565	*
phi 1	-0.6054	0.3259	0.3259	0.0632	*
theta 1	0.7912	0.2287	3.459	0.0005	***
ΔBOD	-0.4266	0.1964	-2.172	0.0299	**
ΔpH	-1.2873	0.4133	-3.114	0.0018	***
temperature	-0.0552	0.0261	-2.114	0.0345	**
		Information Criteria Values			
Log-likelihood	-38.9825	Akaike	91.9650	Schwarz	102.4405
		Hannan-Quinn	95.4897		

In the table, Δ is used to indicate differenced variable, such as ΔBOD in the ARIMAX framework. Also the greater the number of asterisks (*), the smaller the p-value to point out the level of significance of the variable (last column).

8.2 Testing of Residuals

To check the model specification, as was stated earlier, usually Doornik-Hansen test for the normality of residuals or errors is used. The test statistic, Chi-square (0.808) value, does not fall in the rejection region. The test is unable to reject the null hypothesis, because p-value is 0.668, implying that the errors are normally distributed. To further check if autoregressive conditional heteroscedasticy (ARCH) effects are present in the model because of conditional heteroscedasticity, we use LM test. With the estimated LM statistic 1.116 and p-value 0.290, we rule out the presence of ARCH effects in the model. We conclude that the residuals are homoscedastic. This lends support to our fitted model.

8.3 Short-Run ARIMAX for BOD

In an application of VECM model, we could not find a relationship for BOD. Estimated coefficients of differenced DO, BOD, turbidity, pH and even temperature and time were all insignificant. According to VAR model, BOD did not yield any long term relationship either. Therefore, as an alternative to the above models, for possible relationship, we use the differenced ARMAX model. We find that ΔDO influence the ΔBOD since coefficient of ΔDO is significant. ΔDO is inversely related to the ΔBOD which is consistent with the theory. Not only ϕ_1 but also θ_1 is significant parameter. It shows that there is short term persistence. Time trend turns out to be statistically no different from zero either. Similarly, Δ pH and Δ turbidity are found to be nonsignificant variables. Differenced turbidity reflected inverse relation to ΔBOD. As a result, we dropped these variables from the specification. The ARIMAX model output is displayed in the Table 8.

Table 8: ARIMAX Model for BOD

$$\Delta x_t = \alpha + \phi_1 \Delta x_{t-1} + \psi(\Delta z_t) + \theta(L)\epsilon_t$$

	Coefficient	std. error	t-ratio	p-value	Indicator of low p-value
constant	0.0364	0.0218	1.666	0.0957	*
phi 1	0.5199	0.1661	3.130	0.0018	***
theta 1	-1.000	0.0822	-12.16	5.04e-034	***
ΔDO	-0.5169	0.1128	-4.582	4.59e-06	***
		Information Criteria Values			
Log-likelihood	-34.1750	Akaike	78.3500	Schwarz	85.8325
		Hannan-Quinn	80.8676		

In the table, Δ is used to indicate differenced variable, such as ΔDO in the ARIMAX framework. Also the greater the number of asterisks (*), the smaller the p-value to point out the level of significance of the variable (last column).

8.4 Hypothesis Testing for Residuals

Similar to the above model, we check if the residuals of the model are normally distributed. Ljung-Box test shows that the residual are normally distributed because the Chi-square (2.682) does not exceed the critical value, while p-value reaches 0.262. Thus, we rule out the misspecification of the model. Likewise, the LM test rules out the ARCH effects because test statistic (1.436) does not fall in the rejection region with p-value 0.231. Absence of conditional heteroscedasticity seems to confirm that the model is correctly specified.

8.5 ARIMAX Model for Turbidity

In cointegration modeling we did not find an adequate relationship. However, in the differenced ARMAX modeling, we find that pH is affecting the turbidity. In addition, not only ϕ_1 but also θ_1 is significant parameter. This shows persistence because turbidity is affected by the previous period turbidity. Since the cointegration modeling showed that the rank of the system is at most 2, it implies that some of the variables cannot be normalized under the cointegration framework. Differenced ARMAX model seems to provide an alternative explanation of dynamic behavior for the variables that could not be assumed to have unit coefficient. Coefficients of ΔDO, ΔBOD and temperature are inversely related to the turbidity. However, not only ΔDO and ΔBOD but also time and temperature were irrelevant (nonsignificant) variables. Therefore, we dropped them. However, we found an autoregressive component problematic for the residuals of the model to behave normally and therefore dropped it from the final output presented in Table 9.

Table 9: ARIMAX Model for Turbidity

$$\Delta x_t = \alpha + \phi_1 \Delta x_{t-1} + \psi(\Delta z_t) + \theta(L)\epsilon_t$$

	Coefficient	std. error	t-ratio	p-value	Indicator of low p-value
constant	-0.7639	0.1885	-4.053	5.07e-05	***
theta 1	-1.0000	0.1002	-9.975	1.96e-023	***
ΔpH	-11.870	5.495	-2.160	0.0308	**
		Information Criteria Values			
Log-likelihood	-126.442	Akaike	260.884	Schwarz	260.884
		Hannan-Quinn	262.898		

In the table, Δ is used to indicate differenced variable, such as ΔpH in the ARIMAX framework. Also the greater the number of asterisks (*), the smaller the p-value to point out the level of significance of the coefficient or variable (last column).

8.6 Test of Residuals

Similar to the other ARIMA models presented in this article, we test the normality of the residuals in the turbidity model with Ljung-Box test. The null hypothesis with respect to normality of errors is not rejected by the test statistic 5.047 with p-value 0.080. Nor did we find any ARCH effects in the model because Chi-square 0.088 with p-value 0.776 of LM test both in combination rejected it. So we find no evidence of the conditional heteroscedasticity of error term. As a result, misspecification of the model is rejected supporting the model fit.

8.7 ARIMAX Model for pH

To trace the dynamic path of pH, we used differenced ARMAX (1,1). Higher order model fared worse. Information criteria were minimized by the ARMAX (1,1). Differenced DO and turbidity are significant variables. P-value indicates that constant is no different from zero, at five percent significance level, although at 10 percent probability level the parameter reverses in significance. The model did not find relevance of the ΔBOD and temperature for ΔpH. ARIMAX output is shown in Table 10.

Table 10: ARIMAX Model for pH

$$\Delta x_t = \alpha + \phi_1 \Delta x_{t-1} + \psi(\Delta z_t) + \gamma t + \theta(L)\epsilon_t$$

	Coefficient	std. error	t-ratio	p-value	Indicator of low p-value
constant	-0.0637	0.02667	-2.388	0.0170	**
phi 1	0.4377	0.1770	2.473	0.0134	**
theta 1	-1.0000	0.0839	-11.91	1.02e-032	***
ΔDO	-0.1353	0.0397	-3.412	0.0006	***
Δturbidity	-0.0105	0.0037	-2.811	0.0049	***
Time	0.0030	0.0014	2.087	0.0369	**
		Information Criteria Value			
Log-likelihood	1.2634	Akaike	11.4731	Schwarz	21.9487
		Hannan-Quinn	14.9979		

In the table, Δ is used to indicate a differenced variable, such as ΔDO in the ARIMAX framework. Also the greater the number of asterisks (*), the smaller the p-value to point out the level of significance of the variable (last column).

8.8 Residual Hypothesis Testing

Following the same approach we followed for other ARIMA model we test the normality assumption of pH model. The test statistic does not exceed the critical value; χ^2, 0.128, with p-value (0.938) indicates errors are normally distributed. Similarly, the LM test (0.088) does not support null of ARCH effects in the model with p-value 0.766. So the conditional heteroscedasticity of error terms in the model is rejected.

9 Conclusions

Cointegration methodology is usually used in macroeconomics for establishing relationships between time series variables, such as money supply, interest rate, inflation, price and gross domestic product. It is used even in the development economics to test inverted Kuznets curve hypothesis. According to inverted Kuznets curve, environmental degredation is positive in the initial stage of growth and reverses to negative once the per capita income exceeds a certain threshold level in a country. Although this methodology has been used by many studies in other fields, it has not been applied even rarely in the environmental field. This study fills this gap in the literature. In hydrology, the Box and Jenkins methodology has been used. To deal with the nonstationarity the Box and Jenkins methodology dictates differencing of the time series variables. However, differencing excludes the level variables, which may not be always useful practice. It may be important, even crucial, to use both the short and long run variations for analysis. We are not aware of any study in the water quality arena that has used the error correction model. Our study used the error correction modeling to determine the role of the external shocks to the system in the water quality arena. It differs from other cointegration studies if any in the water quality arena, because it uses the Johansen's eigenvector approach. Because of the climate change, population growth, technological changes, institutional developments such as the internet, regulations and the change in individual behavior towards environments, the environmental time series cannot be apriorily considered as stationary; they might very well be nonstationary due to these kinds of factors. Therefore, taking into account nonstationarity is essential in establishing relationships between variables.

Johansen's methodology provides the long run relationships and the magnitude of adjustment if the system deviates from the long run relationship. On the other hand, the error correction representation provides the relationships on the basis of differenced variables in the presence of level variables. It also determines the magnitude of the error correction term if the system converges to a long run equilibrium from the short run variations. We use cointegration and vector error correction modeling for dissolved oxygen, biological oxygen demand, turbidity and the potential hydrogen time series to establish functional relationships. We treat them as endogenous variables. To begin with, we used the Cointegration Regression Dicky Fuller test for unit root. The test statistic rejects the existence of unit root in the residuals. The test statistic falls in the acceptance region for the null of no unit root. We used the Johansen trace test to determine the number of cointegrating vectors. It rejected not only the null of zero but also one for the rank of the coefficient matrix. The test statistic suggested the value of two for the rank. However, the second integration relationship may not be very well supported or very weakly supported by the data. For the biological oxygen demand, turbidity and potential hydrogen equations not only the coefficients of the explanatory variables but also the error correction terms are no different from zero. However, the coefficients of disolved oxygen are significant. Only not the trend but also the error correction term is significant.Temperature is a significant exogenous variable in the disolved oxygen equation. The sign of the error correction term is negative. It implies that the relationship is corrected whenever the system overshoots or undershoots due to external shocks. This may seem to suggest there is equilibrium between the variables.

To extend the dynamic behavior analysis, we also used the ARIMAX. Interestingly, contrary to the cointegration, which did not show the relationship, ARIMAX showed a strong relationship between differenced biological oxygen demand and differenced dissolved oxygen, where none existed according to cointegration modeling. Similarly, it showed existence of a functional relationship of differenced turbidity when none was shown by the error correction model. For turbidity autoregressive part was found to result in nonnormal residuals therefore we dropped it in the final outcome. Differenced turbidity is indicated to be influenced by ΔpH only and ΔpH is influenced by turbidity and ΔDO both. A lag order of 1 is common to all the variables, showing persistence.

We also applied the transitory error correction model. The results were not very different from that of other error correction model. Cointegration approach seems to be very useful for determining parameters for water quality analysis. It could be used in water quality models to control pollutants. It could be utilized in determining the total maximum loadings of pollutants and determining resultant interaction between the water quality variables such as dissolved oxygen, biological oxygen demand, turbidity and potential hydrogen. Above all, it could be employed to identify the long run and the short run parameters and importantly to determine disequilibrium if any between certain environmental variables. Although, Johansen's methodology compared to error correction is difficult to interpret, it is needed while useful to determine the dependent variables in an interdependent simultaneous equations set up and to estimate the independent relationship vectors for analyzing nonstationary time varying environmental processes.

Acknowledgements: The author is thankful to Elizabeth Southerland, Thomas Wall, James Pendergast, Deborah Nagle and Julie Hewitt for encouragement. In addition he wishes to thank Lisa E. Faeth for inspiration.

References

[1] C.W.J. Granger and P. Newbold, Spurious regression in econometrics, *Journal of Econometrics*, **2**(2), (1974), 111-120.

[2] D.F. Hendry and K. Juselius, Explaining cointegration analysis: part 1, *Energy Journal*, **21**(1), (2000), 75-120.

[3] G. Grossman, and A. Kruger, Economic growth and the environment, *Quarterly Journal of Economics*, **110**(2), (1995), 353-377.

[4] T. Panayotou, Empirical tests and policy analysis of environmental degredation at different stages of economic development, Geneva, Switzerland, International Labor Office, *Working Paper*, (1993), WP238.

[5] T. Selden, and D. Song, Environmental quality and development: is there a Kuznets curve for air pollution emissions? *Journal of Environmental Economics and Management*, **27**(3), (1994), 147-162.

[6] D. E. Giles, and C. Mosk, Ruminant eructation and a long term Kuznets curve for enteric methane in New Zealand: conventional and fuzzy regression, *Econometric Working Paper*, **0306**, (2003), Canada, Department of Economics, University of Victoria.

[7] M. Nasir, and F. U. Rehman, Environmental Kuznets curve for carbon emission in Pakistan: an empirical investigation, *Energy Policy*, **39**(3), (2011), 1857–1864.

[8] D. Holtz-Eakin, and T.M. Selden, Stoking the fire? CO_2 emissions and economic growth, *Journal of Public Economics*, **57**, (May, 1995), 85-101.

[9] H. Hettige, M. Mani, and D. Wheeler, Industrial pollution in economic development: Kuznets revisited, *World Bank Working Dissection Paper*, **402**, (1999), Washington, D.C: World Bank.

[10] S. M. De Bruyn, J. C. J. M. Van Den Bergh and J. B. Opshoor, Economic growth and emissions; reconsidering the empirical basis of environmental Kuznets curve, *Ecological Economics*, **25**(2), (1998), 161-173.

[11] J. Roca, E. Padilla, M. Ferri, and V. Gelleto, Economic growth and environmental pollution in Spain: discussing the environmental Kuznets curve hypothesis, *Ecological Economics*, **39**(1), (2001), 85-99.

[12] U.F. Akpan, and A. Chuku, Economic growth and environmental degredation in Nigeria: beyond the environmental Kuznets curve, (2011), Department of Economics, University of Uyo, Nigeria.

[13] H. Sun, and M. Koch, Analysis and forecasting salinity in the Apalachicola Bay, Florida: use of ARIMA models, *Journal of Hydraulic Engineering* **127**(10), (2001),718-727.

[14] R. Govindsamy, Univariate Box-Jenkins forecasts of water discharge in Missouri River, *International Journal of Water Resources Development*, **7**(3), (1991), 168-175.

[15] J. D. Salas, *Analysis and modeling of hydrological time series*, In *Handbook Hydrology*, Maidment (edt), McGraw Hill, New York, 1993.

[16] O. F. Durdu, Stochastic approaches for time series forecasting of boron: a case study of Western Turkey, Environmental *Monitoring and Assessment* **169**, (2010) 687-701.

[17] S. Ahmad, I. H. Khan, and B.P. Parida, Performance of stochastic approaches for forecasting river water quality, *Water Research*, **35**(18), (2001), 4261-4266.

[18] Y.A. K. Kurunc, and O. Cevik, Performance of two stochastic approaches for forecasting water quality and streamflow data from Yesiltrmak River Turkey, *Environmental Modeling and Software*, **20**, (2005), 1195-1200.

[19] A.J. Ragavan, and G. C. Fernandez, Modeling water quality trend in long-term time Series, In *Proceedings of SAS Users Group Meeting (SUGI 31)*, Paper 205-31, (2006).

[20] L. Benyaha, A. St-Hilaire, T.B.M.J. Quarda, B. Bobee, A.N. Behrouz, and B.A. Nedushan, Modeling of water temperatures based on stochastic approaches: case study of the Deschutes River, Journal *of Environmental Engineering and Science*, **6**(4), (2007), 437-448.

[21] S. Hassanzadeh, F. Hosseinibalam, and R. Alizadeh, Statistical models and time series forecasting of sulphur dioxide, a case study Tehran, *Environmental Monitoring and Assessment*, **155**(1-4), (2008), 149-155.

[22] G. Ali, ARCH, GARCH and ARMAX models for forecasting pathogen indicators and advisories at marine recreational sites, Marine *Resource Economics*, **26**(3), (2011), 211-224.

[23] G. Ali, EGARCH, GJR-GARCH, TGARCH, AVGARCHN, NGARCH, IGARCH and APARCH models for pathogens at marine recreational sites, *Journal of Statistical and Econometric Methods*, **2**(3), (2013), 271-273.

[24] H. Lutkepohl, *Econometric analysis with vector autoregressive models*, Chapter 8 in *Handbook of Computational Economics* (Belsley DA, Kontoghiorghes EJ (eds), Wiley Online Library, 2009.

[25] K. Juselius, *The Cointegrated VAR Model: Methodology and Applications*, New York: Oxford University Press Inc., 2006.

[26] S. Johansen, Statistical analysis of cointegration vectors., *Journal of Economic Dynamics and Control*, **12**(2-3), (1988), 231-254.

[27] MÜ Taner, B. Üstün, and A. Erdinçler, A simple tool for the assessment of water quality in polluted lagoon systems: a case study for Küçükçekmece Lagoon, Turkey, *Journal of Ecological Indicators*, **11**(2), (2011), 749-756.

[28] C. Sims, Macroeconomics and reality, *Econometrica*, 48(1), (1980), 1-48.

[29] B. Pfaff, VAR, SVAR and SVEC models: implementation within R package vars, *Journal of Statistical Software*, 27(4), (2008), 1-32.