MATEMATIKA: MJIAM, 2019, Volume 35, Number 2, 139–147

Non-transformed Principal Component Technique on Weekly Construction Stock Market Price

[1]Yusrina Andu, [2]Muhammad Hisyam Lee* and [3]Zakariya Yahya Algamal

[1,2]Department of Mathematical Sciences, Universiti Teknologi Malaysia
81310 Johor Bharu, Johor, Malaysia

[3]Statistics and Informatics Department, University of Mosul
Mosul, IRAQ

*Corresponding author: mhl@utm.my

Article history
Received: 3 October 2018
Received in revised form: 5 March 2019
Accepted: 18 March 2019
Published online: 1 August 2019

Abstract The fast-growing urbanization has contributed to the construction sector becoming one of the major sectors traded in the world stock market. In general, nonstationarity is highly related to most of the stock market price pattern. Even though stationarity transformation is a common approach, yet this may prompt to originality loss of the data. Hence, the non-transformation technique using a generalized dynamic principal component (GDPC) were considered for this study. Comparison of GDPC was performed with two transformed principal component techniques. This is pertinent as to observe a larger perspective of both techniques. Thus, the latest weekly two-years observations of nine constructions stock market price from seven different countries were applied. The data was tested for stationarity before performing the analysis. As a result, the mean squared error in the non-transformed technique shows eight lowest values. Similarly, eight construction stock market prices had the highest percentage of explained variance. In conclusion, a non-transformed technique can also present a better result outcome without the stationarity transformation.

Keywords Construction stock market price; nonstationary; non-transformed; stationarity test; time series data

Mathematics Subject Classification 62H25, 62M10, 62P99

1 Introduction

Buildings constructions and materials suppliers have led to the construction sector as one of the key contributors towards a country development. Remarkably, construction stock market price is listed as one of the notable sectors worldwide. In general, non-stationarity is a common pattern in the stock market price, which also includes the construction sectors. Indeed, stationarity transformation is the popular approach in handling such datasets, such as first difference

[1] linear [2] and log-transformation [3]. Although it is a common practice, however, it may prompt to losing originality of the stock market price data.

Nevertheless, one of the approaches to overcome this is through principal component technique. In this study, ordinary principal component (OPC) were used as a baseline [4]. Noteworthy, this technique has the capability to reduce the dimension and simultaneously keep any existing variation in the data as much as possible. There is a strong connection between OPC and factor models because they both evaluate variances, although with different execution. Thus, the advancement in factor models also was incorporated in OPC. Therefore, following the dynamic factor model, the principal component was extended to Brillinger dynamic principal component (BDPC) [5]. Unlike OPC, BDPC used spectral density matrix instead of covariance matrices to find the closest approximation to spectral density matrix of a given reduced rank. Despite the widely used of both methods, they are mostly for the transformed series.

An alternative approach has been recently introduced, using a non-transformed technique namely generalized dynamic principal component (GDPC) [4]. The main precedence of GDPC lies in its ability to examine any mixed pattern of a time series data. Certainly, not all statistical approaches have superiority over other approaches as it also has its own limitations. However, the GDPC technique used in the studymay provide a better adaptation to the non-transformed time series.

Thus, the comparisons of mean squared error (MSE) and the percentage of explained variance were carried out on the transformed and non-transformed principal component techniques [6]. Organization of this article was as follows. Methods related to this study was layout in Section 2. Further details on the construction stock market price and comparison results were shown in Section 3. Conclusions of the findings were in Section 4.

2 Methodology

2.1 Methods Outline

The following subsections will lay the methodology that were used throughout the study. It begins with the stationary test in Subsection 2.2 which is essential in identifying the stationarity pattern of the construction stock market price series. Subsection 2.3 discussed on the principal component technique and is divided into two parts. Firstly, on the transformed technique, consisting of OPC and BDPC. Secondly on the non-transformed technique that is GDPC. When the construction stock market price is nonstationary, both OPC and BDPC are restricted in handling the data directly. Thus, for this reason, they had to be transformed prior to any further analysis. The stationary transformation was carried out using first difference and log-transformation for both OPC and BDPC construction stock market price. As for GDPC, it has this added advantage of using the nonstationary construction stock market price directly without any prior transformation. Finally, this section is concluded with information criteria method for determining the models that will be chosen in the construction stock market price.

2.2 Augmented Dickey-Fuller (ADF) Test

Testing of stationarity is vital for understanding the time series pattern. There are few stationarity tests available however, Augmented Dickey-Fuller(ADF) [7] is the most widely used. The main reason is because this test is more accessible to larger and complex time series datasets. One can identify the series pattern through the existence of unit root, using hypothesis testing. Following the assumption that time series follows an ARMA structure, null hypothesis, H_0 is tested against alternative hypothesis, H_a. Let the observations of stock market price series be $y_1, y_2, \dots, y_n$, hence the coefficients significance calculated as follows:

$$H_0 : \gamma = 0$$

$$H_a : \gamma < 0$$

$$\Delta y_t = \gamma y_{t-1} + \sum\nolimits_{j=1}^{p} (\delta_j \Delta y_{t-j}) + \varepsilon_t, \quad \varepsilon \sim (0, \sigma^2) \tag{1}$$

where Δy_t is the ADF test equation, define as the changes of the stock market price series, p is the number of lagged used in the difference term, δ_j. Meanwhile, γ is test statistic and ε_t is the error terms which is set to be serially uncorrelated. The presence of unit root shows that y is nonstationary, where the critical values is less than γ in the null hypothesis.

2.3 Transformed Technique

2.3.1 Ordinary Principal Component (OPC)

Taking into consideration that the vector of time series, $\mathbf{z}_t = (z_{1,t}, \dots, z_{m,t})$, where $1 \leq t \leq T$ and assume for simplicity that $\bar{z} = T^{-1} \sum_{t=1}^{T} \mathbf{z}_{\mathrm{t}}$, estimation of mean is valued zero and stationary. Given also $\mathbf{Z}$ is the $T \times m$ matrix, with rows $z_1, \dots, z_m$. Thus, the matrix of sample covariance, C is calculated for OPC as the following:

$$C = \frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t^{'} \tag{2}$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ is the eigenvalues of C.

OPC may reduce the dimension, however, they may be limited in representing the stock market price in construction when non-stationarity is presence.

2.3.2 Brillinger Dynamic Principal Component (BDPC)

In Brillinger [4], discussed on reconstructing the time series as follows. Given zero mean of m dimensional stationary process, z_t, $-\infty < t < \infty$, the dynamic principal components can be found for $m \times 1$ vectors c_k, $-\infty < h < \infty$ and β_j, $-\infty < j < \infty$. It is worth noting that h and j here are the dynamic principal component, respectively and is the first principal component. Hence, the linear combination becomes

$$f_t = \sum_{h=-\infty}^{\infty} c_h^{'} z_{t-h} \tag{3}$$

and subsequently

$$E\left[\left(z_t - \sum\nolimits_{j=-\infty}^{\infty} \beta_j f_{t+j}\right)^{'} \left(z_t - \sum\nolimits_{j=-\infty}^{\infty} \beta_j f_{t+j}\right)\right] \tag{4}$$

is minimum.

The principal components in cross spectral matrices is given by c_h. This is the inverse Fourier transform for each frequency. Meanwhile, from the same principal components conjugates also we can obtain the inverse Fourier transform, β_j, in which more details can be found in [4].

For both OPC and BDPC, the mean squared error (MSE) were calculated using this following formula:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}} \sum\nolimits_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{5}$$

where n is the number of sample observations, y_i is the observed values and $\hat{y}_i$ is the predicted value. It should be noted that BDPC are best with stationary series and can also work around nonstationary series. Despite that, a best minimum MSE values may be difficult to achieve.

2.4 Non-transformed Technique

2.4.1 Generalized Dynamic Principal Component (GDPC)

Supposed that $z_{j,\,t}$, $1 \leq j \leq m$, $1 \leq t \leq T$ and consider two integer numbers $k_1 \geq 0$ and $k_2 \geq 0$, as the lags and leads. Hence the first dynamic principal component as a vector $\mathbf{f} = (f_t)_{-k_1+1 \leq t \leq T+k_2}$, in which the reconstruction of series $z_{j,t}$, were linear combination of $(f_{t-k_1},\, f_{t-k_1+1},\, \ldots,\, f_t,\, f_{t+1},\, \ldots,\, f_{t+k_2})$ is optimum given the criterion of MSE. Also, $\mathbf{f}$ which is a potential factor of a m $\times$ $(k_1 + k_2)$ matrix of coefficients $\gamma = (\gamma_{j,\,i})_{1 \leq j \leq m,\, -k_1 \leq i \leq k_2}$, and $\alpha = (\alpha_1, \ldots, \alpha_m)$. Thus, initial series of $z_{j,t}$ reconstructed can be defined as

$$\hat{z}_{j,t} = \sum\nolimits_{i=-k_1}^{k_2} \gamma_{j,i} f_{t+i} + \alpha_j$$

When $k = k_1 + k_2$

$$f_t^* = f_{t-k_1}, \quad 1 \leq t \leq T + k, \quad \beta_{j,h}^* = \gamma_{j,h-k_1-1}, \quad 1 \leq h \leq k+1$$

$$\begin{aligned} f_t^{**} &= f_{t+k}^*, \qquad 1-k \leq t \leq T \\ \beta_{j,h}^{**} &= \beta_{j,k+2-h,}^{**} \quad 1 \leq h \leq k+1. \end{aligned} \tag{6}$$

This reconstruction can be achieved as

$$\hat{z}_{j,t} = \sum_{i=-k_1}^{k} \beta_{j,i} f_{t+i+k_1} + \alpha_j = \sum_{h=0}^{k} \beta_{j,h+1}^* f_{t+h}^* + \alpha_j = \sum_{h=0}^{k} \beta_{j,h+1}^{**} f_{t-h}^* + \alpha_j$$

The series reconstruction can be obtained from k lags or k leads of the principal component. Acquiring optimal forward solution will lead to backward solution as well as seen in Equation (7). Meanwhile, MSE loss function is through reconstructing the m series using k leads, by letting $\mathbf{f} = (f_1,\, \ldots,\, f_{T+k})'$, $\beta = (\beta_{j,i})_{1 \leq j \leq m, 1 \leq i \leq k+1}$ and $\alpha = (\alpha_1, \ldots, \alpha_m)$,

$$\mathrm{MSE}\left(\mathbf{f}, \beta, \alpha\right) = \frac{1}{Tm} \sum_{j=1}^{m} \sum_{t=1}^{T} \left(z_{j,t} - \sum_{i=0}^{k} \beta_{j,i+1} f_{t+i} - \alpha_j \right)^2 \tag{7}$$

The optimal options of $\mathbf{f} = (f_1, \ldots, f_{T+k})'$ and $\beta = (\beta_{j,i})_{1\leq j\leq m, 1\leq i\leq k+1}$, $\alpha = (\alpha_1, \ldots, \alpha_m)$, are defined by

$$\left(\hat{f}, \hat{\beta}, \hat{\alpha}\right) = \arg_{f\in R^{T+k}, \beta\in R^{m\times(k+1)}, \alpha R^m} \min \mathrm{MSE}(\mathbf{f}, \beta, \alpha) \tag{8}$$

It should be noted that if $\mathbf{f}$ is optimal, clearly $\gamma\mathbf{f} + \delta$ is optimal too. Hence, $\mathbf{f}$ is chosen in order that $\sum_{t=1}^{T+k} f_t = 0$ and $(1/(T+k))\sum_{t=1}^{T+k} f_t^2 = 1$. From $\mathbf{z}_1, \ldots, \mathbf{z}_t$ observations, we can obtained the first GDPC of order k, given by $\hat{f}$. Meanwhile, GDPC of order 0 represents the first regular principal component

When $\mathbf{C}_j(\alpha_j) = (c_{j,t,q}(\alpha_j))_{1\leq t\leq T+k, 1\leq q\leq k+1}$ the $(T+k) \times (k+1)$ matrix can be shown as

$$\mathrm{c}_{j,t,q}(\alpha_j) = \begin{cases} (z_{j,t-q+1} - \alpha_j), & 1 \vee (t-T+1) \leq q \leq (k+1) \wedge t \\ 0 \quad , & \text{otherwise} \end{cases} \tag{9}$$

such as $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. Then, $\mathbf{D}_j(\beta_j) = (d_{t,j,q}(\beta_j))$ is $(T+k) \times (T+k)$ becomes

$$d_{t,j,q}(\beta_j) = \sum_{v=(t-k)\vee 1}^{t\wedge T} \beta_{j,q-v+1}\beta_{j,t-v+1}$$

if $(t-k) \vee 1 \leq q \leq (t+k) \wedge (T+k)$ and 0 otherwise. Let

$$\mathbf{D}(\beta) = \sum_{j=1}^{m} \mathbf{D}_j(\beta_j) \tag{10}$$

Through f_t differentiation in Equation (8), therefore

$$\mathrm{f} = \mathbf{D}(\beta)^{-1} \sum_{j=1}^{m} C_j(\alpha)(\beta_j). \tag{11}$$

As well as using least-square estimators, β_j and α_j coefficients with $1 \leq j \leq m$ are as follows

$$\begin{pmatrix} \beta_j \\ \alpha_j \end{pmatrix} = \left(\mathbf{F}(\mathrm{f})' \mathbf{F}(\mathrm{f})\right)^{-1} \mathbf{F}(\mathrm{f})' \mathbf{z}^{(\mathrm{j})} \tag{12}$$

where $z^{(j)} = (z_{j,1}, \ldots, z_{j,T})'$ and $\mathbf{F}(\mathbf{f})$ is the $T \times (k+2)$ matrix with t^{th} row $(f_t, f_{t+1}, \ldots, f_{t+k}, 1)$. Finally Equation (11) and Equation (12) define the first GDPC.

2.5 Information Criterion Method

The deviance must be defined prior to describing the methods for model comparison [8]. This deviance is also known as log-likelihood (ratio) statistics [9]. Therefore, the two most widely used log-likelihood model are Akaike Information Criteria (AIC) [10] and Bayesian Information Criteria (BIC) [11]. In order to choose model formulation that have the smallest value, it can be determine using information criteria methods. AIC and BIC equations are following the standard; hence, they are not specific to OPC, BDPC and GDPC. Therefore, the equations that were used to calculate AIC and BIC in all three principal component techniques as in Equation (13):

$$\begin{aligned} AIC &= -2\log L + 2w \\ BIC &= -2\log L + \log(n)\cdot w \end{aligned} \tag{13}$$

where w is the number of estimated parameters in the model, L, is the maximum values of the likelihood function for the model and n is the number of observations.

3 Result Analysis

The construction stock market price consists of nine listed companies in this sector, obtained from seven countries. This stock market prices selected were 1800.HK (Hong Kong), 3336.KL and 4677.KL (Malaysia), FBU.NZ (New Zealand), 600170.SS and 600266.SS (Shanghai), IMPN.SW (Switzerland), CCP.BK (Thailand) and STRL (U.S.). It is worth noting that the observations were weekly-based of a two-year period from 1st January 2016 until 1st January 2018. The construction stock market price weekly plot showed in Figure 1. The stationarity result test is in Table 1. Table 2 shows the AIC and BIC for the construction stock market price. Meanwhile, the MSE and percentage of explained variance of both transformed and non-transformed technique is shown in Table 3.

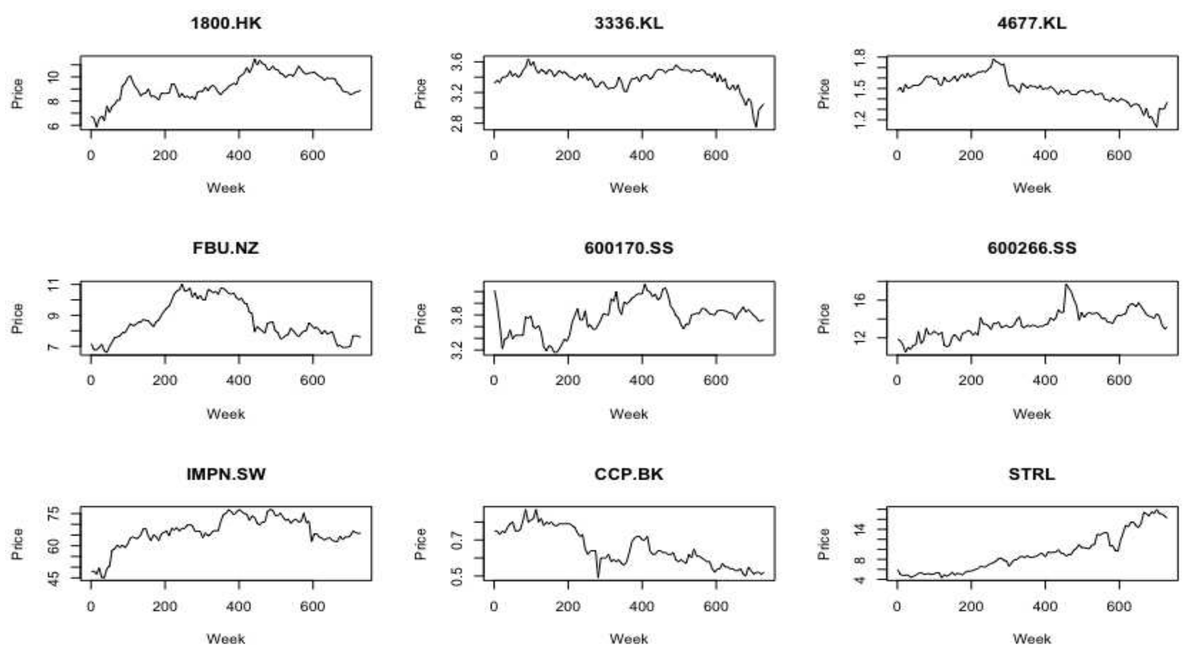

Figure 1: Nine Weekly Construction Stock Market Prices between 1^{st} January 2016 to 1^{st} January 2018

The context of this study is focused only on non-stationarity. Therefore, the non-stationarity pattern was verified using the ADF test. It was found out that the test statistics values were higher than the 95% critical value in all nine constructions stock market price as shown in Table 1. It is worth noting that the 95% critical value is at -1.95. This also shows that all of the nine constructions stock market price fails to reject the H_0 indicating the presence of unit root. Therefore, the series can be concluded as non-stationary.

Table 1: Stationarity Result using ADF Test

Construction stock market price	Test statistics
1800.HK	0.4247
3336.KL	-0.6060
4677.KL	-0.5232
FBU.NZ	0.0819
600170.SS	-0.3594
600266.SS	0.0487
IMPN.SW	0.6438
CCP.BK	-1.0534
STRL	1.4018

Table 2: AIC and BIC at Original Stock Market Price and Lags $k = 3$ Model.

Construction stock market price	AIC	BIC
1800.HK	971.28	981.74
3336.KL	969.62	980.08
4677.KL	975.07	985.53
FBU.NZ	970.56	981.02
600170.SS	974.64	985.10
600266.SS	974.55	985.02
IMPN.SW	970.96	981.42
CCP.BK	975.75	986.21
STRL	974.78	985.24

Next, the models were tested through AIC and BIC model assessment, whereby those that have smaller deviance indicates a better fit to the data. It should be noted that a number of models were tested using different lags. Also, comparison of the models was made using both AIC and BIC. Hence, the models that were chosen throughout this consist of the original value and at lags $k = 3$ as shown in Table 2. From Table 2, it can be observed that AIC had smaller deviance values in its nine constructions stock market price as compared to BIC.

Meanwhile, as can be observed in Table 3, the MSE for the non-transformed GDPC had the lowest values in its eight constructions stock market price. This is followed by transformed BDPC, where eight of its stock market prices were lower than OPC. Nevertheless, transformed OPC had the highest MSE among the three methods. Hence, this shows that when non-stationarity is presence, the non-transformed technique of GDPC has better performance model as compared to the transformed techniques.

As can be observed in Table 3, eight constructions stock market price in the non-transformed

Table 3: Mean Squared Error and The Percentage of Explained Variance of Transformed and Non-transformed Method

Construction stock market price	**MSE**			**Percentage of Explained Variance**		
	OPC	**BDPC**	**GDPC**	**OPC**	**BDPC**	**GDPC**
1800.HK	13.042	6.240	0.082	57.1	83.3	93.4
3336.KL	1.135	1.714	0.002	59.9	87.2	89.6
4677.KL	10.962	1.378	0.001	57.5	67.7	94.7
FBU.NZ	11.387	4.091	0.066	50.9	82.4	95.8
600170.SS	7.180	3.556	0.011	59.3	94.7	85.4
600266.SS	11.420	1.786	0.211	50.3	76.6	88.9
IMPN.SW	6.520	2.573	2.914	52.1	79.1	93.9
CCP.BK	20.846	5.835	0.001	53.1	77.7	94.9
STRL	42.643	26.556	0.329	52.3	92.4	97.7

GDPC had the highest percentage of explained variance as compared to the other two methods. It should be highlight here that this values shown here is only for the first component, as for demonstration [6]. In the transformed technique of BDPC, only 600170.SS and STRL construction stock market price showed higher percentage of explained variance of 90% above, whereas the remaining are less than 90%. As for the transformed OPC, the average of explained variance percentage is about 50% only. This can be seen in the comparison of FBU.NZ in transformed OPC which had almost a gap of 50% of percentage of explained variance as compared to in the non-transformed GDPC. Hence, from Table 3, the comparison between transformed and non-transformed technique had showed that much more information can be gain from the non-transformed technique.

4 Conclusion

In conclusion, the non-transformed technique of GDPC has the capability of tackling the non-stationarity which are present in the stock market price. This is achieved through direct application of the series using principal component technique. Furthermore, the non-transformed technique of GDPC had shown that it had better performance than of the transformed techniques of OPC and BDPC. Furthermore, the ability of giving higher percentage of explained variance in the first component of the construction stock market price showed that much information can be gain only in the first component. As for the construction sector benefit, it is hope that the findings will provide a baseline for modelling and prediction of the stock market price in their respected field. Thus, the subject of future study may be extended to other sectors in the stock market price.

Acknowledgement

The first author would like to thank Universiti Malaysia Kelantan and the Ministry of Education, Malaysia for a study leave in UTM under SLAI-KPT scholarship. The authors also would like to express their appreciation for the support of the sponsors with Project No 01M30.

References

[1] P. Chatvorawit, P. Sattayatham, and B. Premanode. Improving stock price prediction with SVM by simple transformation: The sample of stock exchange of Thailand (SET). *Thai J. Math*. 2016. 14(3): 553–563.

[2] M. Á. Sánchez, J. E. Trinidad, J. García, and M. Fernández. The effect of the underlying distribution in hurst exponent estimation. *PLoS One*. 2015. 10(5): 1–17.

[3] P. Caraiani. Evidence of multifractality from emerging European stock markets. *PLoS One*. 2012. 7(7): 1–9.

[4] D. Peña and V. J. Yohai. Generalized dynamic principal components. *J. Am. Stat. Assoc*. 2016. 111(515): 1121–1131.

[5] D. R. Brillinger. *Time Series: Data Analysis and Theory*. SIAM. 1981.

[6] Y. Andu, M. H. Lee, and Z. Y. Algamal. Generalized dynamic principal component for monthly nonstationary stock market price in technology sector. *J. Phys. Conf. Ser*. 2018. 1132: 012076.

[7] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc*. 1979. 74(366a): 427–431.

[8] J. A. Nelder and R. W. Wedderburn. Generalized linear models. *J. R. Stat. Soc. Ser. A*. 1972. 3(135): 370–384.

[9] A. J. Dobson and Barnett A.G. *An Introduction to Generalized Linear Models*. 3rd Edition Chapman & Hall/CRC, Boca Raton. 2008.

[10] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Contr*. 1974. 6(19): 716–723.

[11] H. Akaike. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*. 1979. 2(66): 237–242.