International Journal of Research in Business, Economics and Management

Vol.3 Issue 3 May-June 2019

www.ijrbem.com

# A Method Adjusting Time Series Data to Improve Quality of Forecast Results of Consumer Price Index in Vietnam

**Nguyen Van Giap[1], Nguyen Thi Kim Ngan[2], Ly Thu Trang[1]**

[1]University of Information and communication technology – Thai Nguyen University - Vietnam

[2]Thai Nguyen ethnic boarding vocational school – Vietnam

## ABSTRACT

In this paper, we propose a method to process the time series to improve the quality of the original time series to forecast. We conducted experimentation with forecasting Consumer Price Index (CPI) of Vietnam for the first six months of the year in 2019 with time series dataset from 1/2007 to 12/2018. We used the Augmented Dickey-Fuller (ADF) test to test non-stationarity of time series and ARIMA model to predict CPI values. In this research, we propose a method to adjust the time series data by punding outliers with value – $\tau$ through processing conduction. The threshold is based on $\tau$ value ($\varepsilon$). We predicted CPI of Vietnam for the first six months of 2019 with three situations: Original time series – The first situation; Time series with $\varepsilon$=1.5 is the second situation and time series with $\varepsilon$=0.5 is the third situation. The result of this research will suggest a method for pre-processing time series before using it. This method improves our accuracy forecast values significantly. We found that choosing threshold optimation directly affects our quality prediction. The adjusted time series by $\varepsilon$ higher will gain better forecast results. The adjusted time series by $\varepsilon$ higher will gain better forecast results. The value should be less than two and more than one that is suitable for CPI monthly situation in Vietnam.

**Keywords:** *Time series, outliers, analysis, analyzing data, processing time series, CPI forecast, prediction.*

## INTRODUCTION

Vietnam is a developing economy in Southeast Asia. According to the World Bank's report in the last 3 years, Vietnam's Annual GDP Growth (%) is increasing with a high level in the world: 6.2 (2016); 6.8 (2017); 6.8 (2018). In the coming time, Vietnam's economy will still grow up. The forecast of the World Bank, the value will be 6.6 (2019) and 6.5 (2020). According to the organization's report, Vietnam's annual inflation cut down fast: 18.7% (2011); 9.1% (2012); 6.6 (2013); 4.7 (2014); and 0.9 (2015). One of the important factors which strongly affect the growing is CPI. Researching the monthly CPI time series in Vietnam is important and popular. The macroeconomic and macroeconomic studies all use this index for the analysis report. Forecast of the consumer price index is always concerned by researchers because of its important role in the economy in the near future. The accuracy of the consumer price index forecast will directly affect the quality of research results of micro and macroeconomics researchers. According to this, our research will focus on research about CPI time series and answer two below questions:

- Does the replacement of outliers by the $\tau$ value in the time series make the prediction accuracy better?

- How is the value of the error (ε) set?

## LITERATURE REVIEW

Recent relevant researches on CPI have been conducted with some different methods. The research results about CPI of the UK used daily means of log-price changes to adjust its forecasts (Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J., & Winton, J. (2018)). In the US, researchers proposed to use MIDAS models to a significant upgrade in the forecast accuracy of official consumer price inflation (Harchaoui, T. M., & Janssen, R. V. (2018)). In China, they used an integrated (ARIMA)–backpropagation (BP) algorithm for consumer price index (CPI) forecasting (Wang, J., Sun, L., Zhao, H., & Wang, Y. (2016)). The other research found that good forecasts can be made with simple rules of thumb that are extrapolations of a single data series (Faisal, F. (2012)). The careful studies of Siregar, R., and Nguyen, T. (2013), used the MS-VAR approach to know the relationship between inflation and gold prices in Vietnam. These studies have not mentioned the harmful effects of outliers on our research results even though they have pruned with the different techniques. The normal way is used Sliding Average Method to gain the time series. But the received time series will face too much influenced by the weight of the outliers. We will propose a method to handle a good time series before applying the ARIMA model to analyze the next steps. We conduct to adjust the time series with a new value that is called τ. The detail method is shown below part.

## III. Proposal method

### *a. Set up time series*

We set up amount value of time series Yt ={yij} with i=1 to n; j=1 to m and t=nm; m={3,6,9,12}. We propose one of the values of m={3,6,9,12} these are suitable for the period of time per year. If we set m=3 we will gain three months per group and fours group data per year. As the same, we also choose m=6, 9, or 12 that depends on each economy.

### *b. τ value*

We propose τ_ij that is a value for adjusting outliers. The value is shown below:

$$\tau_{ij} = \frac{|y_{ij} - \gamma_i|}{\gamma_i} \times 100\% \qquad (3.1)$$

i= 1 to n; j =1 to m

while $\gamma_i$ is given by (3.2) i=1 to n

$$\gamma_i = \frac{\sum_{j=1}^{m} y_{ij}}{m} \qquad (3.2)$$

### *c. Collected value of time series method*

We propose ε value to prund the outliers. It is determined the threshold of τ to gather values for the time series. Our experiments show that ε should be chosen depending on the real time series. We proposal $\varepsilon \in [0,2]$.

The new time series is gained to be $\hat{Y} = \{\hat{y}_{ij}\}$ with $i = \overline{1,n}\ and\ j = \overline{1,m}$

each $\hat{y}_{ij}$ is given by

$$\hat{y}_{ij} = y_{ij}\ if\ \tau_{ij} < \varepsilon \quad (3.3)$$

or

$$\hat{y}_{ij} = \gamma_i\ if\ \tau_{ij} \geq \varepsilon \quad (3.4)$$

### *d. Proposal processing*

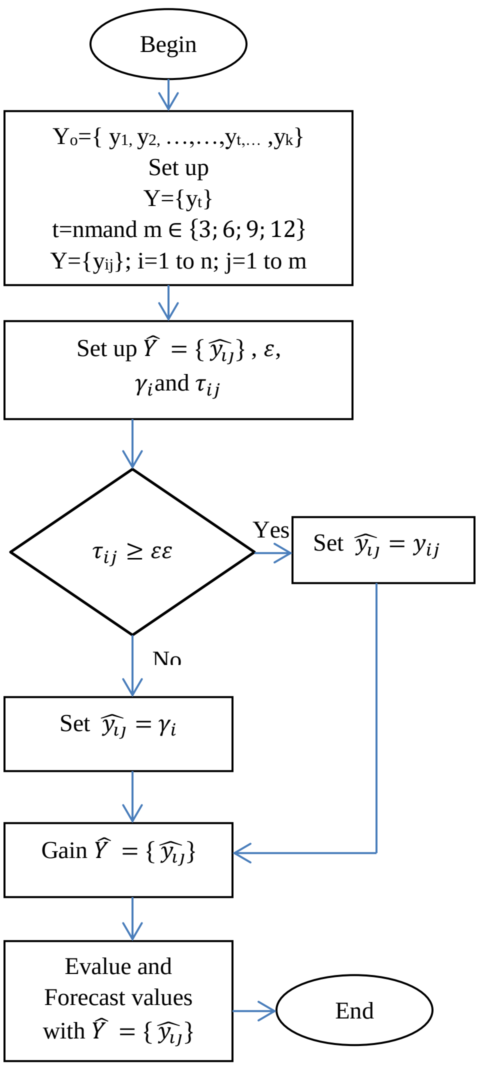

*Figure 1: The poposal processing*

***Step 1:*** Setting up time series {y} and prunding the time series to {$y_t$} with t=nm. It is important to choose the value of m which is the numbers of the analyzed period. The period should be 3, 6, 9, or 12. Because of the fact of the case countries, for example in Vietnam, there are four distinct seasons with every 3 months. The consumption demand for these four seasons in Vietnam has a big difference. In this period, the consumer price index has quite a similarity.

***Step 2:*** We have the original time series. We set the value $\varepsilon$, $\gamma_i$ and calculate $\tau_{ij}$. The value of $\varepsilon$ is the threshold of permissible error and the value of $\tau_{ij}$ is the calculated values.

***Step 3:*** Replacing the value of original $\widehat{y_{ij}}$:

- if $\tau_{ij} \geq \varepsilon$ then $\widehat{y_{ij}} = y_{ij}$ (3.5)

- else $\widehat{y_{ij}} = \gamma_i$ (3.6)

***Step 4:*** We gain new time series $\widehat{Y} = \{\widehat{y_{ij}}\}$ and use the time series to analyze and predict.

**III. Experimentation**

In this research, we collected the CPI data in Vietnam during the period from January 2007 to January 2019. The time series is collected from the General Statistics Office and the State Bank of Vietnam. This time series of data was collected monthly method and we used Eviews 8.0 software to analyze. We gathered 144 samples from 1/2007 to 12/2018 in Table 1. We experiment with $\varepsilon = 1.5$ Table 2 and $\varepsilon = 0.5$ Table 3.

*Table 1: CPI time series of Vietnam (%) (1/2007 to 12/2018)*

| month | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1.02 | 2.38 | 1.9 | 1.36 | 1.73 | 1 | 1.25 | 0.69 | -0.2 | 0 | 0.46 | 0.51 |
| **2** | 2.2 | 3.61 | 0.22 | 1.96 | 2.09 | 1.37 | 1.32 | 0.55 | -0.1 | 0.42 | 0.23 | 0.73 |
| **3** | -0.3 | 2.94 | -0.15 | 0.75 | 2.17 | 0.16 | -0.2 | -0.4 | 0.15 | 0.57 | 0 | -0.27 |
| **4** | 0.54 | 2.26 | -0.3 | 0.14 | 3.32 | 0.06 | 0.02 | 0.08 | 0.14 | 0.33 | 0.21 | 0.08 |
| **5** | 0.71 | 3.9 | 0.44 | 0.27 | 2.21 | 0.18 | -0.1 | 0.2 | 0.16 | 0.54 | -0.53 | 0.55 |
| **6** | 0.89 | 2.12 | 0.55 | 0.22 | 1.09 | -0.3 | 0.05 | 0.3 | 0.35 | 0.46 | -0.17 | 0.61 |
| **7** | 0.97 | 1.11 | 0.52 | 0.06 | 1.17 | -0.3 | 0.03 | 0.23 | 0.13 | 0.13 | 0.11 | -0.09 |
| **8** | 0.52 | 1.58 | 0.24 | 0.23 | 0.93 | 0.63 | 0.08 | 0.12 | -0.1 | 0.1 | 0.92 | 0.45 |
| **9** | 0.52 | 0.2 | 0.62 | 1.31 | 0.82 | 2.2 | 1.06 | 0.4 | -0.2 | 0.54 | 0.59 | 0.59 |
| **10** | 0.78 | -0.2 | 0.37 | 1.05 | 0.36 | 0.85 | 0.49 | 0.11 | 0.11 | 0.51 | 0.41 | 0.33 |
| **11** | 1.2 | -0.7 | 0.55 | 1.86 | 0.39 | 0.47 | 0.34 | 0.27 | 0.07 | 0.48 | 0.13 | - |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | 0.29 |
| **12** | 2.87 | -0.7 | 1.38 | 1.99 | 0.53 | 0.27 | 0.51 | -0.2 | 0.02 | 0.23 | 0.21 | -0.25 |

*Table 2: CPI Adjusted time series* ($\varepsilon = 1.5$)

| month | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1.02 | 2.38 | 1.9 | 1.36 | 1.73 | 1 | 1.25 | 0.69 | -0.2 | 0 | 0.46 | 0.51 |
| **2** | 2.2 | 1.39 | 0.22 | 1.96 | 2.09 | 1.37 | 1.32 | 0.55 | -0.1 | 0.42 | 0.23 | 0.73 |
| **3** | -0.3 | 0.7 | -0.15 | 0.75 | 0.7 | 0.16 | -0.2 | 0.7 | 0.15 | 0.57 | 0 | -0.27 |
| **4** | 0.54 | 2.26 | -0.3 | 0.14 | 0.91 | 0.06 | 0.02 | 0.08 | 0.14 | 0.33 | 0.21 | 0.08 |
| **5** | 0.71 | 1.13 | 0.44 | 0.27 | 2.21 | 0.18 | -0.1 | 0.2 | 0.16 | 0.54 | -0.53 | 0.55 |
| **6** | 0.89 | 2.12 | 0.55 | 0.22 | 1.09 | -0.3 | 0.05 | 0.3 | 0.35 | 0.46 | -0.17 | 0.61 |
| **7** | 0.97 | 1.11 | 0.52 | 0.06 | 1.17 | -0.3 | 0.03 | 0.23 | 0.13 | 0.13 | 0.11 | -0.09 |
| **8** | 0.52 | 1.58 | 0.24 | 0.23 | 0.93 | 0.63 | 0.08 | 0.12 | -0.1 | 0.1 | 0.92 | 0.45 |
| **9** | 0.52 | 0.2 | 0.62 | 1.31 | 0.82 | 2.2 | 1.06 | 0.4 | -0.2 | 0.54 | 0.59 | 0.59 |
| **10** | 0.78 | -0.2 | 0.37 | 1.05 | 0.36 | 0.85 | 0.49 | 0.11 | 0.11 | 0.51 | 0.41 | 0.33 |
| **11** | 1.2 | 1.31 | 0.55 | 1.86 | 0.39 | 0.47 | 0.34 | 0.27 | 0.07 | 0.48 | 0.13 | -0.29 |
| **12** | 2.87 | -0.7 | 1.38 | 1.99 | 0.53 | 0.27 | 0.51 | -0.2 | 0.02 | 0.23 | 0.21 | -0.25 |

*Table 3: CPI Adjused time series* ($\varepsilon = 0.5$)

| month | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1.02 | 1.09 | 1.09 | 1.36 | 1.09 | 1 | 1.25 | 0.69 | 1.09 | 1.09 | 1.09 | 1.09 |
| **2** | 1.39 | 1.39 | 1.39 | 1.96 | 1.39 | 1.37 | 1.32 | 1.39 | 1.39 | 1.39 | 1.39 | 0.73 |
| **3** | 0.7 | 0.7 | 0.7 | 0.75 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.57 | 0.7 | 0.7 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **4** | 0.54 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 |
| **5** | 0.71 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 | 1.13 |
| **6** | 0.89 | 1.02 | 0.55 | 1.02 | 1.09 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 0.61 |
| **7** | 0.97 | 1.11 | 0.52 | 0.92 | 1.17 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 |
| **8** | 1.14 | 1.58 | 1.14 | 1.14 | 0.93 | 0.63 | 1.14 | 1.14 | 1.14 | 1.14 | 0.92 | 1.14 |
| **9** | 1.47 | 1.47 | 1.47 | 1.31 | 0.82 | 2.2 | 1.06 | 1.47 | 1.47 | 1.47 | 1.47 | 1.47 |
| **10** | 0.78 | 1.26 | 1.26 | 1.05 | 1.26 | 0.85 | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 | 1.26 |
| **11** | 1.2 | 1.31 | 1.31 | 1.86 | 1.31 | 1.31 | 1.31 | 1.31 | 1.31 | 1.31 | 1.31 | 1.31 |
| **12** | 1.57 | 1.57 | 1.38 | 1.99 | 1.57 | 1.57 | 1.57 | 1.57 | 1.57 | 1.57 | 1.57 | 1.57 |

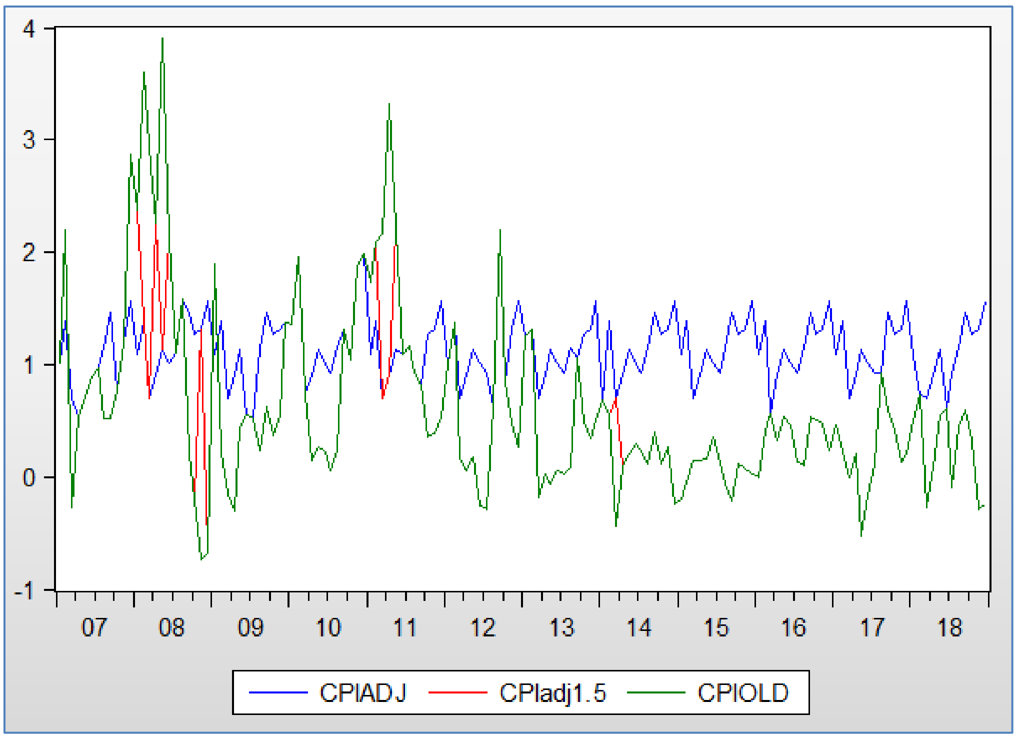


*Figure 2: The graph of 3 situation experiments (The Blue is $\varepsilon = 0.5$; The Red is $\varepsilon = 1.5$ and The Green is original data)*

Figure 2 shows the lines of three situation time series data: The Green line is the original data – the first situation; The Red line is shown the second situation and the Blue is shown the third situation data.

Figure 3 shows the results of a non-stationary time series test based on ADF test value against the critical Test values. According to the method, the time series is stationary when the value of the ADF test is greater than the absolute value of critical Test values at significant levels. Figure (a), (b) (Pro$^*$=0.000<<0.05) satisfy the condition: 5.06 is more than all 3.48; 2.88; and 2.57, 5.19 is more than 3.47; 2.88; and 2.57. Figure (c) does not satisfy the stationary time series: 3.07 is less than 3.48. These means the original time series and the time series is obtained with $\varepsilon$ =1.5 are stationary time series while the other is gain by $\varepsilon$ =0.5 is non-stationary time series. Adjusting non-stationary time series becomes stationary time series by the $1^{sth}$ difference which result is shown at (d) with Prob=0.000<<0.05.

Applying the ARIMA method we have results from the three-time series as follows:

***a. Original time series (The first situation):***

Models are set:

According the probility of the original time series and

LS dcpiold c Ar(1) ar(2) ar(6) ar(9) ar(29) ma(1) ma(5) ma(6) ma(9) ma(14)

LS dcpiold c ar(2) ar(29) ma(6)

LS dcpiold c ma(6)

***b. The second time series ($\boldsymbol{\varepsilon = 1.5}$) (The second situation)***

Models are set :

LS dcpiadj1_5 c ar(1)ar(2)ar(3)ar(4)ar(6)ar(10) ma(1)

LS dcpiadj1_5 c ar(6)ar(10) ma(1)

LS dcpiadj1_5 c ar(10) ma(1)

c. The third time series ($\boldsymbol{\varepsilon = 0.5}$) (The third situation)

Models are set::

LS dcpiadj c ar(1)ar(2)ar(6)ar(9)ar(11)ar(15)ar(23)ma(1)ma(11)ma(12)ma(13)

ma(23)ma(24)ma(25)ma(30)ma(35)ma(36)

LS dcpiadj c ar(1)ar(2)ar(23) ma(11)ma(13)ma(30)ma(35)ma(36)

LS dcpiadj c ar(1)ar(2)ar(23) ma(36)

Null Hypothesis: CPIOLD has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=13)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -5.066247 | 0.0000 |
| Test critical values: | 1% level | -3.476472 | |
| | 5% level | -2.881685 | |
| | 10% level | -2.577591 | |

(a)

Null Hypothesis: CPIADJ1_5 has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=13)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -5.191635 | 0.0000 |
| Test critical values: | 1% level | -3.476805 | |
| | 5% level | -2.881830 | |
| | 10% level | -2.577668 | |

(b)

Null Hypothesis: CPIADJ has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic - based on SIC, maxlag=13)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -3.077032 | 0.0307 |
| Test critical values: | 1% level | -3.480425 | |
| | 5% level | -2.883408 | |
| | 10% level | -2.578510 | |

(c)

Null Hypothesis: D(CPIADJ) has a unit root
Exogenous: Constant
Lag Length: 10 (Automatic - based on SIC, maxlag=13)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -12.44842 | 0.0000 |
| Test critical values: | 1% level | -3.480425 | |
| | 5% level | -2.883408 | |
| | 10% level | -2.578510 | |

(d)

*Figure 3: Test stationary time series with ADF model: (a) Original time series; (b) adjusted time series with $\varepsilon = 1.5$; (c) $\varepsilon = 0.5$; (d) adjusted time series ($\varepsilon = 0.5$) to become stationary time series*

The final models (3rd Models) are selected to ensure statistical significance with Pro * = 0.000 << 0.05 (Figure 4) and they are set by stationary time series (DCPIOLD; DCPIADJ1_5; DCPIADJ) which are tested by ADF model test, above (Figure 3).

The final model for the first situation: LS dcpiold c ma(6)

The final model for the second situation: LS dcpiadj1_5 c ar(10) ma(1)

The final model for the third situation: LS dcpiadj c ar(1)ar(2)ar(23) ma(36)

Dependent Variable: DCPIOLD
Method: Least Squares
Date: 04/26/19 Time: 00:45
Sample (adjusted): 2018M06 2018M12
Included observations: 7 after adjustments
Convergence achieved after 10 iterations
MA Backcast: 2017M12 2018M05

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.040002 | 1.60E-05 | 2506.669 | 0.0000 |
| MA(6) | 0.999985 | 8.29E-10 | 1.21E+09 | 0.0000 |

(a)

Dependent Variable: DCPIADJ
Method: Least Squares
Date: 04/26/19 Time: 00:19
Sample (adjusted): 2009M01 2018M12
Included observations: 120 after adjustments
Convergence achieved after 8 iterations
MA Backcast: 2006M01 2008M12

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -0.000377 | 0.016580 | -0.022724 | 0.9819 |
| AR(1) | -0.646206 | 0.088556 | -7.297163 | 0.0000 |
| AR(2) | -0.256134 | 0.087051 | -2.942356 | 0.0039 |
| AR(23) | -0.196942 | 0.068697 | -2.866824 | 0.0049 |
| MA(36) | 0.883256 | 0.020055 | 44.04149 | 0.0000 |

(b)

Dependent Variable: DCPIADJ1_5
Method: Least Squares
Date: 04/25/19 Time: 23:29
Sample (adjusted): 2007M12 2018M12
Included observations: 133 after adjustments
Convergence achieved after 10 iterations
MA Backcast: 2007M11

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -0.012655 | 0.012227 | -1.035059 | 0.3026 |
| AR(10) | -0.170987 | 0.082097 | -2.082750 | 0.0392 |
| MA(1) | -0.714502 | 0.062235 | -11.48072 | 0.0000 |

(c)

*Figure 4: Test stationary time series with ADF model: (a) Original time series; (b) adjusted time series with $\varepsilon = 1.5$; (c) $\varepsilon = 0.5$; (d) adjusted time series ($\varepsilon = 0.5$) to become stationary time series*

## IV. Results and Discussion

We conduct to evaluate the accuracy of forecast data through comparison them with actual values. We propose comparative values: ss, SSE, and AVSSE. They are shown in the formulas below. If the value of AVSSE is low, it means good accuracy of time series which was used to predict. Conversely, if the AVSSE value is higher, it means that the accuracy is not good accuracy of the time series.

$$ss = CPI_f - CPI_g \quad (4.1)$$

$CPI_f$ the forecast value of CPI; $CPI_g$ the original value of CPI

$$SSE = \sum_{i=1}^{n} ss_i^2 \quad (4.2)$$

$ss_i$ is ss of month i; i=1 to n

n is the number of sample experimentation

$$AVSSE = \frac{SSE}{n} \quad (4.3)$$

We use experimental data about CPI of Vietnam from 6/2018 to 12/2018 to assess the relevance of the case studies. They have gained by the three best models from the three case studies. Results for Tables 4 and Figure 5 below:

*Table 4: Forecast data from three situation experiments*

| **Month** | **Time series** | | | |
|---|---|---|---|---|
| | **Org** | **Old** | **1_5** | **0_5** |
| 2018M6 | 0.61 | 0.06 | -0.46 | -0.35 |
| 2018M7 | -0.09 | -0.70 | 0.04 | 0.10 |
| 2018M8 | 0.45 | 0.54 | 0.02 | 0.33 |
| 2018M9 | 0.59 | 0.14 | 0.03 | -0.10 |
| 2018M10 | 0.33 | -0.26 | -0.03 | 0.14 |
| 2018M11 | -0.29 | -0.62 | -0.07 | -0.07 |
| 2018M12 | -0.25 | 0.04 | -0.05 | 0.17 |

*Table 5: Comparative values for the period 6/2018 – 12/2018*

| **Old** | | **1_5** | | **0_5** | |
|---|---|---|---|---|---|
| **ss1** | **ss1*ss1** | **ss2** | **ss2*ss2** | **ss3** | **ss3*ss3** |
| -0.90 | 0.81 | -1.75 | 3.07 | -1.58 | 2.48 |
| 6.78 | 45.94 | -1.46 | 2.14 | -2.07 | 4.29 |
| 0.20 | 0.04 | -0.96 | 0.93 | -0.26 | 0.07 |
| -0.76 | 0.58 | -0.94 | 0.89 | -1.16 | 1.35 |
| -1.79 | 3.20 | -1.09 | 1.18 | -0.57 | 0.33 |
| 1.14 | 1.29 | -0.77 | 0.60 | -0.75 | 0.57 |
| -1.16 | 1.35 | -0.79 | 0.62 | -1.67 | 2.80 |
| 0.88 | 0.78 | 0.56 | 0.32 | -3.17 | 10.07 |
| SSE | 53.99 | | 9.75 | | 21.96 |
| **AVSSE** | **0.92** | | **0.39** | | **0.59** |

The evaluation results are shown in Table 5. and Figure 5. below (see the period 6/2018-12/2018: The Ord is real values; Old is the line of predictive values with the original time series, 1_5 is the line of predicted values with gained time series by $\varepsilon$ =1.5, and 0_5 is the line of predicted values with gained time series by =0.5). According to these results, the second model with $\varepsilon$ = 1.5 gives the best results with AVSSE = 0.39; following by it is the third model with $\varepsilon$ = 0.5 (AVSSE = 0.59); and finally, the Old model (AVSSE = 0.92).

In the final step, the CPI values for the first six months of 2019 are forecasted and shown in tables and graphs (see the period 1/2019-6/2019), below:

*Table 6: Final forecast data from three situation experiments*

| Month | Time series | | | |
|---|---|---|---|---|
| | **Org** | **Old** | **1_5** | **0_5** |
| 2019M1 | N/A | **0.19** | **0.16** | **-0.22** |
| 2019M2 | N/A | **-0.08** | **-0.07** | **0.33** |
| 2019M3 | N/A | **-0.14** | **-0.10** | **-0.46** |
| 2019M4 | N/A | **-0.03** | **0.06** | **0.11** |
| 2019M5 | N/A | **-0.01** | **-0.02** | **0.11** |
| 2019M6 | N/A | **-0.05** | **-0.02** | **-0.05** |

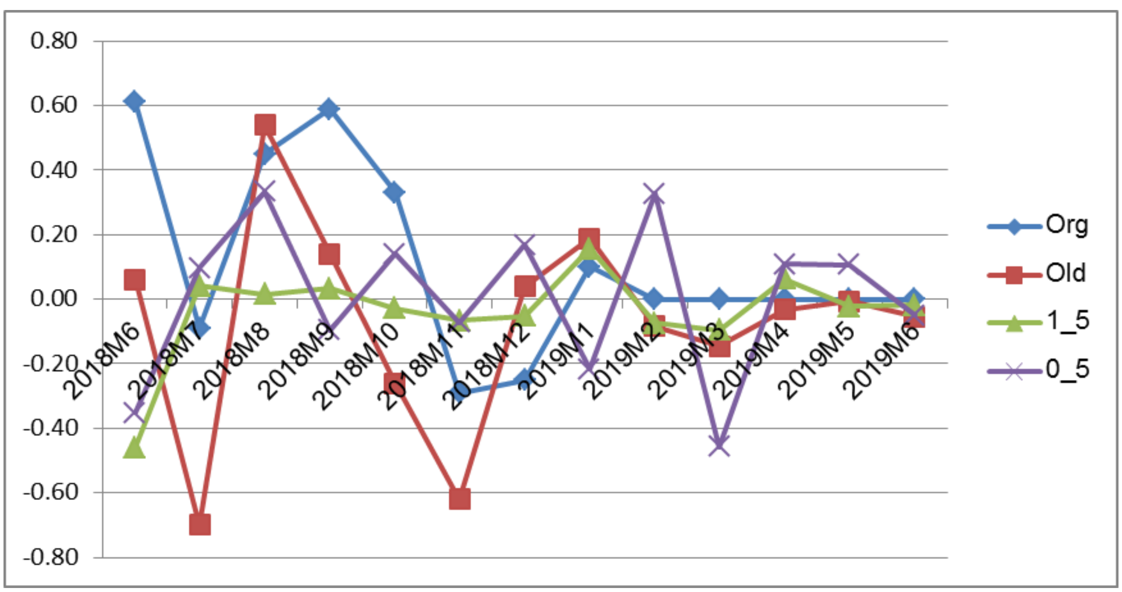

*Figure 5: Test stationary time series with ADF model: (a) Original time series; (b) adjusted time series with* $\varepsilon = 1.5$*; (c)* $\varepsilon = 0.5$*; (d) adjusted time series (*$\varepsilon = 0.5$*) to become stationary time series*

**IV. Conclution**

We have proposed a method of adjusting the time series for Vietnam's CPI from 1/2007 to 12/2018 with t value and error $\varepsilon$. We find a significant improvement in forecasting results when using this method (for answering the first research question). The initial selection of $\varepsilon$ value is important and directly affects the quality of the forecast results. In the cases of the study of Vietnam CPI, we find that the value of $\varepsilon = 1.5$, which will be based on gaining a time series, is better than $\varepsilon = 0.5$ (to answer to the second research question). However, the study of affected factors to choose a $\varepsilon$ value should be studied in the future to improve the forecasted results. We also find an issue when using the ARIMA forecast model when forecasting farther values. The accuracy of the values will be reduced. In the future, we will continuously focus on research on these issues which will improve our time series.

**Acknowledgment**

This research is supported by University Information and Communication Technology (ICTU) – Thai Nguyen University under project number T2018 – 07 – 22.

**References**

Das, S. (1994). *Time series analysis*. Princeton University Press, Princeton, NJ.

Wei, W. W. (2006). Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Trang, P. T. T. (2009). Các yếu tố tác động tới lạm phát tại Việt Nam–Phân tích chuỗi thời gian phi tuyến. *Tạp chí Kinh tế và Dự báo*.

Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). *Introduction to time series and forecasting* (Vol. 2). New York: springer.

Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, *15*(1), 125-139.

Wang, J., Sun, L., Zhao, H., & Wang, Y. (2016). ARIMA-BP integrated intelligent algorithm for China’s consumer price index forecasting and its applications. *Journal of Intelligent & Fuzzy Systems*, *31*(4), 2187-2193.

Harchaoui, T. M., & Janssen, R. V. (2018). How can big data enhance the timeliness of official statistics?: The case of the US consumer price index. *International Journal of Forecasting*, *34*(2), 225-234.

Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J., & Winton, J. (2018). Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *181*(3), 737-756.

Faisal, F. (2012). Forecasting Bangladesh's Inflation Using Time Series ARIMA Models. *World Review of Business Research*, *2*(3), 100-117.