Chapter 3

Research Methodology

3.1. Descriptive Statistic Analysis

To study on statistical data of mean, median, maximum, minimum, standard deviation, variance, coefficient of variation, and Ambient Air Standard of five basic air pollutants in Chiang Mai. To highlight the critical air pollutants which exceed the National Ambient Air Standard and impact to human health, environment, and economy.

3.2 Time Series Analysis

To graphical analysis of time series plots in order to understand time behavior of these air pollution's series (to see trends and volatility). Moreover, to highlight the critical air pollutants and period of severe problem occurs.

ARIMA modeling under Box-Jenkins method

1) Unit Root Test to check stationarity of data by employing Augmented Dickey–Fuller Test (ADF-Test). If necessary, transform the series to induce stationarity.

2) ARIMA process (four steps as below)

(i) Identification: From the autocorrelation properties of the transformed series, choose a few ARMA specifications for estimation and testing in order to arrive at a preferred specification with white noise residuals.

(ii) Estimation: Take the preferred ARMA models from Identification step to estimate parameter and examine statistical significant by testing T-statistic.

(iii) Diagnostic Checking: Check autocorrelation from estimated error if any white noise by testing Q-statistic.

(iv) Forecasting: To predict the future air pollutant concentration, there are three phases of forecasting.

(a) Historical Forecast: To forecast the air pollutant concentration in the past from January 1st, 1997 to April 30th, 2010.

(b) Ex-post Forecast: To forecast the air pollutant concentration in the past (shorter period; one year backcast) from May 1st, 2009 to April 30th, 2010 and compare with the actual data. The best fit model should provide the minimum Root Mean Square Error (RMSE), Theil's Inequality Coefficient (U), and Akaike Information Criterion (AIC).

(c) Ex-ante Forecast: Take the best fit model to forecast the future air pollutant concentration (the next one year period) from May 1st, 2010 to April 30th, 2011.

Unit root test: Augmented Dicky-Fuller (ADF) test;

Null hypothesis; $H_0 : \rho = 1$ from the below equation.

$$X_t = \rho X_{t-1} + \varepsilon_t$$

If $|\rho| = 1$, X_t is stationary.

If $\rho = 1$, X_t is nonstationary

Or we can test from this equation as well as the above: $\Delta X_t = \theta X_{t-1} + \varepsilon_t$

Where;

$\rho = (1+\theta)$

If $\theta =$ negative value, then ρ from the above equation is lower than one.

To reject $H_0 : \theta = 0$ results to accept $H_a : \theta < 0$, this means $\rho < 1$ and X_t has integration of order zero or X_t stationary. On the other hand, if we cannot reject $H_0 : \theta = 0$, this mean is nonstationary.

Cointegration test from residuals of regression equation;

$$\Delta e_t = \hat{\gamma} e_{t-1} + v_t \tag{3.1}$$

After that to get t-statistic from ratio: $\frac{\hat{\gamma}}{S.E.\hat{\gamma}}$ and compare with MacKinnon critical values according to null hypothesis of no cointegration: H_0: $\gamma = 0$, significant negative t-statistic will force to reject H_0 and lead to the conclusion that nonstationary variables are cointegrated.

Situations on Unit Root Test

Situation 1: no constant term and no linear trend in the model

The equation that will be estimated for performing the test is:

$$Y_t = \beta_1 Y_{t-1} + u_t \tag{3.2}$$

We want to test the null hypothesis,
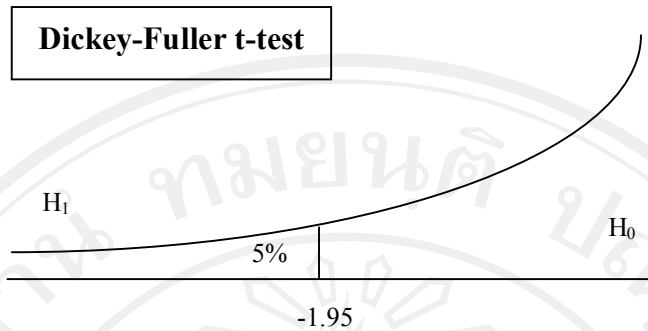
$$H_0: Y_t \sim I(1) \ or \ H_0 : \beta_1 = 1,$$

against

$$H_1: Y_t \sim I(0) \ or \ H_1 : 0 < \beta_1 < 1,$$

It is convenient to subtract Y_{t-1} from both sides of equation (3.2) and to estimate the parameter in the equation:

$$\nabla Y_t = (\beta_1 - 1)Y_{t-1} + u_t \tag{3.3}$$

Write the model with $\theta = (\beta_1 - 1)$,

$$\nabla Y_t = \theta Y_{t-1} + u_t \tag{3.4}$$

Source: Vogelvang, Ben (Econometrics: Theory and Applications with Eviews, 2005)

Figure 3.1: Unit Root Test- no constant term and no linear trend

$$Y_t = (\beta_1 - 1)Y_{t-1} + u_t$$

estimate θ with OLS and compute the t-statistic $t_{\hat{\theta}}$:

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{\text{s.e}(\hat{\theta})}$$

This t-statistic is directly given in the computer output of the unit-root test. Then the null and alternative hypotheses are identical to:

$$H_0 : \theta = 0$$
$$H_1 : \theta < 0$$

This t-statistic follows a *DF-t-distribution* under the null hypothesis of a unit root. For example, suppose we have a series with 100 observations, then the critical value is -1.95 at the 5% significance level. The testing of the null hypothesis has been illustrated in Figure 3.1.

Situation 2: a constant term but not a linear trend in the model

The omission of the constant term *in* the equation (3.2) can be rather restrictive. With a constant term the equation that will be estimated is:
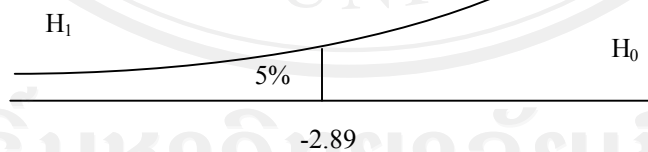
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t \tag{3.5}$$

Again this equation is rewritten to make it more suitable to test directly the null hypothesis $H_0: Y_t \sim I(1) \; versus \; H_1: Y_t \sim I(0)$,

$$\nabla Y_t = \beta_0 + (\beta_1 - 1)Y_{t-1} + u_t \tag{3.6}$$

and with $\theta = (\beta_1 - 1)$, the equation is:

$$\nabla Y_t = \beta_0 + \theta Y_{t-1} + u_t \tag{3.7}$$

Dickey-Fuller t-test

H₁

H₀

5%

-2.89

Source: Vogelvang, Ben (Econometrics: Theory and Applications with Eviews, 2005)

Figure 3.2: Unit Root Test- a constant term but no linear trend

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

estimate the parameters and compute the t-statistic $t_{\hat{\theta}}$:

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{s.e\,(\hat{\theta})}$$

This t-statistic follows a *DF-t-distribution* under the null hypothesis of a unit root. In EViews, the critical values for the test are computed. Using the same example as in the first situation of a time-series with 100 observations, the 5% critical value is -2.89. The picture that illustrates the test is given in Figure 12.4. The critical value of - 2.89 clearly deviates from the 5% critical value of the Student's t-distribution: $t_{0.05}(100) = -1.66$.

<u>Situation 3: a constant term and a deterministic linear trend in the model</u>

The last situation that we consider is a situation that a linear deterministic trend has also to be included in the equation. Then the equation is:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_t + u_t \qquad (3.8)$$

The null hypothesis of a unit root will be tested in the rewritten equation:

$$\nabla Y_t = \beta_0 + (\beta_1 - 1)Y_{t-1} + \gamma_t + u_t \qquad (3.9)$$

$$\nabla Y_t = \beta_0 + \theta Y_{t-1} + \gamma_t + u_t \qquad (3.10)$$

Just as before, estimate the θ with OLS and test $H_0: Y_t \sim I(1)\ versus\ H_1: Y_t \sim I(0)$, with the t-statistic $t_{\hat{\theta}}$:

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{\text{s.e}\,(\hat{\theta})}$$

by using a *DF-t-distribution*. For example, the 5% critical value for *n = 100* is -3.45.