11-18-2008

Time Series Forecasting Model for Chinese Future Marketing Price of Copper and Aluminum

Zhejin Hu

TIME SERIES FORECASTING MODEL FOR CHINESE FUTURE MARKETING PRICE OF COPPER AND ALUMINUM

by

ZHEJIN HU

Under the Direction of Dr. Yu-Sheng Hsu

ABSTRACT

This thesis presents a comparison for modeling and forecasting Chinese futures market of copper and aluminum with single time series and multivariate time series under linear restrictions. For single time series, data transformation for stationary purpose has been tested and performed before ARIMA model was built. For multivariate time series, co-integration rank test has been performed and included before VECM model was built. Based on selected models, the forecasting shows multivariate time series analysis has a better result than single time series, which indicates utilizing the relationships among the series can improve the accuracy of time series forecasting.

INDEX WORDS: ARIMA model, Stationary, VECM model, Cointegration

TIME SERIES FORECASTING MODEL FOR CHINESE FUTURE MARKETING PRICE OF COPPER AND ALUMINUM

by

ZHEJIN HU

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2008

TIME SERIES FORECASTING MODEL FOR CHINESE FUTURE MARKETING PRICE OF COPPER AND ALUMINUM

by

ZHEJIN HU

Major Professor: Dr. Yu-Sheng Hsu

Committee: Dr. Yichuan Zhao
Dr. Yuanhui Xiao

Electronic Version Approved:

Office of Graduate Studies
College of Art and Sciences
Georgia State University
December 2008

ACKNOWLEDGEMENTS

First and foremost, my deepest gratitude goes to my advisor, Dr. Yu-Sheng Hsu, professor and graduate director of Statistics. This thesis would not have been possible without his encouragement and guidance.

I would also like to thank my committee members, Dr. Yichuan Zhao and Dr. Yuanhui Xiao. I sincerely appreciate their time to go through my thesis and also their insightful comments.

I would also like to thank Dr. Hung-Gay Fung, professor in Chinese studies at University of Missouri, St. Louis, for providing the data and the background of the data.

Lastly, I would like to thank my family, friends and my employer, who have shown tremendous support during this education process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARIMA	Auto Regressive and Integrated Moving Average
VECM	Vector Error Correction Model
STS	Single Time Series
MTS	Multivariate Time Series
ADF	Augmented Dickey-Fuller Test
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
IACF	Inverse Autocorrelation Function
SAC	Sample Autocorrelation Function
SPAC	Sample Partial Autocorrelation Function
ESACF	Extended Sample Autocorrelation Function
MINIC	MINimum Information Criterion
SCAN	Smallest CANonical
OLS	Ordinary Least Square
ML	Maximum Likelihood

CHAPTER ONE: INTRODUCTION

1.1 Background

In 1978, the economic revolution in China began when the Government contributed to turning a planed/centralized economy to one that operates on the principles of a market economy. Since then, The Chinese economy becomes one of the fastest growing economies in the world. China has exhibited a consistent growth of more than 9 percent with the per capita income swelling by four times since 2001. The economy of China is expected to grow in the range of 7 to 9 percent over the next decade, which can be termed as a strong growth rate compared to leading economies like America and Japan.

One of several groundbreaking initiatives in the financial markets is the establishment of Chinese future market. On Oct. 12, 1990, the establishment of the China Zhengzhou Grain Wholesale Market was approved to run on an experimental basis by the State Council, and while this new exchange mainly focused on spot trading, it also introduced the futures trading for the first time in China. On June 10, 1991, the Shenzhen Metal Exchange was established, and it started trading on Jan. 18, 1992. About four months later, the Shanghai Metal Exchange also started trading on May 28, and by the second half of 1993, the number of futures exchanges topped 50. China's first futures brokerage company, Guangdong Wantong, was established in September of 1992.

In 1995, with the agreement of the State Council, the China Securities Regulatory Commission (CSRC) approved 15 pilot futures exchanges. Then, following the State Council's

Notice on Further Rearranging and Standardizing Futures Markets, the 13 existing exchanges were consolidated into the Shanghai Futures Exchange (SHFE), whereas the Zhengzhou Commodity Exchange (ZCE) and Dalian Commodity Exchange (DCE) were kept as they were as independent exchanges. The Shanghai Futures Exchange (SHFE) ranked third among China's three active derivatives platforms by contract volume in 2007, but it was by far the largest in terms of value, listing contracts in copper, aluminum, natural rubber, fuel oil, zinc and gold.

The topic of this thesis is about forecasting China's copper and aluminum futures markets, which are the third largest in the world.

1.2 Source of Data

Data used in this thesis is the daily Chinese future market open/close price of copper and aluminum from 2003/12/11 to 2007/04/16.

Open – opening price of the futures contracts during the day

Close – closing price of the futures contracts during the day

There are 833 records in the original data of each series excluding holidays and weekends. For the missing data, we use procedure, which combines time trend regression with an autoregressive model and uses a stepwise method to select the lags to use for the autoregressive process. After filling in all the missing data, the number of records is 1214 for each series. When building model, the first 1204 records are taken. The last 10 records is used for calculating the sum of square of the residuals between the actual value and forecasting value.

Variable name conventions used in the thesis are, c_open stands for copper open price, c_close stands for copper close price, a_open stands for aluminum open price and a_close stands for aluminum close price.

CHAPTER TWO: METHOD OF ANALYSIS

In this chapter, we discuss and review the techniques of Univariate (Single) Time Series analysis and Multivariate time Series analysis. By definition, a time series is a sequence of data points, measured typically at successive times, spaced at (often uniform) time intervals. There are two main goals of time series analysis: (a) identifying the nature of the phenomenon represented by the sequence of observations, and (b) forecasting (predicting future values of the time series variable). Both of these goals require that the pattern of observed time series data is identified and more or less formally described; as a result, model identification is the most critical part of time series analysis.

2.1 Univariate (Single) Time Series

Univariate Time Series refers to a time series that consists of single observations recorded sequentially over equal time increments. In our case, we have four Univariate Time Series, and they are copper open price, copper close price, aluminum open price and aluminum close price.

2.1.1 Box and Jenkins Method

The ARIMA methodology developed by Box and Jenkins (1976) has gained enormous popularity in many areas and research practice and its power and flexibility in time series analysis have been widely proved. And this will be the methodology we are going to use for our forecasting on single time series. In general, Box_Jenkins methodology consists of a four-step iterative procedure.

I. Model Identification: using historical data to tentatively identify an appropriate Box Jenkins Model.

II. Model Estimation: once the tentative model is identified, historical data is used again to estimate the parameters of the model.

III. Model Diagnose: various diagnostics are used to check the adequacy of the tentatively identified model, such as autocorrelation check of residuals. And if it is necessary, an improved model is suggested, it then regarded as a new tentatively identified model.

IV. Forecasting: Once a final model is built, it can be used to forecast future time series values.

As we mentioned earlier, among these four steps, model identification is most critical one, we will talk about the details of this step in section 2.1.4.

2.1.2 Data Stationary and Unit Root Test

It is very important to understand the Box-Jenkins approach is under the assumption that the time series is stationary. Thus, in order to tentatively identify a Box-Jenkins model. We need first identify whether or not the time series under study is stationary, if not, transformation or differencing is needed to make the time series stationary, which means the first two moments are preserved through the time.

To determine whether the time series is stationary or not, a plot of the series is the first step to see if the mean and the variance remain the same through the time. Secondly, the plot of ACF and PACF are also very helpful to examine if a difference is needed. Normally, if the plot

of ACF decays very slowly, while PACF cuts off after lag 1, it indicates that difference is needed. Additional statistical test such as Augmented Dickey-Fuller (ADF) test can then be performed to see if there is a unit root, To further explain the Augmented Dickey-Fuller unit roots test, let's look at the general AR(p) model,

$$\Phi_p(B)Z_t = a_t \tag{2.1}$$

where we need to test if $\Phi_p(B) = 1 - \Phi_1 B - ... - \Phi_p B^p$ may contain a unit root. Let $\Phi_{p-1}(B) = 1 - \Phi_1 B - ... - \Phi_{p-1} B^{p-1}$ and assume all the roots of $\Phi_{p-1}(B)$ are outside the unit circle. Thus, we can rewrite equation (2.1) as,

$$(1-B)\Phi_{p-1}(B)Z_t = (Z_t - Z_{t-1}) - \sum_{j=1}^{p-1} \Phi_j (Z_{t-j} - Z_{t-j-1}) = a_t \tag{2.2}$$

The testing for unit root becomes to test if $\Phi = 1$ in equation (2.3)

$$Z_t = \Phi Z_{t-1} + \sum_{j=1}^{p-1} \Phi_j \Delta Z_{t-j} + a_t \tag{2.3}$$

The null hypothesis of the Augmented Dickey-Fuller t-test is that there exists unit root versus the alternative hypothesis of no unit roots exist. If we cannot reject the null hypothesis that there exists unit root, it means the time series is not stationary.

Based on what discussed above, if the time series turns out to be not stationary, we can often transform it to become stationary with one of the following techniques.

I. We can difference the data. That is, given the series Z_t, we create the new series

$$Y_i = Z_i - Z_{i-1}$$

The differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.

II. For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points. The above transformations are actually the special cases of the Box-Cox transformation. The transformation can be represented by,

$$T(Z_t) = \frac{Z_t^{l} - 1}{l} \quad (2.4)$$

Table 1 shows some common Box-Cox transformations, where Z_t represents the original time series.

Table 1 Common Box-Cox Transformations

Value of l	Transformation
-2	$1/Z_t^2$
-1	$1/Z_t$
-0.5	$1/\sqrt{Z_t}$
0	$\log Z_t$
0.5	$\sqrt{Z_t}$
1	No Transformation
2	Z_t^2

2.1.3 Components of ARIMA Model

Once the differenced/transformed series is stationary, we can follow the general stationary ARMA (p,q) procedure to build Auto Regressive Integrated Moving Average (ARIMA) Model. ARIMA model contains three parts, and is usually written as ARIMA (p,d,q), in which p represents the auto regressive part of the model, d represents the order of the

difference taken on the data to make it stationary, q represents the moving average part of the model. It's like the regression model but the independent variables are the past values of the time series and past/current random terms. The general form of ARIMA (p,d,q) model is,

$$f_p(B)(1-B)^d Z_t = q_0 + q_q(B)a_t$$

where

$f_p(B)$ represents the AR part: $1 - f_1 B - \cdots - f_p B^p$

$q_q(B)$ represents the MA part: $1 - q_1 B - \cdots - q_q B^q$

a_t represents a zero mean white noise process with constant variance

Once the differenced/transformed series is stationary, we can follow the general stationary ARMA (p,q) guidelines and use Sample Autocorrelation Function (SAC) and Sample Partial Autocorrelation Function (SPAC) to identify the model.

In real case, the situation could be much more complicate than what has been covered by the guidelines above, we should combine the above the guidelines with other techniques, such as Extended Sample Autocorrelation Function (ESACF) method, MINimum Information Criterion (MINIC) method and Smallest CANonical (SCAN) correlation method. All together, the final model chosen should pass the model diagnostics. As mentioned earlier, the time series model selection is an iterative process for improved model to get the best result.

2.2 Multivariate Time Series

In many real cases, one event may consist of multiple time series variables. Often, these variables are not only contemporaneously correlated to each other, they are also correlated to

each other's past values. Among these time series, cross relationship may exist and could impact the behavior of each individual time series. By analyzing multiple time series jointly, we can utilize the additional information in the event and find out the dynamic relationships over time among the series. These will help to understand more about the pattern of the series and at the same time to get more accurate forecasting results.

The basic procedures of model building between univariate time series and multivariate time series are about the same. But multivariate time series do have some unique features other than univariate time series, such as cointegration between dependents.

2.2.1 Cointegration

As discussed earlier, many nonstationary univariate time series can be made stationary by appropriate differencing before ARMA models are fitted to the differenced series. However, for nonstationary multivariate time series, the situation starts to get more complicated. Since the dynamic of a multivariate time series is multidimensional, even if we can make individual variable stationary by appropriate differencing, the vector process of the differenced variables may still be nonstationary.

Engle and Granger (1987) formally demonstrated that it is possible for some linear combinations of the variables of nonstationary multivariate time series to be stationary. Later, this phenomenon is called Co-Integration. The concept of cointegration turned out to be extremely important in the modeling and analysis of non-stationary multivariate time series in economics and finances. Although economic/financial variables individually may exhibit

disequilibrium behaviors, in many cases, due to the nature of economic forces, these disequilibrium economic variables jointly form a dynamic equilibrium relationship. Specifically, certain linear combinations among the variables of a nonstationary multivariate time series might appear to be stationary.

From Single Time Series, we have a definition for integration as, a time series is said to be integrated of order $I(d)$, if it has a stationary, invertible, non-deterministic ARMA representation after differencing d times. Following Engle and Granger (1987), consider two time series y_{1t} and y_{2t}, which are both be integrated of order $I(d)$ individually, if there exists a vector $a = (1, -b)'$, such that the linear combination

$$Z_t = y_{1t} - by_{2t}$$

is $I(d-b)$, where $d \geq b > 0$, then y_{1t} and y_{2t} are defined as cointegrated of order (d, b), denoted as $y_t = (y_{1t}, y_{2t})' \sim CI(d, b)$, and vector a is called cointegrating vector. If Z_t has k components, then there could exsit more than one a. It is assumed that there are r linearly independent cointegrating vectors with $r < k$, and r is called the cointegration rank of Z_t.

2.2.2 VARMA Model and VECM Model

For a stationary Multivariable Time Series, the multivariate form of the Box-Jenkins univariate models is similar as the ARMA model for single time series. A VARMA(p,q) has the form as,

$$\Phi_p(B)\dot{Z}_t = \Theta_q(B)a_t$$

where

$\dot{Z}_t$ represents the vector or multivariate time series

$\Phi_p(B)$ represents the AR part: $1-\Phi_1 B-\cdots-\Phi_p B^p$

$\Theta_q(B)$ represents the MA part: $1-\Theta_1 B-\cdots-\Theta_q B^q$

a_t represents a sequence of k dimensional white noise random vectors

There are a few interesting properties associated with the AR parameter matrices. Consider the following example for a bivariate series with k =2, p = 2, and q = 0. The VARMA (2,0) model is,

$$\begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} \Phi_{1,11} & \Phi_{1,12} \\ \Phi_{1,21} & \Phi_{1,22} \end{pmatrix} \begin{pmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{pmatrix} + \begin{pmatrix} \Phi_{2,11} & \Phi_{2,12} \\ \Phi_{2,21} & \Phi_{2,22} \end{pmatrix} \begin{pmatrix} Z_{1,t-2} \\ Z_{2,t-2} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix}$$

Clearly, this model can also be written as two univariate regression equations,

$$Z_{1,t} = \Phi_{1,11} Z_{1,t-1} + \Phi_{2,11} Z_{1,t-2} + \Phi_{1,12} Z_{2,t-1} + \Phi_{2,12} Z_{2,t-2} + a_{1t}$$

$$Z_{2,t} = \Phi_{1,21} Z_{1,t-1} + \Phi_{2,21} Z_{1,t-2} + \Phi_{1,22} Z_{2,t-1} + \Phi_{2,22} Z_{2,t-2} + a_{2t}$$

When performing multivariate time series analysis on economics and financial data, in order to achieve equilibrium, there exists a dynamic system that consists of short-run changes leading to long-run equilibrium. This dynamic system is now known as the Vector Error Correction Model (VECM).

The concept of an error correction mechanism was first discussed by Phillips (1957), followed by Sargan (1964). After estimating the long-run relationship in the form of a cointegrating vector, short-run dynamics have to be incorporated to ensure that the system has

reached equilibrium. Vector Error Correction Model (VECM) was first proposed by Davidson et al. (1978). Since then, it has been widely used on economic studies, as it can lead to a better understanding of the nature of any nonstationary among the different component series and can also improve longer term forecasting over an unconstrained model.

The VECM(p) form is written as,

$$\Delta \dot{Z}_t = d + \Pi \dot{Z}_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta \dot{Z}_{t-i} + e_t$$

where $\Delta \dot{Z}_t = \dot{Z}_t - \dot{Z}_{t-1}$; $\Pi = a b'$, with a and b are $k \times r$ matrices; Φ_i^* is a $k \times k$ matrix. k is the number of the components in $\dot{Z}_t$ and r is the rank of cointegration $(r \le k)$. It has an equivalent VAR(p) representation as follows,

$$\dot{Z}_t = d + (I_k + \Pi + \Phi_1^*)\dot{Z}_{t-1} + \sum_{i=2}^{p-1} (\Phi_i^* - \Phi_{i-1}^*)\dot{Z}_{t-i} - \Phi_{p-1}^* \dot{Z}_{t-p} + e_t$$

Where I_k is a $k \times k$ identity matrix. The detail derivation can be seen in reference [8].

2.2.3 Testing and Estimating Cointegration

2.2.3.1 Engle and Granger (1987) Two-Step Procedure

Based on VECM, Engle and Granger (1987) proposed a residual-based method to test the existence of the cointegration, and a two-step procedure to estimate the cointegration.

For cointegration test, consider the cointegration regression, $Z_{1t} = a + bZ_{2t} + e_t$, If Z_{1t} and Z_{2t} are I(1) variables and are cointegrated in the ordinary least square (OLS) regression,

then the error term, $e_t = Z_{1t} - a - bZ_{2t}$, should be I(0). Thus, Engle and Granger suggest that the residuals are tested for unit roots, where we could perform Dickey-Fuller test, Augmented Dickey-Fuller test. If there exist unit roots, we can conclude cointegration exists.

The cointegration estimation from Engle and Granger (1987) consists of two steps. First, the long-run equation is to be obtained by regressing the cointegrating variables (OLS regression). The dependent variable is selected according to theory. As only one cointegration relationship can be estimated by this method, one regression equation is estimated, and the residuals are saved for the second step. In the second step, the ECM is analysed by regressing the differenced variables with the lag values of the residual of the long-run regression.

But this method has many drawbacks. Such as, the results of the ECM depend on the results of the first long-run equation. If there is a misspecification in the long-run equation, it will affect the results of the ECM. In addition, if there is more than one cointegrating relationship, this method cannot be employed.

2.2.3.2 Johansen's Approach

Johansen (1988) suggests a Maximum Likelihood (ML) procedure to obtain a and b by decomposing the matrix Π of the VECM, $\Pi = a b'$. The most critical part of Johansen's approach is to identify the rank of cointegrations. To do that, a likelihood ratio test of hypotheses procedure is used. The Johansen procedure consists of three steps,

I. All variables are pre-tested to assess the order of integration. It is easier to detect the possible trends when a series is plotted. The order of integration is important,

because variables with different orders of integration pose problems in setting the cointegration relationship. Order of integration is detected by the unit root tests. The lag length test is used to find the number of lag values that should be included in the model.

II. Model estimation and determination of the rank. There are two test statistics produced by the Johansen Maximum Likelihood (ML) procedure. They are the Trace test and maximal eigenvalue test. Both of the tests can be used to determine the number of distinct cointegrating vectors, although they don't always indicate the same number of cointegrating vectors. The Trace test is a joint test, the null hypothesis is that the number of cointegrating vectors is less than or equal to r, against a general alternative hypothesis that there are more then r. The Maximal Eigenvalue test conducts separate tests on each eigenvalue. The null hypothesis is that there are r cointegrating vectors present against the alternative that there are (r + 1) present.

III. Estimating the VECM, incorporating the cointegrating relations from the previous step. If the variables are cointegrated, after selecting the rank, the normalized b vector and the short-run dynamics are estimated.

The Johansen procedure is preferred to the residual-based approach because it makes it possible to analyze with more complex models. Johansen's approach allows us to deal with models with several endogenous variables. The procedure begins with an unrestricted VAR involving potentially non-stationary variables. Also, a key aspect of the approach is isolating and

identifying the r cointegrating combinations among a set of k integrated variables and merging them into an empirical model.

CHAPTER THREE: RESULTS AND DISCUSSIONS

Now, let’s look at the empirical results when we apply the techniques mentioned in Chapter two to forecast the daily open/close price of Chinese future market of copper and aluminum. At the end of this Chapter, we will conduct a comparison between the forecasting results of using STS and MTS.

3.1 Single Time Series Forecasting

3.1.1 Single Time Series Forecasting Copper

As discussed, plots of time series is very important and straight forward to get the first overall trend of a time series, by examining the plots carefully, we can get some ideas of time series stationary in both mean and variance for further test and research.

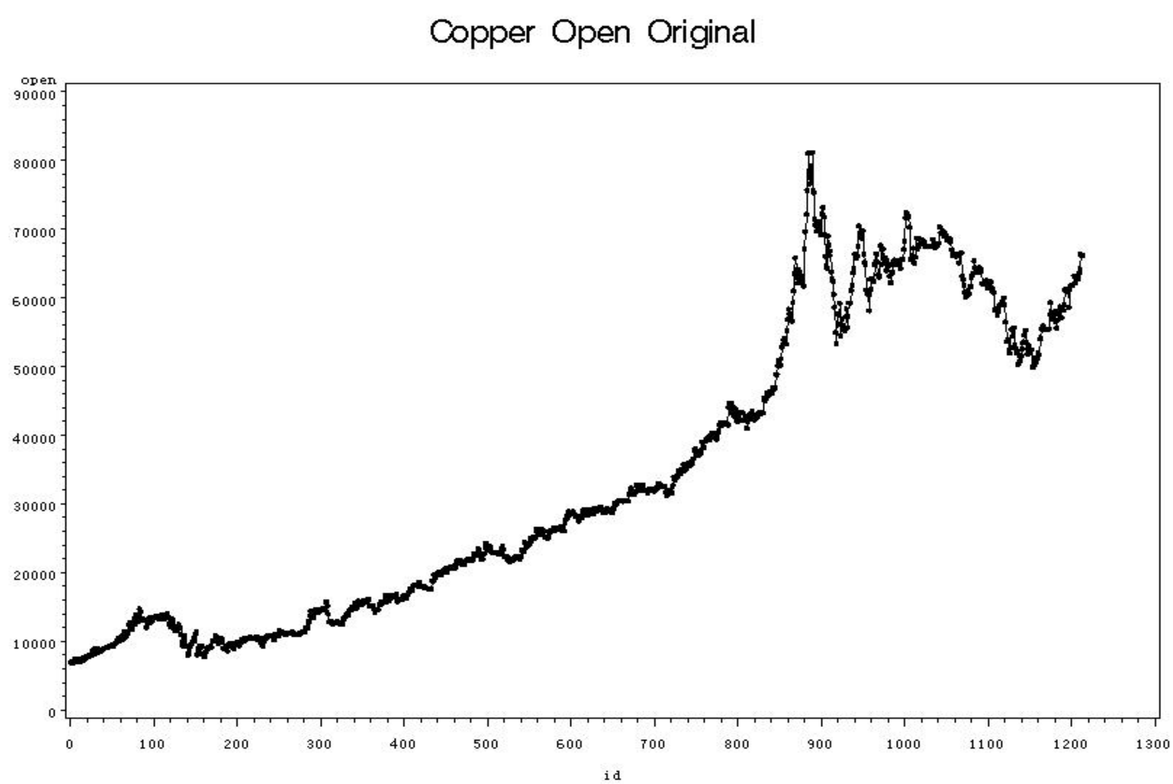

Figure 1 Open Price for Copper

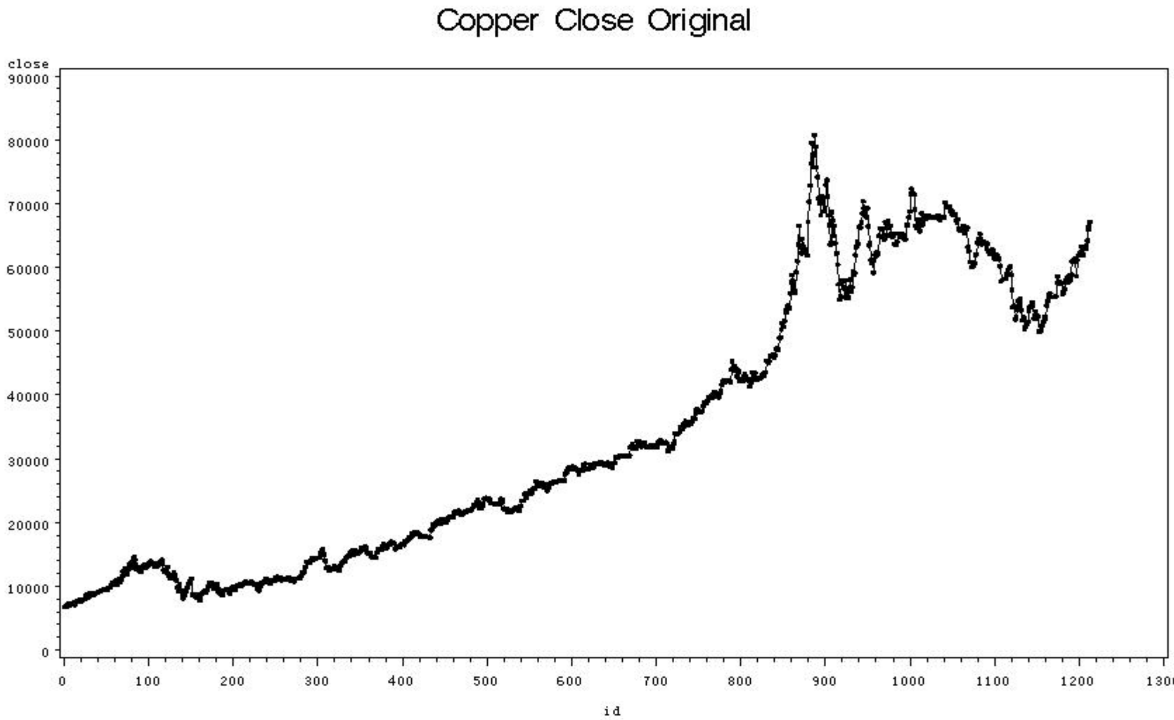

Figure 2 Close Price for Copper

Obviously, both series have very similar pattern overall, but within each series, there exists different trend patterns. As we know, all time series analysis needs to be stationary, or at least stationary after transformation. It could be very difficult to model the time series if it's in totally different pattern and can not be transformed to stationary series no matter what. At the same time, we know the most recent history will have more impact on forecasting then the remote history. After examining the plots carefully, we think both of the series for copper should be divided into three parts, first part includes the first 150 observations (2003/12/11 ~ 2004/05/08), the second part includes observations between 151 and 849 (2004/05/09 ~ 2006/04/07), the third part includes observations 850 (2006/04/08) after for the purpose of stationarity. As mentioned in Chapter one, for result comparison purpose, we will leave the last 10 observation out, and focus on observations between 850 and 1213 (2006/04/08 ~ 2007/04/06). Thus, we will have 364 observations for model identification.

According to the plots, for observations after 850, looks like the variance are not quite the same, we need to consider if Box-Cox Transformations are needed first. Table 2 lists the outputs from SAS. It recommends no transformation is needed.

Table 2 Box-Cox Transformation for STS Copper

Open Price			**Close Price**		
Lambda	**R-Square**	**Log Like**	**Lambda**	**R-Square**	**Log Like**
-2	0.13	-3167.92	-2	0.14	-3164.65
-1.5	0.13	-3161.56	-1.5	0.14	-3158.33
-1	0.14	-3156.24	-1	0.14	-3152.98
-0.5	0.14	-3151.99	-0.5	0.14	-3148.65
0	0.14	-3148.84	0	0.15	-3145.36
0.5	0.15	-3146.82 *	0.5	0.15	-3143.13 *
1.0 +	**0.15**	**-3145.99** <	**1.0** +	**0.15**	**-3142.01** <
1.5	0.15	-3146.36 *	1.5	0.15	-3142.02 *
2	0.15	-3147.97	2	0.15	-3143.19 *

< - Best Lambda * - Confidence Interval + - Convenient Lambda

Next, let's check the ACF, PACF plots of the original open/close price of copper and the ADF test for unit roots to see if a difference is needed.

I. ACF, PACF of original open price of copper (c_open)

```
                               Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1    Std Error

  0    37501011      1.00000     |                    |********************|            0
  1    36257926      0.96685     |                  . |******************* |     0.052414
  2    34943644      0.93181     |                .   |******************* |     0.088789
  3    33591985      0.89576     |                .   |******************  |     0.112491
  4    32154894      0.85744     |               .    |*****************   |     0.130625
  5    30921290      0.82455     |              .     |****************    |     0.145267
  6    29547201      0.78790     |              .     |****************    |     0.157601
  7    28309281      0.75489     |             .      |***************     |     0.168075
  8    27004243      0.72009     |             .      |**************      |     0.177144
  9    25742971      0.68646     |             .      |**************      |     0.185012
 10    24747996      0.65993     |            .       |*************       |     0.191881

                           Partial Autocorrelations

            Lag   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

              1     0.96685     |                  . |******************* |
              2    -0.04597     |                  .*|  .                 |
              3    -0.03266     |                  .*|  .                 |
              4    -0.05326     |                  .*|  .                 |
              5     0.06509     |                  . |* .                 |
              6    -0.07992     |                  **|  .                 |
              7     0.03877     |                  . |* .                 |
              8    -0.05538     |                  .*|  .                 |
              9     0.01076     |                  . |  .                 |
             10     0.07539     |                  . |**                  |
```

II. ACF, PACF of original close price of copper (c_close)

```
                                        Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error
  0      36833714        1.00000    |                    |********************|              0
  1      35741483        0.97035    |                  . |******************* |       0.052414
  2      34506460        0.93682    |                .   |******************* |       0.088998
  3      33282190        0.90358    |               .    |******************  |       0.112884
  4      31948913        0.86738    |               .    |*****************   |       0.131259
  5      30594566        0.83061    |              .     |*****************   |       0.146160
  6      29344192        0.79667    |              .     |****************    |       0.158598
  7      28205179        0.76574    |             .      |***************     |       0.169236
  8      26827558        0.72834    |             .      |***************     |       0.178501
  9      25695788        0.69762    |             .      |**************      |       0.186486
 10      24737645        0.67160    |            .       |*************       |       0.193523
```

```
                                    Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
                1        0.97035    |                  . |******************* |
                2       -0.08140    |                  **| .                  |
                3       -0.00806    |                  . | .                  |
                4       -0.06938    |                  .*| .                  |
                5       -0.02270    |                  . | .                  |
                6        0.02865    |                  . |*.                  |
                7        0.02969    |                  . |*.                  |
                8       -0.13475    |                 ***| .                  |
                9        0.10720    |                  . |**                  |
               10        0.04005    |                  . |*.                  |
```

Table 3 ADF Test for STS Copper

a. ADF test for original open price							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	0.1376	0.7149	0.31	0.7762		
	1	0.1358	0.7144	0.3	0.7726		
	2	0.1158	0.7096	0.25	0.758		
Single Mean	0	-11.6581	0.0885	-2.63	0.0884	3.62	0.1427
	1	-12.773	0.0671	-2.73	0.0701	3.9	0.0935
	2	-13.6091	0.0545	-2.75	0.0675	3.91	0.0926
Trend	0	-14.7886	0.187	-3.08	0.1127	4.88	0.1969
	1	-16.425	0.1366	-3.22	0.0816	5.34	0.1049
	2	-17.6141	0.1079	-3.24	0.0792	5.34	0.104
b. ADF test for original close price							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-331.845	0.0001	-17.47	<.0001		
	1	-326.959	0.0001	-12.75	<.0001		
	2	-244.569	0.0001	-9.52	<.0001		
Single Mean	0	-332.21	0.0001	-17.47	<.0001	152.53	0.001
	1	-327.858	0.0001	-12.75	<.0001	81.24	0.001
	2	-245.522	0.0001	-9.52	<.0001	45.32	0.001
Trend	0	-332.431	0.0001	-17.45	<.0001	152.28	0.001
	1	-328.398	0.0001	-12.74	<.0001	81.11	0.001
	2	-245.998	0.0001	-9.51	<.0001	45.22	0.001

Obviously, the plots of ACF for both c_open and c_close decay very slowly, while PACF cuts off after lag 1, it indicates that difference is needed. Also, the ADF tests indicate both series for copper have unit roots. Now, let's check the ACF, PACF plots of the open/close price of copper after the first order difference and the ADF test for unit roots as well. To help identify the model, IACF is also listed.

I. ACF, IACF and PACF of open price of copper (c_open) after first order difference

```
                                    Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0       2061924        1.00000    |                    |********************|             0
  1     60252.664        0.02922    |                  . |*.                  |      0.052486
  2     64710.271        0.03138    |                  . |*.                  |      0.052531
  3        161591        0.07837    |                  . |**                  |      0.052583
  4       -187707        -.09103    |                  **| .                  |      0.052904
  5        170595        0.08274    |                  . |**                  |      0.053333
  6       -130135        -.06311    |                  .*| .                  |      0.053686
  7     53849.482        0.02612    |                  . |*.                  |      0.053890
  8    -52566.917        -.02549    |                  .*| .                  |      0.053925
  9       -205527        -.09968    |                  **| .                  |      0.053958
 10       -248664        -.12060    |                  **| .                  |      0.054463
```

```
                                Inverse Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1       -0.09473    |                  **| .                  |
                2        0.04106    |                  . |*.                  |
                3       -0.14662    |                 ***| .                  |
                4        0.10807    |                  . |**                  |
                5       -0.20274    |                ****| .                  |
                6        0.14023    |                  . |***                 |
                7       -0.08651    |                  **| .                  |
                8        0.08712    |                  . |**                  |
                9        0.00836    |                  . | .                  |
               10        0.18791    |                  . |****                |
```

```
                                Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1        0.02922    |                  . |*.                  |
                2        0.03056    |                  . |*.                  |
                3        0.07672    |                  . |**                  |
                4       -0.09695    |                  **| .                  |
                5        0.08513    |                  . |**                  |
                6       -0.07142    |                  .*| .                  |
                7        0.04344    |                  . |*.                  |
                8       -0.04971    |                  .*| .                  |
                9       -0.07238    |                  .*| .                  |
               10       -0.14243    |                 ***| .                  |
```

II. ACF, IACF and PACF of close price of copper (c_close) after first order difference

```
                                   Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0       1764630        1.00000    |                    |********************|             0
   1        145106        0.08223    |                  . |**                  |      0.052486
   2     20760.562        0.01176    |                  . | .                  |      0.052840
   3        180990        0.10257    |                  . |**                  |      0.052847
   4     22476.173        0.01274    |                  . | .                  |      0.053393
   5    -77666.701        -.04401    |                  .*| .                  |      0.053401
   6    -99408.068        -.05633    |                  .*| .                  |      0.053501
   7        229395        0.13000    |                  . |***                 |      0.053664
   8       -242046        -.13717    |                 ***| .                  |      0.054525
   9       -102582        -.05813    |                  .*| .                  |      0.055467
  10       -162036        -.09182    |                  **| .                  |      0.055635
```

```
                        Inverse Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1       -0.12241    |                  **| .                  |
                2        0.01950    |                  . | .                  |
                3       -0.15074    |                 ***| .                  |
                4        0.04864    |                  . |*.                  |
                5       -0.06456    |                  .*| .                  |
                6        0.08652    |                  . |**                  |
                7       -0.13821    |                 ***| .                  |
                8        0.15041    |                  . |***                 |
                9       -0.01218    |                  . | .                  |
               10        0.11575    |                  . |**                  |
```

```
                        Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1        0.08223    |                  . |**                  |
                2        0.00504    |                  . | .                  |
                3        0.10188    |                  . |**                  |
                4       -0.00401    |                  . | .                  |
                5       -0.04638    |                  .*| .                  |
                6       -0.06058    |                  .*| .                  |
                7        0.14183    |                  . |***                 |
                8       -0.15633    |                 ***| .                  |
                9       -0.02043    |                  . | .                  |
               10       -0.12049    |                  **| .                  |
```

From the plots of ACF and PACF, we can conclude the single series of open/close price of copper become stationary after first order difference. The ADF result in Table 4 also shows the stationary of the series.

Table 4 ADF Test for STS Copper after First Order Difference

a. ADF test for open price after first order difference							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-351.098	0.0001	-18.44	<.0001		
	1	-328.7	0.0001	-12.79	<.0001		
	2	-263.505	0.0001	-9.81	<.0001		
Single Mean	0	-351.422	0.0001	-18.43	<.0001	169.81	0.001
	1	-329.519	0.0001	-12.79	<.0001	81.76	0.001
	2	-264.441	0.0001	-9.8	<.0001	48.06	0.001
Trend	0	-351.693	0.0001	-18.42	<.0001	169.6	0.001
	1	-330.155	0.0001	-12.78	<.0001	81.67	0.001
	2	-265.159	0.0001	-9.79	<.0001	47.98	0.001
b. ADF test for close price after first order difference							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-331.845	0.0001	-17.47	<.0001		
	1	-326.959	0.0001	-12.75	<.0001		
	2	-244.569	0.0001	-9.52	<.0001		
Single Mean	0	-332.21	0.0001	-17.47	<.0001	152.53	0.001
	1	-327.858	0.0001	-12.75	<.0001	81.24	0.001
	2	-245.522	0.0001	-9.52	<.0001	45.32	0.001
Trend	0	-332.431	0.0001	-17.45	<.0001	152.28	0.001
	1	-328.398	0.0001	-12.74	<.0001	81.11	0.001
	2	-245.998	0.0001	-9.51	<.0001	45.22	0.001

For open price of copper, the output of SCAN and ESACF is as below,

```
      ARMA(p+d,q) Tentative
      Order Selection Tests

    ---SCAN--    --ESACF--
    p+d     q    p+d     q
      0     0      0     0
                   2     1
                   3     1
                   4     1
                   5     1
```

Combine with the plots of ACF, IACF, PACF, and also check the diagnose of residuals

```
                              Autocorrelation Check of Residuals
  To        Chi-             Pr >
 Lag      Square     DF     ChiSq    --------------------Autocorrelations--------------------

   6        0.07      1    0.7865     0.000     0.001    -0.002    -0.009     0.009    -0.005
  12        9.59      7    0.2128    -0.003     0.010    -0.118    -0.103     0.026    -0.006
  18       19.47     13    0.1093     0.080     0.129     0.029    -0.030    -0.034     0.009
  24       29.20     19    0.0629    -0.030     0.012     0.089    -0.070    -0.104     0.018
  30       33.00     25    0.1310     0.038     0.024    -0.017     0.073    -0.015    -0.041
  36       38.70     31    0.1611     0.021    -0.034    -0.106    -0.021     0.002    -0.031
  42       40.09     37    0.3347     0.015    -0.019     0.017    -0.023     0.037     0.025
  48       50.00     43    0.2154    -0.047    -0.093     0.012    -0.045     0.016    -0.102
```

Final model for open price of copper (c_open) is

$$(1+0.84701B-0.05514B^2-0.10579B^3+0.01716B^4)(1-B)c\_open=(1+0.91548B)a_t$$

For close price of copper, the output of SCAN and ESACF don't give much useful information.

```
ARMA(p+d,q) Tentative
Order Selection Tests

---SCAN--    --ESACF--
p+d     q    p+d     q

  0     0      0     0
```

We examine the plots of ACF, IACF and PACF, and it shows, there are spikes at leg 1,3 7,8. And the diagnose of residuals shows a good fit.

```
                         Autocorrelation Check of Residuals

 To        Chi-             Pr >
Lag      Square     DF     ChiSq    --------------------Autocorrelations--------------------

  6        0.07      1    0.7865     0.000     0.001    -0.002    -0.009     0.009    -0.005
 12        9.59      7    0.2128    -0.003     0.010    -0.118    -0.103     0.026    -0.006
 18       19.47     13    0.1093     0.080     0.129     0.029    -0.030    -0.034     0.009
 24       29.20     19    0.0629    -0.030     0.012     0.089    -0.070    -0.104     0.018
 30       33.00     25    0.1310     0.038     0.024    -0.017     0.073    -0.015    -0.041
 36       38.70     31    0.1611     0.021    -0.034    -0.106    -0.021     0.002    -0.031
 42       40.09     37    0.3347     0.015    -0.019     0.017    -0.023     0.037     0.025
 48       50.00     43    0.2154    -0.047    -0.093     0.012    -0.045     0.016    -0.102
```

Final model for close price of copper (c_close) is

$$(1-0.1109B-0.09315B^3-0.14707B^7+0.15716B^8)(1-B)c\_close=a_t$$

Based on the models, the forecasting result is list in Table 5

Table 5 Forecasting for STS Copper

a. Forecasting for open price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
365	66423.4292	1412.7945	63654.4029	69192.4555
366	66164.4947	2067.5291	62112.2122	70216.7773
367	66411.0117	2557.5752	61398.2564	71423.767
368	66211.0824	3051.0832	60231.0692	72191.0955
369	66362.6174	3406.7186	59685.5716	73039.6632
370	66253.7656	3784.2463	58836.7792	73670.752
371	66328.9376	4087.2292	58318.1157	74339.7596
372	66278.7268	4396.845	57661.0689	74896.3848
373	66311.2843	4667.6808	57162.798	75459.7705
374	66290.7601	4935.4313	56617.4926	75964.0276

b. Forecasting for close price				
Obs	Forecast	Std Error	95% Confidence Limits	
365	67256.61	1295.524	64717.43	69795.79
366	67299.1	1936.407	63503.81	71094.39
367	67480.7	2422.206	62733.26	72228.14
368	67504.01	2890.649	61838.44	73169.57
369	67659.78	3306.119	61179.9	74139.65
370	67497.18	3676.87	60290.65	74703.71
371	67510.1	4018.304	59634.37	75385.83
372	67424.05	4409.312	58781.96	76066.14
373	67392	4709.599	58161.35	76622.64
374	67409.67	4979.734	57649.57	77169.77

3.1.2 Single Time Series Forecasting for Aluminum

Similar as what we did with the open/close price of copper, we first plot the open/close price of Aluminum as showed in Figure 3 and Figure 4. Again, both series have very similar pattern overall, but within each series, there are exists different trend patterns.

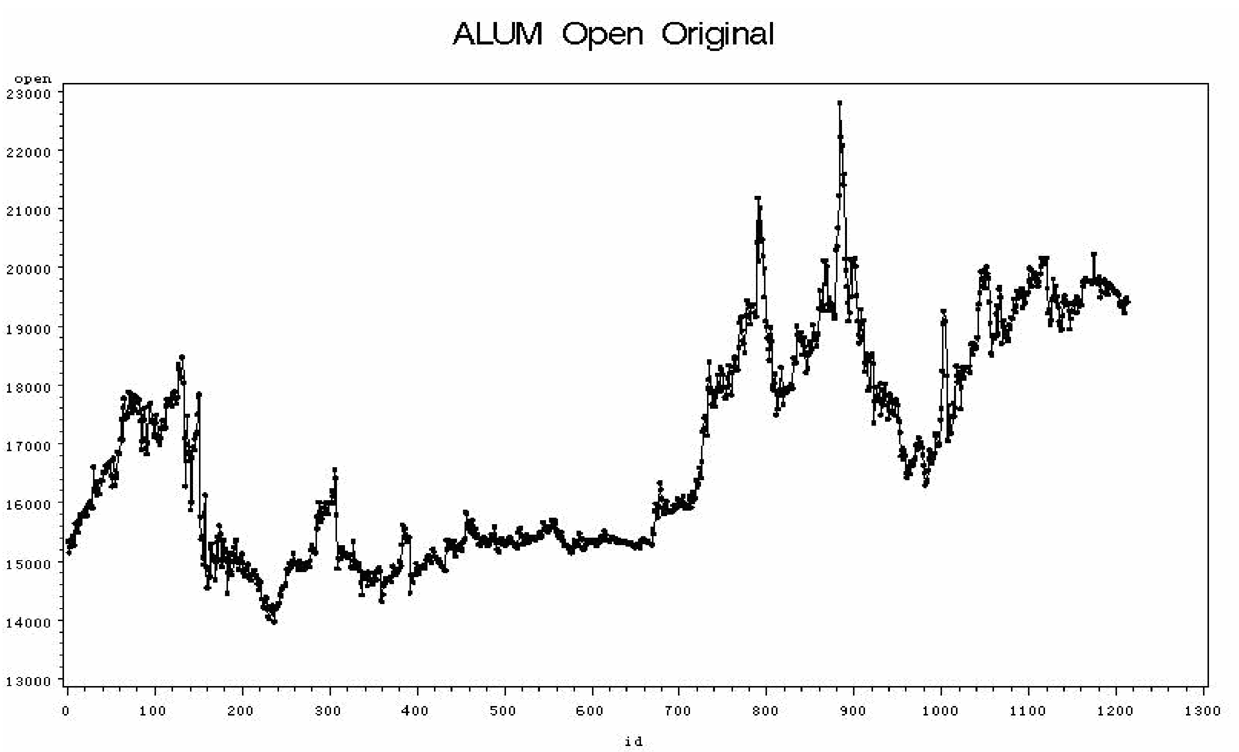

Figure 3 Open Price for Aluminum

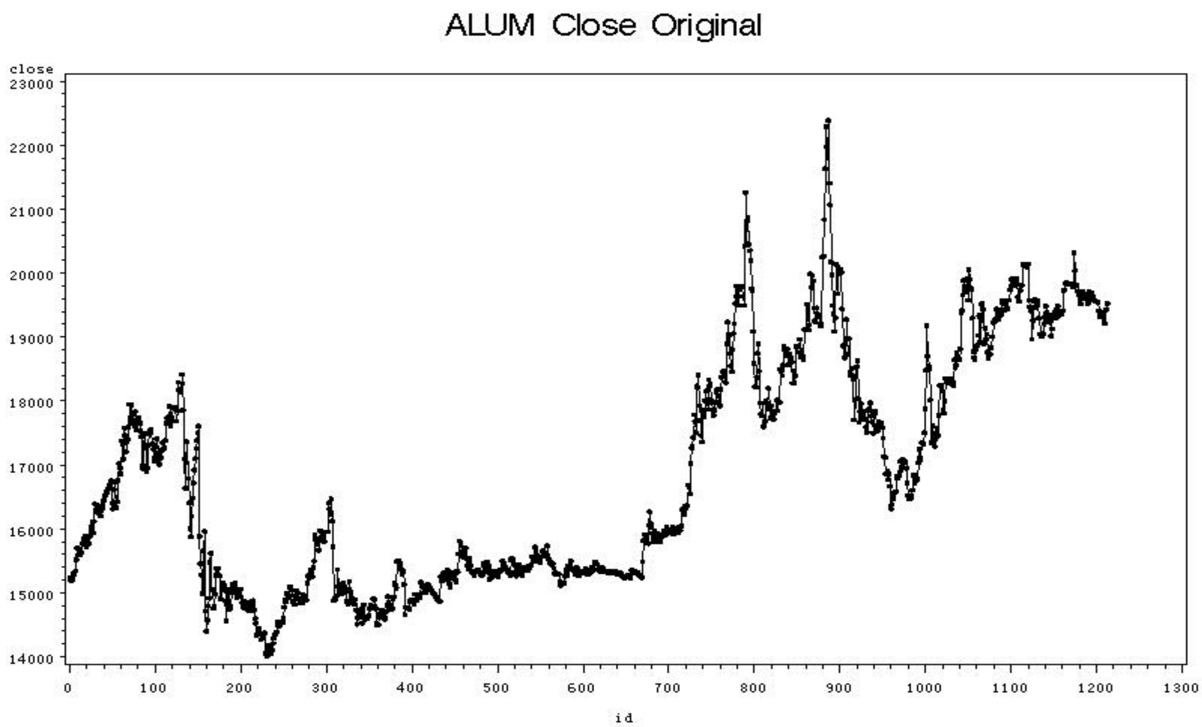

Figure 4 Close Price for Aluminum

According to the plots, we decide both of the series for aluminum can be also divided into three parts, first part includes the first 350 observations (2003/12/11 ~ 2004/11/24), the second part includes observations between 351 and 749 (2004/11/25 ~ 2005/12/28), the third part includes observations 750 (2005/12/29) after. By leaving the last 10 observation out for forecasting purpose, we will build model focus on observations between 750 and 1213 (2005/12/29 ~ 2007/04/06). Thus, we will have 464 observations for model identification.

The plots of both open/close price of aluminum also show both of them are not stationary in variance. Transformation may be needed before considering taking any difference. Table 6 gives the result of BoxCox transformation recommendations. Not surprisely, both series have similar results. Although, the results yield a range of lambda from 0.5 to 2, combine with the plots, we decide to choose the square root transformation to make the variance smaller.

Table 6 Box-Cox Transformation for STS Aluminum

Open Price			**Close Price**		
Lambda	**R-Square**	**Log Like**	**Lambda**	**R-Square**	**Log Like**
-2	0.06	-3239.55	-2	0.06	-3233.63
-1.5	0.06	-3235.36	-1.5	0.06	-3229.66
-1	0.06	-3231.73	-1	0.06	-3226.22
-0.5	0.06	-3228.68	-0.5	0.06	-3223.34
0	0.06	-3226.22	0	0.06	-3221.01
0.5	**0.06**	**-3224.36 ***	**0.5**	**0.06**	**-3219.25 ***
1.0 +	0.06	-3223.13 *	1.0 +	0.06	-3218.08 *
1.5	0.06	-3222.53 <	1.5	0.06	-3217.50 <
2	0.06	-3222.60 *	2	0.06	-3217.52 *
< - Best Lambda		* - Confidence Interval		+ - Convenient Lambda	

Next, let's check if a difference is needed on the transformed series. The ACF, PACF plots of the transformed open/close price of aluminum and the ADF test for unit roots is as below,

I. ACF, PACF of original open price of aluminum (a_open)

```
                                  Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0     15.355565        1.00000    |                    |********************|              0
  1     14.783072        0.96272    |                  . |******************* |       0.046424
  2     14.258449        0.92855    |                 .  |******************* |       0.078423
  3     13.790869        0.89810    |                .   |******************  |       0.099330
  4     13.226996        0.86138    |               .    |*****************   |       0.115513
  5     12.672237        0.82525    |               .    |*****************   |       0.128613
  6     12.221217        0.79588    |              .     |****************    |       0.139560
  7     11.866373        0.77277    |              .     |***************     |       0.149021
  8     11.559424        0.75278    |              .     |***************     |       0.157421
  9     11.224891        0.73100    |             .      |***************     |       0.164997
 10     10.981138        0.71512    |             .      |**************      |       0.171835


                              Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1        0.96272    |                  . |******************* |
                2        0.02361    |                  . | .                  |
                3        0.03480    |                  . |*.                  |
                4       -0.09770    |                  **| .                  |
                5       -0.01573    |                  . | .                  |
                6        0.06584    |                  . |*.                  |
                7        0.08468    |                  . |**                  |
                8        0.04502    |                  . |*.                  |
                9       -0.03918    |                  .*| .                  |
               10        0.05681    |                  . |*.                  |
```

II. ACF, PACF of original close price of aluminum (a_close)

```
                                  Autocorrelations

Lag     Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      15.008095        1.00000    |                    |********************|             0
  1      14.607531        0.97331    |                  . |******************* |      0.046424
  2      14.142135        0.94230    |                 .  |******************* |      0.078984
  3      13.685998        0.91191    |                .   |******************  |      0.100328
  4      13.162191        0.87701    |               .    |******************  |      0.116834
  5      12.659677        0.84352    |               .    |*****************   |      0.130251
  6      12.235499        0.81526    |              .     |****************    |      0.141536
  7      11.878218        0.79145    |              .     |****************    |      0.151318
  8      11.516934        0.76738    |              .     |***************     |      0.159991
  9      11.238105        0.74880    |             .      |***************     |      0.167736
 10      10.981670        0.73172    |             .      |***************     |      0.174792
```

```
                             Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1        0.97331    |                  . |******************* |
                2       -0.09555    |                  **| .                  |
                3        0.00236    |                  . | .                  |
                4       -0.10472    |                  **| .                  |
                5        0.02322    |                  . | .                  |
                6        0.07387    |                  . |*.                  |
                7        0.06419    |                  . |*.                  |
                8       -0.03436    |                  .*| .                  |
                9        0.08233    |                  . |**                  |
               10       -0.00831    |                  . | .                  |
```

Table 7 ADF Test for Transformed STS Aluminum

a. ADF test for transformed open price							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	0.0188	0.6871	0.11	0.7184		
	1	0.0255	0.6886	0.16	0.7327		
	2	0.031	0.6899	0.21	0.7459		
Single Mean	0	-16.9463	0.0238	-2.93	0.0432	4.32	0.0679
	1	-16.0716	0.0296	-2.84	0.0545	4.05	0.0839
	2	-14.9583	0.0392	-2.73	0.0708	3.76	0.1084
Trend	0	-18.0589	0.0998	-3.02	0.1267	4.57	0.2582
	1	-17.063	0.1217	-2.91	0.1595	4.24	0.325
	2	-15.8166	0.155	-2.78	0.2039	3.88	0.3976
b. ADF test for transformed close price							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	0.0307	0.6898	0.22	0.7508		
	1	0.0303	0.6897	0.2	0.7447		
	2	0.0341	0.6906	0.23	0.7534		
Single Mean	0	-11.8907	0.0841	-2.45	0.1284	3.05	0.2901
	1	-14.5333	0.0436	-2.69	0.0766	3.66	0.1324
	2	-14.3834	0.0453	-2.67	0.0813	3.6	0.1484
Trend	0	-12.6389	0.2786	-2.52	0.3161	3.19	0.5381
	1	-15.4502	0.1663	-2.77	0.2094	3.84	0.407
	2	-15.2454	0.1729	-2.73	0.2244	3.73	0.4279

From the above, the plots of ACF for both square roots of a_open and a_close decay very slowly, while PACF cuts off after lag 1, it indicates that difference is needed. Also, the ADF tests indicate both series for aluminum have unit roots. To identify the order of the difference and the model, let's check the ACF, IACF and PACF plots of the open/close price of copper after the first order difference and the ADF test for unit roots as well.

I. ACF, IACF and PACF of square roots of open price of aluminum (a_open) after first order difference

```
                                     Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0      1.125626        1.00000    |                    |********************|              0
   1     -0.053092        -.04717    |                  .*| .                  |       0.046474
   2     -0.058853        -.05229    |                  .*| .                  |       0.046577
   3      0.099040        0.08799    |                  . |**                  |       0.046704
   4    -0.0092608        -.00823    |                  . | .                  |       0.047060
   5     -0.103981        -.09238    |                  **| .                  |       0.047064
   6     -0.093558        -.08312    |                  **| .                  |       0.047454
   7     -0.043945        -.03904    |                  .*| .                  |       0.047767
   8      0.022114        0.01965    |                  . | .                  |       0.047836
   9     -0.094632        -.08407    |                  **| .                  |       0.047853
  10     -0.196923        -.17495    |                 ***| .                  |       0.048171

                                 Inverse Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1        0.07942    |                  . |**                  |
                2        0.00937    |                  . | .                  |
                3       -0.05357    |                  .*| .                  |
                4        0.05449    |                  . |*.                  |
                5        0.10143    |                  . |**                  |
                6        0.06112    |                  . |*.                  |
                7        0.01796    |                  . | .                  |
                8        0.00628    |                  . | .                  |
                9        0.08986    |                  . |**                  |
               10        0.17010    |                  . |***                 |

                                 Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1       -0.04717    |                  .*| .                  |
                2       -0.05463    |                  .*| .                  |
                3        0.08324    |                  . |**                  |
                4       -0.00300    |                  . | .                  |
                5       -0.08512    |                  **| .                  |
                6       -0.10090    |                  **| .                  |
                7       -0.05701    |                  .*| .                  |
                8        0.02182    |                  . | .                  |
                9       -0.07405    |                  .*| .                  |
               10       -0.19083    |                ****| .                  |
```

II. ACF, IACF and PACF of square roots of close price of aluminum (a_close) after first order difference

```
                                 Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      0.769890        1.00000    |                    |********************|              0
  1      0.065996        0.08572    |                  . |**                  |       0.046474
  2     -0.012199        -.01585    |                  . | .                  |       0.046814
  3      0.070253        0.09125    |                  . |**                  |       0.046826
  4     -0.020362        -.02645    |                  .*| .                  |       0.047208
  5     -0.078588        -.10208    |                  **| .                  |       0.047240
  6     -0.063118        -.08198    |                  **| .                  |       0.047714
  7     0.0051081        0.00663    |                  . | .                  |       0.048018
  8     -0.084608        -.10990    |                  **| .                  |       0.048020
  9     -0.024692        -.03207    |                  .*| .                  |       0.048560
 10     -0.083661        -.10867    |                  **| .                  |       0.048605
```

```
                          Inverse Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1       -0.08950    |                  **| .                  |
                2        0.05979    |                  . |*.                  |
                3       -0.11055    |                  **| .                  |
                4        0.04914    |                  . |*.                  |
                5        0.04320    |                  . |*.                  |
                6        0.09460    |                  . |**                  |
                7       -0.06456    |                  .*| .                  |
                8        0.09082    |                  . |**                  |
                9       -0.01498    |                  . | .                  |
               10        0.11279    |                  . |**                  |
```

```
                          Partial Autocorrelations

              Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

                1        0.08572    |                  . |**                  |
                2       -0.02337    |                  . | .                  |
                3        0.09540    |                  . |**                  |
                4       -0.04408    |                  .*| .                  |
                5       -0.09296    |                  **| .                  |
                6       -0.07690    |                  **| .                  |
                7        0.02283    |                  . | .                  |
                8       -0.10168    |                  **| .                  |
                9       -0.00442    |                  . | .                  |
               10       -0.13338    |                 ***| .                  |
```

From the plots of ACF and PACF, we can conclude the single series of transformed open/close price of aluminum become stationary after first order difference. The ADF result in Table 8 also shows the stationary of the series.

Table 8 ADF Test for Transformed STS Aluminum after First Order Difference

a. ADF test for transformed open price after first order difference							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-483.744	0.0001	-22.52	<.0001		
	1	-538.909	0.0001	-16.4	<.0001		
	2	-417.724	0.0001	-11.96	<.0001		
Single Mean	0	-483.794	0.0001	-22.5	<.0001	253.05	0.001
	1	-539.117	0.0001	-16.38	<.0001	134.22	0.001
	2	-418.045	0.0001	-11.95	<.0001	71.36	0.001
Trend	0	-483.793	0.0001	-22.47	<.0001	252.5	0.001
	1	-539.124	0.0001	-16.37	<.0001	133.94	0.001
	2	-418.088	0.0001	-11.93	<.0001	71.21	0.001
b. ADF test for transformed close price after first order difference							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-422.308	0.0001	-19.7	<.0001		
	1	-441.491	0.0001	-14.83	<.0001		
	2	-335.597	0.0001	-11.05	<.0001		
Single Mean	0	-422.383	0.0001	-19.68	<.0001	193.62	0.001
	1	-441.725	0.0001	-14.82	<.0001	109.77	0.001
	2	-335.915	0.0001	-11.04	<.0001	60.98	0.001
Trend	0	-422.384	0.0001	-19.66	<.0001	193.2	0.001
	1	-441.735	0.0001	-14.8	<.0001	109.53	0.001
	2	-335.947	0.0001	-11.03	<.0001	60.85	0.001

For the transformed open price of aluminum, the output of SCAN and ESACF is as following,

```
 ARMA(p+d,q) Tentative
Order Selection Tests

---SCAN--    --ESACF--
p+d     q    p+d     q

  1     1      1     1
  0     5      0     5
```

The autocorrelation of residuals of the above recommendations show no good fit. We examine the plots of ACF, IACF and PACF, and it shows there are spikes at leg 3, 5, 6 and 10. The residual diagnostics is as below,

```
                        Autocorrelation Check of Residuals

 To        Chi-             Pr >
Lag      Square     DF     ChiSq    --------------------Autocorrelations--------------------

  6        1.10      1    0.2947     0.002    -0.026    -0.010    -0.030     0.008     0.025
 12        8.26      7    0.3099    -0.035     0.012    -0.080     0.000     0.038     0.077
```

18	12.87	13	0.4582	-0.038	0.049	0.040	0.048	-0.037	-0.020
24	15.63	19	0.6821	-0.007	-0.032	0.005	0.045	-0.050	-0.001
30	17.53	25	0.8617	0.025	0.015	0.010	-0.010	-0.041	-0.034
36	17.83	31	0.9717	-0.019	0.003	-0.014	0.000	-0.006	-0.004
42	21.87	37	0.9772	-0.033	-0.041	-0.053	-0.012	0.032	-0.034
48	25.07	43	0.9868	0.014	-0.048	0.011	0.013	0.056	-0.015

Final model for open price of aluminum (a_open) is

$$(1-0.07879B^{3}+0.11152B^{5}+0.09796B^{6}+0.18313B^{10})(1-B)\sqrt{a\_open}=(1-0.0707B)a_{t}$$

For the transformed close price of aluminum, the output of SCAN and ESACF is as following,

ARMA(p+d,q) Tentative
Order Selection Tests

---SCAN--		--ESACF--	
p+d	q	p+d	q
2	2	1	1
4	1	5	4
1	4	0	5
5	0		
0	5		

The autocorrelation of residuals of the above recommendations shows no good fit either. We examine the plots of ACF, IACF and PACF, and it shows there are spikes at leg 1, 5 and 6. And the diagnose of residuals indicates a good fit,

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	1.84	2	0.3981	0.005	0.037	0.047	-0.005	-0.017	-0.001
12	10.50	8	0.2315	-0.034	-0.060	-0.038	-0.093	-0.044	0.040
18	15.78	14	0.3269	0.023	0.060	0.063	0.020	-0.036	0.035
24	18.14	20	0.5781	-0.032	-0.006	0.026	0.007	-0.055	-0.007
30	25.96	26	0.4655	0.056	-0.054	0.034	0.051	-0.067	-0.040
36	34.43	32	0.3522	-0.072	0.010	-0.061	0.080	0.024	-0.031
42	37.48	38	0.4933	-0.035	-0.025	-0.048	-0.037	-0.022	-0.005
48	42.44	44	0.5387	-0.036	-0.048	-0.005	0.060	0.044	0.024

Final model for square root of close price of aluminum (a _close) is

$$(1+0.71595B+0.11037B^{5}+0.14465B^{6})(1-B)\sqrt{a\_close}=(1+0.80959B)a_{t}$$

Based on the models, the forecasting result is list in Table 9

Table 9 Forecasting for Transformed STS Aluminum

a. Forecasting for square root of open price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
465	139.5232	1.0331	137.4984	141.548
466	139.6458	1.4103	136.8817	142.41
467	139.5527	1.706	136.2089	142.8965
468	139.4659	1.9988	135.5484	143.3835
469	139.4638	2.2512	135.0516	143.876
470	139.5458	2.4321	134.779	144.3126
471	139.545	2.571	134.5058	144.5842
472	139.4312	2.705	134.1296	144.7328
473	139.3973	2.8272	133.8561	144.9385
474	139.452	2.9405	133.6888	145.2151
b. Forecasting for square root of close price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
465	139.8556	0.8641	138.1619	141.5493
466	139.8601	1.2806	137.3502	142.3699
467	139.7944	1.5578	136.7411	142.8477
468	139.7322	1.8136	136.1776	143.2867
469	139.7372	2.0241	135.7699	143.7044
470	139.6701	2.1857	135.3861	143.954
471	139.7023	2.3089	135.177	144.2276
472	139.6858	2.4424	134.8988	144.4728
473	139.714	2.5587	134.699	144.729
474	139.7023	2.6762	134.4571	144.9475

3.2 Multivariate Time Series Forecasting

3.2.1 Multivariate Time Series Forecasting for Copper

In section 3.1.1, we talked about the stationary test of open/close price of copper as separate single time series. The results indicate both single series are not stationary before taking the first order difference. But both of them became stationary after first order difference, where the DF test shows H_0 can be rejected. Recall the discussion in section 2.2.1, we can conclude that open/close price of copper are integrated of order $I(1)$ as single time series.

Now let's treat open/close price of copper together as a vector and the results of their DF test are shown in Table 10. It's clear that the first order difference on the vector is necessary to make the multivariate to be stationary.

Table 10 DF Test for MTS Copper

a. DF test for copper before difference					
Variable	**Type**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**
open	Zero Mean	0.14	0.7144	0.3	0.7726
	Single Mean	-12.77	0.0671	-2.73	0.0701
	Trend	-16.43	0.1366	-3.22	0.0816
close	Zero Mean	0.15	0.7181	0.34	0.7838
	Single Mean	-12.13	0.0787	-2.64	0.0866
	Trend	-15.53	0.1625	-3.09	0.1104
b. DF test for copper after first order difference					
Variable	**Type**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**
open	Zero Mean	-328.7	0.0001	-12.79	<.0001
	Single Mean	-329.52	0.0001	-12.79	<.0001
	Trend	-330.15	0.0001	-12.78	<.0001
close	Zero Mean	-326.96	0.0001	-12.75	<.0001
	Single Mean	-327.86	0.0001	-12.75	<.0001
	Trend	-328.4	0.0001	-12.74	<.0001

To determine the order of cointegration between the open/close price of copper, the Johansen's approach of cointegration rank test using trace shows the result in Table 11. Since we use option NOINT, Drift in ECM shows NOINT, which means there is separate drift in the VECM model. While Drift in Process shows Constant, which means there is a constant drift before differencing. Row 1 in Table 11 tests if r = 0 against r > 0, row 2 tests if r = 1 against r >1. as in row 1, the 5% critical value (12.21) is less than the trace value (60.8318), we move to row 2, which shows the 5% critical value (4.13) is greater than the trace value (0.0002), we conclude the multivariate series of open/close price of copper is integrated at order 1.

Table 11 Cointegration Rank Test for MTS Copper

Cointegration Rank Test Using Trace						
H0: Rank=r	**H1: Rank>r**	**Eigenvalue**	**Trace**	**5% Critical Value**	**Drift in ECM**	**Drift in Process**
0	0	0.155	60.5318	12.21	NOINT	Constant
1	1	0.0002	0.0541	4.14		

To determine the AR order of the model, Table 12 lists the schematic representations of partial autoregression, partial cross autoregression and the result of partial canonical correlations. Combine these information together, VECM(5) will be used.

Table 12 VECM Model Identification for MTS Copper

a. Schematic Representation of Partial Autoregression								
Variable/Lag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
open	-+	-+	..	-+	..	..	.+	..
close	..	..	..	-+	+-	..	.+	..
+ is > 2*std error, - is < -2*std error, . is between								

b. Schematic Representation of Partial Cross Autoregression								
Variable/Lag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
open	-+	-.	.+	-.	..	..	..	..
close	+.	+-	+.	+.	+-	..	.+	+-
+ is > 2*std error, - is < -2*std error, . is between								

c. Partial Canonical Correlations					
Lag	**Correlation1**	**Correlation2**	**DF**	**Chi-Square**	**Pr > ChiSq**
1	0.38615	0.07547	4	56.04	<.0001
2	0.3383	0.05464	4	42.39	<.0001
3	0.13606	0.12007	4	11.85	0.0185
4	0.27314	0.03868	4	27.32	<.0001
5	0.27147	0.02837	4	26.67	<.0001
6	0.09364	0.06869	4	4.81	0.3069
7	0.1868	0.07054	4	14.19	0.0067
8	0.23889	0.09064	4	23.18	0.0001

For model diagnostics purpose, after fitting VECM(5) model, the ANOVA diagnostics shows each univariate model is significant the residual is checked for normality and unequal variance of autoregressive conditional heterosedastic (ARCH), both of them show significant at 5% level. The Durbin Watson (DW) statistics are near 2 for both residual series. At the same time, the autoregressive models fit to these residuals up to lag 3 show no significance at 5% significant level. All these tests together indicates there is no correlation among the residuals, the model has a good fit. Details are in Table 13.

Table 13 Model Diagnostics for MTS Copper

a. Univariate Model ANOVA Diagnostics				
Variable	**R-Square**	**Standard Deviation**	**F Value**	**Pr > F**
open	0.5001	1422.40006	38.69	<.0001
close	0.3999	1400.34736	25.76	<.0001

b. Univariate Model White Noise Diagnostics					
Variable	**Durbin**	**Normality**		**ARCH**	
	Watson	**Chi-Square**	**Pr > ChiSq**	**F Value**	**Pr > F**
open	2.03119	86.47	<.0001	4.23	0.0404
close	2.0205	91.72	<.0001	7.27	0.0073

c. Univariate Model AR Diagnostics								
Variable	**AR1**		**AR2**		**AR3**		**AR4**	
	F Value	**Pr > F**	**F Value**	**Pr > F**	**F Value**	**Pr > F**	**F Value**	**Pr > F**
open	0.09	0.7616	0.56	0.5728	1.17	0.32	2.68	0.0315
close	0.04	0.8466	0.34	0.7125	1.07	0.3625	2.28	0.0601

The final model for multivariate time series of open/close price of copper is:

$$\Delta\begin{pmatrix} c\_open \\ c\_close \end{pmatrix}_t = \begin{pmatrix} -2.9035 & 2.98468 \\ 0.08243 & -0.08474 \end{pmatrix}\begin{pmatrix} c\_open \\ c\_close \end{pmatrix}_{t-1} + \begin{pmatrix} 1.19976 & -2.02204 \\ -0.1897 & -0.54228 \end{pmatrix}\Delta\begin{pmatrix} c\_open \\ c\_close \end{pmatrix}_{t-1}$$
$$+\begin{pmatrix} 0.51594 & -1.13919 \\ -0.44201 & -0.15628 \end{pmatrix}\Delta\begin{pmatrix} c\_open \\ c\_close \end{pmatrix}_{t-2} + \begin{pmatrix} 0.21385 & -0.52411 \\ -0.38758 & 0.10332 \end{pmatrix}\Delta\begin{pmatrix} c\_open \\ c\_close \end{pmatrix}_{t-3}$$
$$+\begin{pmatrix} -0.18786 & 0.00754 \\ -0.41664 & 0.29653 \end{pmatrix}\Delta\begin{pmatrix} c\_open \\ c\_close \end{pmatrix}_{t-4} + e_t$$

Based on above model, the forecasting for copper is listed in Table 14.

Table 14 Forecasting for MTS Copper

a. Forecasting for open price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
365	68100.53854	1422.40006	65312.68566	70888.3914
366	69190.77078	2252.16365	64776.61113	73604.9304
367	70460.35948	3112.98254	64359.02581	76561.6932
368	71181.62482	4096.53501	63152.56373	79210.6859
369	71545.62357	5160.04052	61432.13	81659.1172
370	72588.52076	6392.4671	60059.51547	85117.526
371	73589.29471	7657.46619	58580.93676	88597.6527
372	74627.09114	9000.28563	56986.85547	92267.3268
373	75626.9725	10436.28234	55172.23498	96081.71
374	76449.46114	11918.29385	53090.03443	99808.8879

b. Forecasting for close price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
365	68502.19734	1400.34736	65757.56695	71246.8277
366	69513.29008	2266.35839	65071.30927	73955.2709
367	70482.21476	3084.75289	64436.2102	76528.2193
368	71145.96384	4056.05361	63196.24485	79095.6828
369	71714.56569	5136.2346	61647.73086	81781.4005
370	72847.10033	6312.39335	60475.0367	85219.164
371	73801.57383	7556.32895	58991.44123	88611.7064
372	74740.60293	8877.83448	57340.3671	92140.8388
373	75649.1856	10276.53576	55507.54563	95790.8256
374	76455.24792	11727.72913	53469.32121	99441.1746

3.2.2 Multivariate Time Series Forecasting for Aluminum

From the result of our study on the stationary test of open/close price of aluminum after square root transformation as separate single time series, it indicate both single series are not stationary before taking the first order difference. But both of them became stationary after first order difference, so that the square root of open/close price of aluminum are integrated of order *I*(*1*) as single time series. Combine the two single time series as a vector time series, Table 15 shows as a vector of square root of open/close price of aluminum, it becomes stationary after first order difference.

Table 15 DF Test for Transformed MTS Aluminum

a. DF test for transformed aluminum before first order difference					
Variable	**Type**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**
sqrt_open	Zero Mean	0.03	0.6886	0.16	0.7327
	Single Mean	-16.07	0.0296	-2.84	0.0545
	Trend	-17.06	0.1217	-2.91	0.1595
sqrt_close	Zero Mean	0.03	0.6897	0.2	0.7447
	Single Mean	-14.53	0.0436	-2.69	0.0766
	Trend	-15.45	0.1663	-2.77	0.2094
b. DF test for transformed aluminum after first order difference					
Variable	**Type**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**
sqrt_open	Zero Mean	-538.91	0.0001	-16.4	<.0001
	Single Mean	-539.12	0.0001	-16.38	<.0001
	Trend	-539.12	0.0001	-16.37	<.0001
sqrt_close	Zero Mean	-441.49	0.0001	-14.83	<.0001
	Single Mean	-441.73	0.0001	-14.82	<.0001
	Trend	-441.73	0.0001	-14.8	<.0001

Same as what we did for copper, the result of cointegration test of transformed prices of aluminum is listed in Table 16. Option NOINT is selected, Drift in ECM shows NOINT, which means there is separate drift in the VECM model. While Drift in Process shows Constant, which means there is a constant drift before differencing. We conclude the multivariate series of square root of open/close price of aluminum is integrated at order 1.

Table 16 Cointegration Rank Test for Transformed MTS Aluminum

Cointegration Rank Test Using Trace						
H0: Rank=r	**H1: Rank>r**	**Eigenvalue**	**Trace**	**5% Critical Value**	**Drift in ECM**	**Drift in Process**
0	0	0.3458	196.4967	12.21	NOINT	Constant
1	1	0.0001	0.052	4.14		

Table 17 VECM Model Identification for Transformed MTS Aluminum

a. Schematic Representation of Partial Autoregression

Variable/Lag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
sqrt_open	-+	-+	.+	-+	..	-.	-+	..
sqrt_close	-+	..	..	..	..	..	..	..

+ is > 2*std error, - is < -2*std error, . is between

b. Schematic Representation of Partial Cross Autoregression

Variable/Lag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
sqrt_open	-.	-.	.+	..	-.	..	-.	..
sqrt_close	+.	+-	+.	+.	..	..	..	.-

+ is > 2*std error, - is < -2*std error, . is between

c. Partial Canonical Correlations

Lag	**Correlation1**	**Correlation2**	**DF**	**Chi-Square**	**Pr > ChiSq**
1	0.35221	0.05389	4	58.65	<.0001
2	0.32352	0.04207	4	49.07	<.0001
3	0.18802	0.11015	4	21.84	0.0002
4	0.13509	0.07266	4	10.8	0.0289
5	0.11159	0.08703	4	9.17	0.0569
6	0.11465	0.03768	4	6.66	0.1552
7	0.13764	0.04625	4	9.61	0.0475
8	0.14549	0.01277	4	9.7	0.0457

Table 17 lists the schematic representations of partial autoregression, partial cross autoregression and the result of partial canonical correlations. It looks VECM(4) will be appropriate.

After fitting VECM(4) model, the ANOVA shows each univariate model inside the VECM model is significant. The residual check for normality and ARCH shows significant at 5% level. The Durbin Watson (DW) statistics is near 2 for both residual series. At the same time, the autoregressive models fit to these residuals up to lag 2 show no significance at 5% significant level. All these tests together the model has a good fit. Details are in Table 18.

Table 18 Model Diagnostics for Transformed MTS Aluminum

a. Univariate Model ANOVA Diagnostics

Variable	R-Square	Standard Deviation	F Value	Pr > F
sqrt_open	0.5461	1.03844	77.5	<.0001
sqrt_close	0.3572	0.95491	35.81	<.0001

b. Univariate Model White Noise Diagnostics

Variable	Durbin	Normality		ARCH	
	Watson	Chi-Square	Pr > ChiSq	F Value	Pr > F
sqrt_open	2.04427	237.36	<.0001	28.45	<.0001
sqrt_close	2.03946	177.69	<.0001	10.06	0.0016

c. Univariate Model AR Diagnostics

Variable	AR1		AR2		AR3		AR4	
	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F
sqrt_open	0.23	0.6308	0.97	0.3793	2.46	0.0618	4.83	0.0008
sqrt_close	0.19	0.6648	1.16	0.3136	2.86	0.0366	6.13	<.0001

The final model for square roots of open/close price of Aluminum is:

$$\Delta\begin{pmatrix}\sqrt{a\_open}\\ \sqrt{a\_close}\end{pmatrix}_t = \begin{pmatrix}-2.70173 & 2.78721\\ -0.41555 & 0.42870\end{pmatrix}\begin{pmatrix}\sqrt{a\_open}\\ \sqrt{a\_close}\end{pmatrix}_{t-1} + \begin{pmatrix}0.94714 & -1.73994\\ 0.16695 & -0.87847\end{pmatrix}\Delta\begin{pmatrix}\sqrt{a\_open}\\ \sqrt{a\_close}\end{pmatrix}_{t-1}$$
$$+\begin{pmatrix}0.13899 & -0.35508\\ 0.00068 & -0.51249\end{pmatrix}\Delta\begin{pmatrix}\sqrt{a\_open}\\ \sqrt{a\_close}\end{pmatrix}_{t-2} + \begin{pmatrix}0.13899 & -0.35508\\ 0.03621 & -0.21277\end{pmatrix}\Delta\begin{pmatrix}\sqrt{a\_open}\\ \sqrt{a\_close}\end{pmatrix}_{t-3} + e_t$$

Based on above model, the forecasting for square root of aluminum is listed in Table 19.

Table 19 Forecasting for Transformed MTS Aluminum

a. Forecasting for square root of open price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
465	139.6	0.9912	137.66	141.54
466	139.79	1.3913	137.06	142.52
467	139.64	1.7214	136.27	143.01
468	139.64	1.9823	135.76	143.53
469	139.68	2.2069	135.35	144
470	139.66	2.4151	134.93	144.4
471	139.66	2.6064	134.55	144.77
472	139.67	2.7837	134.21	145.12
473	139.66	2.9508	133.88	145.45
474	139.66	3.109	133.57	145.76
b. Forecasting for square root of close price				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
465	139.85	0.8715	138.14	141.56
466	139.91	1.2907	137.38	142.44
467	139.86	1.6026	136.72	143
468	139.86	1.8508	136.23	143.48
469	139.87	2.0681	135.82	143.92
470	139.86	2.2673	135.42	144.31
471	139.86	2.4498	135.06	144.66
472	139.86	2.6193	134.73	145
473	139.86	2.7787	134.42	145.31
474	139.86	2.9294	134.12	145.61

3.3 Comparison and Discussion

Table 20 lists the forecasting result for copper from both STS and MTS methods, the residual sum of square from MTS is much smaller than that from STS, and indicates a better forecasting result. Table 21 lists the forecasting result for aluminum from both STS and MTS methods, as the forecasting is originally based on the square root of each elements, the result is transformed back before calculating the residual sum of square. The comparison also shows a better forecasting result of MTS than STS. Therefore, both of cases indicate it is very necessary to take the relationships among the series into consideration to improve the accuracy of time series forecasting.

Table 20 Forecasting Result Comparison for Copper

a. Forecasting for open price Comparison						
Obs	**STS Forecast**	**Actual**	**SSR**	**MTS Forecast**	**Actual**	**SSR**
365	66423.43	66509.06	7332.22	68100.54	66509.06	2532811.57
366	66164.49	66292.59	16408.70	69190.77	66292.59	8399445.15
367	66411.01	68010.00	2556763.58	70460.36	68010.00	6004261.58
368	66211.08	70990.00	22838053.43	71181.62	70990.00	36720.07
369	66362.62	71710.00	28594500.67	71545.62	71710.00	27019.61
370	66253.77	71900.00	31879962.90	72588.52	71900.00	474060.84
371	66328.94	70900.00	20894611.46	73589.29	70900.00	7232306.04
372	66278.73	70709.04	19627716.71	74627.09	70709.04	15351087.89
373	66311.28	70889.62	20961116.92	75626.97	70889.62	22442550.99
374	66290.76	72200.00	34919116.20	76449.46	72200.00	18057919.98
Sum			**182,295,582.80**			**80,558,183.72**
a. Forecasting for close price Comparison						
Obs	**STS Forecast**	**Actual**	**SSR**	**MTS Forecast**	**Actual**	**SSR**
365	67256.61	67377.39	14587.70	68502.20	67377.39	1265192.56
366	67299.10	67361.26	3864.43	69513.29	67361.26	4631213.76
367	67480.70	69150.00	2786562.49	70482.21	69150.00	1774796.17
368	67504.01	71840.00	18800809.28	71145.96	71840.00	481686.19
369	67659.78	70730.00	9426250.85	71714.57	70730.00	969369.60
370	67497.18	72600.00	26038771.95	72847.10	72600.00	61058.57
371	67510.10	71210.00	13689260.01	73801.57	71210.00	6716254.92
372	67424.05	70917.72	12205731.64	74740.60	70917.72	14614432.17
373	67392.00	70777.95	11464689.45	75649.19	70777.95	23728890.16
374	67409.67	73120.00	32607868.71	76455.25	73120.00	11123878.69
Sum			**127,038,396.52**			**65,366,772.79**

Table 21 Forecasting Result Comparison for Aluminum

a. Forecasting for open price Comparison								
Obs	**STS Forecast SQRT**	**STS Forecast**	**Actual**	**SSR**	**MTS Forecas t SQRT**	**MTS Forecast**	**Actual**	**SSR**
465	139.52	19466.72	19444.19	507.59	139.60	19487.82	19444.19	1903.45
466	139.65	19500.95	19457.83	1859.30	139.79	19540.77	19457.83	6878.87
467	139.55	19474.96	19660.00	34241.25	139.64	19499.44	19660.00	25780.89
468	139.47	19450.74	20000.00	301689.55	139.64	19500.01	20000.00	249994.54
469	139.46	19450.15	19860.00	167975.78	139.68	19510.20	19860.00	122357.60
470	139.55	19473.03	19780.00	94230.40	139.66	19505.80	19780.00	75186.59
471	139.55	19472.81	19700.00	51616.65	139.66	19505.06	19700.00	38002.36
472	139.43	19441.06	19639.25	39281.05	139.67	19506.53	19639.25	17616.38
473	139.40	19431.61	19622.50	36438.24	139.66	19506.09	19622.50	13551.12
474	139.45	19446.86	19820.00	139233.23	139.66	19505.89	19820.00	98664.82
Sum				**867,073.04**				**649,936.62**

a. Forecasting for close price Comparison								
Obs	**STS Forecast SQRT**	**STS Forecast**	**Actual**	**SSR**	**MTS Forecast SQRT**	**MTS Forecast**	**Actual**	**SSR**
465	139.86	19559.59	19567.06	55.76	139.85	19558.61	19567.06	71.34
466	139.86	19560.85	19559.85	1.00	139.91	19573.95	19559.85	198.94
467	139.79	19542.47	19990.00	200279.28	139.86	19560.56	19990.00	184416.77
468	139.73	19525.09	20080.00	307927.64	139.86	19559.57	20080.00	270845.23
469	139.74	19526.49	19810.00	80380.72	139.87	19563.24	19810.00	60889.49
470	139.67	19507.74	19870.00	131234.60	139.86	19562.04	19870.00	94839.86
471	139.70	19516.73	19760.00	59179.02	139.86	19561.62	19760.00	39354.78
472	139.69	19512.12	19672.17	25615.18	139.86	19562.11	19672.17	12112.81
473	139.71	19520.00	19646.36	15967.22	139.86	19562.01	19646.36	7115.74
474	139.70	19516.73	19880.00	131963.19	139.86	19561.92	19880.00	101172.02
Sum				**952,603.61**				**771,016.97**

CHAPTER FOUR: CONCLUSION AND FUTURE RESEARCH

In this thesis, we discussed the methodologies of identifying models for both STS and MTS. Later, we applied these methodologies on forecasting Chinese future market open/close price of copper and aluminum. The result indicates that economic/financial elements are not only correlated to each other, they are also correlated to each other's past values. Therefore, incorporating the dynamic relationships among these elements is very important to improve the accuracy of the forecasting.

By comparing with the result from the thesis of Mr. David Yankey, who graduated in 2004 and had performed the similar analysis based on the same set of data of copper and aluminum from July 2000 to May 2003, the conclusion is consistent. This in addition supports the importance of utilizing the relationships among the series to improve the accuracy of time series forecasting.

Other than the elements which are under forecasting themselves (dependents/endogenous variables), there are other elements (independents/exogenous variables) within the environment could impact the behavior of dependents. In our particular case of Chinese future market open/close price of copper and aluminum, those exogenous variables could be the volume of the transactions, highest price of the day, lowest price of the day etc. It will be very interesting to see the result by including these exogenous variables into the model into the future research on this topic.

REFERENCES

[1] Renhai Hua & Baizhu Chen, 2007. "International linkages of the Chinese futures markets," Applied Financial Economics, Taylor and Francis Journals, vol. 17(16), pages 1275-1287.

[2] Wagge, L.L., 1992. "Identifiability of the Arma(p,q)-Model," Papers 403, California Davis - Institute of Governmental Affairs.

[3] Robert A. Yaffee, Monnie McGee, 2000. Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS

[4] Marno Verbeek, 2008. A Guide to Modern Econometrics

[5] Bruce L. Bowerman, Richard T. O'Connell, 2005. Anne B. Koehler, Forecasting, Time Series, and Regression

[6] William W. S. Wei, 2006. Time Series Analysis, Univariate and Multivariate Mehods

[7] Helmut Lütkepohl, Markus Krätzig, 2004. Applied Time Series Econometrics

[8] Christopher Chatfield, 2001. Time-series Forecasting

[9] Björn Schelter, Matthias Winterhalder, Jens Timmer, 2006. Handbook of Time Series Analysis

[10] John Clare Brocklebank, David A. Dickey, 2003. SAS for Forecasting Time Series

[11] Colin P. Hargreaves, 1994. Nonstationary Time Series Analysis and Cointegration

[12] Aman Ullah, David E. A. Giles, 1998. Handbook of Applied Economic Statistics

[13] R. F. Engle, Clive William John Granger, 1991. Long-run Economic Relationships

[14] David Yankey, 2004. Forecasting the Chinese Futures Market Prices of Aluminum and Copper Commodities

[15] SAS Institute, Inc., SAS/ETS User's Guide, The SAS System Version 9

APPENDICES

Appendix A: SAS Code

1. Fill in blanks using Proc forecast

```
data blanks;
input row len @@;
cards;
17 2 22 1 24 2 31 2 38 12 52 2 59 2 66 2 73 2 80 2 87 2 94 2 101 2 108 2 115
2 122 2 129 2 136 2 143 9 157 2 164 2 171 2 178 2 185 2 192 2 199 2 206 2 213
2 220 2 227 2 234 2 241 2 248 2 255 2 262 2 269 2 276 2 283 2 290 2 297 2 304
2 311 2 318 2 325 2 332 2 339 2 346 2 353 2 360 2 367 3 374 2 381 2 388 3 395
2 402 2 409 2 416 2 423 11 437 2 444 2 451 2 458 2 465 2 472 2 479 2 486 2
493 2 500 2 507 9 521 2 528 2 535 2 542 2 549 2 556 2 563 2 570 2 577 2 584 2
591 2 598 2 605 2 612 2 619 2 626 2 633 2 640 2 647 2 654 2 661 9 675 2 682 2
689 2 696 2 703 2 710 2 717 2 724 2 731 2 738 2 745 2 752 3 759 2 766 2 773 2
780 2 783 1 787 2 794 2 801 2 808 2 815 2 822 2 829 2 836 2 843 2 850 2 857 2
864 2 871 2 874 6 885 2 892 2 899 2 906 2 913 2 920 2 927 2 934 2 941 2 948 2
955 2 962 2 969 2 976 2 983 2 990 2 997 2 1004 2 1011 2 1018 2 1025 2 1032 2
1039 2 1046 2 1053 2 1060 2 1067 2 1074 2 1081 2 1088 2 1095 2 1102 2 1109 2
1116 4 1123 2 1130 2 1137 2 1144 2 1151 2 1158 2 1165 2 1172 2 1179 2 1186 2
1193 2 1200 2 1207 2 1214 2 1221 2
;

data copper;
set test.copper_reg;

data temp;
set copper;
if id<10;

%macro makeup(sta=,lag=);
data temp;
set temp;
if id<&sta;

proc forecast data= temp lead=&lag out=pred(drop=_TYPE_ _Lead_);
var open high low close cash volumn open_int toa_vol tot_open_int;
id id;

data temp;
update copper pred;
by id;

data copper;
set temp;

%mend makeup;

data _null_ ;
set blanks;
call symput ("init", row);
```

```
call symput ("le",len);
call execute ('%makeup (sta=&init, lag=&le)');

data test.copper_auto_final;
set temp;
```

2. Single time series for open price of copper

```
options nodate nonumber formdlim="-";

PROC IMPORT OUT= WORK.COPPER_FINAL
            DATAFILE= "D:\Thesis\data\final data\COPPER_AUTO_FINAL.xls"
            DBMS=EXCEL2000 REPLACE ;
     SHEET="COPPER_AUTO_FINAL";
     GETNAMES=YES;
RUN;

data copper_final_TS;
set copper_final;
if id >849 and id<1214;

proc arima data=copper_final_TS;
identify var=open stationarity=(adf);
run;
quit;

proc transreg data=copper_final_TS;
model boxcox(open/lambda=-2 to 2 by 0.5)=identity(id);
output out=open_trans;
run;

proc arima data=copper_final_TS;
identify var=open(1) scan esacf stationarity=(adf);
estimate  p=4 q=1  method=ml noconstant;
outlier type=(ao ls) alpha=0.05 sigma=MSE maxnum=10;
forecast lead=10;
run;
```

3. Single time series for close price of copper

```
options nodate nonumber formdlim="-";

PROC IMPORT OUT= WORK.COPPER_FINAL
            DATAFILE= "D:\Thesis\data\final data\COPPER_AUTO_FINAL.xls"
            DBMS=EXCEL2000 REPLACE ;
     SHEET="COPPER_AUTO_FINAL";
     GETNAMES=YES;
RUN;

data copper_final_TS;
set copper_final;
if id>849 and id<1214;

proc arima data=copper_final_TS;
identify var=close stationarity=(adf);
run;quit;
```

```
proc transreg data=copper_final_TS;
model boxcox(close/lambda=-2 to 2 by 0.5)=identity(id);
output out=copper_close_trans;
run;

proc arima data=copper_final_TS;
identify var=close(1) scan esacf stationarity=(adf);
estimate p=(1,3,7,8)  method=ml noconstant;
outlier type=(ao ls) alpha=0.05 sigma=MSE maxnum=10;
forecast lead=10;
run;
```

4. Single time series for open price of aluminum

```
options nodate nonumber formdlim="-";

PROC IMPORT OUT= WORK.ALUM_FINAL
            DATAFILE= "D:\Thesis\data\final data\ALUM_AUTO_FINAL.xls"
            DBMS=EXCEL2000 REPLACE ;
     SHEET="ALUM_AUTO_FINAL";
     GETNAMES=YES;
RUN;

data ALUM_final_TS;
set ALUM_final;
sqrt_open=sqrt(open);
if id>749 and id<1214;

proc transreg data=ALUM_final_TS;
model boxcox(open/lambda=-2 to 2 by 0.5)=identity(id);
output out=open_trans;
run;

proc arima data=ALUM_final_TS;
identify var=sqrt_open stationarity=(adf);
run;quit;

proc arima data=ALUM_final_TS;
identify var=sqrt_open(1) scan esacf stationarity=(adf);
estimate p=(3,5,6,10) q=1 method=ml noconstant;
outlier type=(ao ls) alpha=0.05 sigma=MSE maxnum=10;
forecast lead=10 out=ALUM_fore;
run;
```

5. Single time series for close price of aluminum

```
options nodate nonumber formdlim="-";

PROC IMPORT OUT= WORK.ALUM_FINAL
            DATAFILE= "D:\Thesis\data\final data\ALUM_AUTO_FINAL.xls"
            DBMS=EXCEL2000 REPLACE ;
     SHEET="ALUM_AUTO_FINAL";
     GETNAMES=YES;
```

```
RUN;

data ALUM_final_TS;
set ALUM_final;
sqrt_close=sqrt(close);
if id>749 and id<1214;

proc transreg data=ALUM_final_TS;
model boxcox(close/lambda=-2 to 2 by 0.5)=identity(id);
output out=ALUM_close_trans;
run;

proc arima data=ALUM_final_TS;
identify var=sqrt_close stationarity=(adf);
run;quit;

proc arima data=ALUM_final_TS;
identify var=sqrt_close(1) scan esacf stationarity=(adf);
estimate p=(1,5,6) q=1 method=ml noconstant;
outlier type=(ao ls) alpha=0.05 sigma=MSE maxnum=10;
forecast lead=10;
run;
```

6. Multivariate time series for copper

```
options nodate nonumber formdlim="-";

PROC IMPORT OUT= WORK.COPPER_FINAL
            DATAFILE= "D:\Thesis\data\final data\COPPER_AUTO_FINAL.xls"
            DBMS=EXCEL2000 REPLACE ;
     SHEET="COPPER_AUTO_FINAL";
     GETNAMES=YES;
RUN;

data copper_final_TS ;
set copper_final;
drop delivery;
if id >849 and id<1214 ;

proc print data=copper_final_TS;run;

symbol1 i=join  v=dot height=0.4;
symbol2 i=join  v=dot height=0.4;

proc gplot data=copper_final_TS;
Title1 "Copper";
      plot open*id = 1 close*id=2/overlay ;
run;

proc varmax data=copper_final_TS;
      model open close  / p=5 noint   dftest
                        cointtest =(johansen);
run;

proc varmax data=copper_final_TS;
```

```
      model open close  /  p=5  noint dify(1) print=(parcoef pcorr pcancorr)
            lagmax=8 cointtest =(johansen) ECM=(rank=1);
output lead=10;
run;
```

7. Multivariate time series for aluminum

```
options nodate nonumber formdlim="-";

PROC IMPORT OUT= WORK.ALUM_FINAL
            DATAFILE= "D:\Thesis\data\final data\ALUM_AUTO_FINAL.xls"
            DBMS=EXCEL2000 REPLACE ;
     SHEET="ALUM_AUTO_FINAL";
     GETNAMES=YES;
RUN;

data alum_final_TS;
set alum_final;
drop delivery;
sqrt_open=sqrt(open);
sqrt_close=sqrt(close);
if id >749 and id<1214;

symbol1 i=join  v=dot height=0.4;
symbol2 i=join  v=dot height=0.4;

proc gplot data=alum_final_TS;
Title1 "Alum";
      plot sqrt_open*id = 1 sqrt_close*id=2/overlay ;
run;

proc varmax data=alum_final_TS;
      model sqrt_open sqrt_close  / noint dftest
                                    cointtest =(johansen);
run;

proc varmax data=alum_final_TS;
      model sqrt_open sqrt_close  /  p=4  noint dify(1) dftest print=(parcoef
pcorr pcancorr)  lagmax=8 cointtest =(johansen) ECM=(rank=1);
output lead=10;
run;
```

Appendix B: SAS Output

1. Outliers for open price of copper

Outlier Details

Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
40	Additive	3648.9	17.20	<.0001
35	Additive	3325.6	15.00	0.0001
74	Shift	-4873.1	14.62	0.0001
59	Shift	4414.8	12.50	0.0004
122	Shift	4298.4	12.28	0.0005
158	Shift	-4221.9	12.26	0.0005
70	Shift	4144.6	12.23	0.0005
220	Shift	-3774.4	10.49	0.0012
43	Shift	-3723.4	10.51	0.0012
31	Shift	3879.7	11.75	0.0006

2. Outliers for close price of copper

Outlier Details

Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
38	Additive	3029.1	13.28	0.0003
46	Additive	-3040.4	13.89	0.0002
59	Shift	4110.3	11.72	0.0006
31	Shift	3913.9	10.99	0.0009
20	Additive	2529.5	10.65	0.0011
348	Additive	-2446.6	10.27	0.0014
62	Additive	2380.4	10.00	0.0016
32	Shift	3236.7	8.45	0.0037
101	Additive	2127.0	8.41	0.0037
103	Additive	-2144.3	8.75	0.0031

3. Outliers for square root of open price of aluminum

Outlier Details

Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
135	Shift	4.71899	22.54	<.0001
41	Additive	3.32837	21.69	<.0001
141	Shift	-3.99694	17.83	<.0001
174	Additive	-2.76542	16.34	<.0001
257	Shift	-3.60010	15.60	<.0001
147	Additive	2.60509	15.55	<.0001
40	Shift	3.44187	15.27	<.0001
163	Shift	-3.42237	15.61	<.0001
58	Shift	-3.36260	15.59	<.0001
49	Additive	2.32504	14.20	0.0002

4. Outliers for square root of close price of aluminum

Outlier Details

Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
41	Additive	2.85898	25.73	<.0001
131	Shift	3.26378	15.97	<.0001
138	Additive	2.31883	18.56	<.0001
147	Additive	2.22732	17.84	<.0001
117	Shift	3.06065	15.76	<.0001
40	Shift	2.91151	14.76	0.0001
141	Shift	-2.87132	14.83	0.0001
316	Additive	-1.89896	14.90	0.0001
168	Shift	-2.68730	13.87	0.0002