**Milagros F. Malaya**

# FORECASTING AND UPDATING PRICES OF SELECTED COMMODITIES IN THE PHILIPPINES FUTURES MARKET

Professor Malaya is a full-time Math faculty member of the Business Management Department of the DLSU College of Business and Economics. She obtained her BS Physics-Math (magna cum laude) from the College of Holy Spirit and her MS in Statistics from the University of the Philippines.

ONE INVESTMENT that has made millionaires overnight is the futures commodities market. It is the proverbial pot of gold for those who have turned richer and the woeful tale of the many others who have lost their money in just the wink of an eye. The Philippines futures market is relatively new, being in existence for a little less than 12 years, with the local commodity exchange inaugurated only in the latter half of 1987. Its intricacies are quite unknown to local businessman, but somehow it is highly attractive to liquid investors, particularly the Chinese.

In contrast to a cash contract where cash is paid for the immediate delivery of a product, a futures contract is an agreement that involves the delivery of a specified amount of a stated commodity at a designated time in the future. The buyer simply pays a deposit or margin, about 5% - 10% of the full contract price. Speculation then becomes encouraging as there is a potential for large rates of return in a short period of time because of the substantial leverage involved, *i.e., controlling a large sum of money* with little cash.

However, it must be made clear that an investor can lose more than his original investment, because the latter is only a marginal deposit, but his losses (or winnings) are computed on the full amount of the contract; hence, the millionaire-to-pauper's tale.

Consequently, strategies and techniques to maximize profits and offset losses are continuously being refined. This paper then attempts to equip the investor with an alternative method to forecasting commodity prices by identifying the behavior pattern of three selected commodities (Tokyo Soy Bean, Hong Kong Soy Bean, and Tokyo Red Bean) using univariate time series analysis. (Data in weekly prices were provided by the Kingly Commodities Research Department.)

**Statistical Methodology: Time Series Analysis and the Box-Jenkins Methodology**

*Basic Terms and Concepts*

A time series is a set of ordered observations at equally spaced discrete time intervals. The mean of a time series is its over-all level, and is computed just like any other mean. The variance of the series is a measure of the dispersion of the time series observations about its mean.

We also make a distinction between a process and realization. An observed time series is a realization of some underlying process. Hence, a process is similar to a population in classical statistics, and a realization is analogous to a sample. This realization is then used to build a model, or a representation of the process expressed as an algebraic equation. The procedures for this model-building are referred to as time series analysis.

A trend is a motion of the series in a specific direction, usually upward or downward. It describes any systematic change in the level of the time series. When a time series does not follow a trend, then it may drift. The real difference that lies between trend and drift is that trend is deterministic (or a fixed function of time), while drift is stochastic, meaning that future values vary in a probabilistic manner. Many time series are a combination of both, which are incorporated as components in the ARIMA models.

ARIMA is the acronym for Auto-Regressive Integrated Moving Average, the family of linear models which are the bases of the Box-Jenkins methodology of model-building for a time process.

---

* This is an abstract of the author's Master's thesis presented to the U.P. Statistical Center, and which discussed in greater detail the ARIMA models and the Box-Jenkins methodology. The statistical theory has been simplified and some of the results have been highlighted.

*Model Equations*

A discrete linear stochastic model is described by the equation

$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (1)

where

$Y_t$ = an observation of a time series taken at time interval t

$\varepsilon_t$ = random error component possessing white noise properties

$E(\varepsilon_t) = 0 ; E(\varepsilon_t^2) . \varepsilon^2$

$E(\varepsilon_t \cdot s) = 0 , t \neq s$

$\mu$ = constant

$\theta_1 \ldots \theta_q$ = parameters

or equivalently (proof shown on thesis) as

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \delta + \delta_t$ (2)

where $Y_t \ldots Y_{t-p}$ = successive values of a time series

$\phi_1 \ldots \phi_p$ = parameters

$\delta$ = constant

A time series model with the form of (1) is referred to as a moving average model of order q, or MA (q), while model equation (2) is called an autoregressive model of order p, or AR(p). This time series is said to be stationary if its statistical properties do not change over time. To fulfill this, certain conditions are imposed on the parameters of the model. For MA models, these are the *stationarity* conditions, and for AR models, these are the invertibility conditions.

Certain time series may combine both of the above models, called the AR-MA(p,q) or

$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$ (3)

To simplify (3), the backshift notation may be used, which is defined as

$B^j Y_t = Y_{t-j}$

Hence, (3) can be rewritten as

$\phi p (B) Y_t = \theta_o + \theta_q (B) \varepsilon t$

where $\phi_p (B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^t)$

$\theta_q (B) = (1 - \theta_1 B - \phi_2 B^2 - \ldots - \theta_q B^q)$

If a time series is nonstationary, this may be due to: (a) changing level of the series (no constant mean); (b) changing variability; or both (a) and (b).

One way to transform a nonstationary time series due to (b) to stationary values is to take the logarithm of these values. If nonstationarity is due to (a), the difference operator is $\nabla$ used, where

$\nabla^d = (1 - B)^d$

for $d = 1,\ \nabla Y_t = (1 - B) Y_t = Y_t - Y_{t-1}$

$d = 2,\ \nabla Y_t = (1 - B)^2 Y_t = Y_t - 2 Y_{t-1} + Y_{t-2}$

and the model equation is now called the ARIMA (p, d, q)

$\phi_p (B) \nabla^d Y_t = \theta_o + \theta_q (B) \varepsilon_T$

If a time series has seasonal variation, then the model equation of the multiplicative type can be expressed as

$\phi_p (B) \phi_p (B^s) (1 - b^s)^d \nabla^d Y_t = \theta_q (B) \theta_q (B^s) \varepsilon_t$

and is called the ARIMA (p, d, q) (P, D, Q)s where

(1) B is the backshift operator.
(2) d is the degree of regular differencing.
(3) D is the degree of seasonal differencing.
(4) $\theta_p$ (B) is the nonseasonal AR operator of order p.
(5) $\sigma_p$ (B) is the seasonal AR operator of order P.
(6) $\theta_q$ (B) is the nonseasonal MA operator of order q.
(7) $\theta$QB is the seasonal MA operator of order Q.
(8) s is the length of seasonality or periodicity.

*The Box Jenkins Methodology*

Statistical properties which describe the relationship between paired observations of a time series are the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

For a given time process, the autocorrelation is a measure of the correlation between each observation $Y_t$ and $Y_t$ lagged K time units forward in time.

The partial autocorrelation is a measure of the correlation between time series observations K time units apart after correlation at intermediate lags has been partialled out.

Each time series has a unique ACF and PACF and can be identified by graphing these properties. For AR models, the time series has an ACF which is characterized by exponential decay and a PACF which cuts off after lag p. Every MA series, on the other hand, has a PACF which decays and an ACF which cuts off after lag q. Table 1 presents a summary of the characteristics of the general and common nonseasonal models.

The Box-Jenkins (B-J) technique is then an experimental method used to identify the appropriate ARIMA model. It advises the use of at least 50 and preferably 100 observations for useful analysis. Basically, it consists of four steps: identification of the model, estimation of parameters, diagnostic checking and forecasting. Since the objective of the B-J technique is to obtain parsimonious models, i.e., models that adequately describe a time series and yet employ relatively few parameters, the selection process necessarily becomes iterative. Figure 1 shows the flowchart for this process.

It is also to be noted that over other estimating procedures, the appropriate ARIMA model produces optimal forecasts, i.e., the least forecast error variance among all linear models.

**Summary Of Results**

Applying the B-J methodology to Tokyo Soy Bean (TSB: January 1983 - October 1986), Hong Kong Soy Bean (HKSB: May 1983 - October 1986), and Tokyo Red Bean (TRB: January 1976 - January 1986), the following results were obtained:

1. The final model selected for the three bean products take the form of ARIMA (0,1,1) (1,0,0)s.
2. All the time series of the three different commodities do not possess a deterministic trend but mainly stochastic components. This explains why many shorter strategies used by commodity agents to aid investors in forecasting are not very successful. These are mainly based on trend lines which are deterministic while analysis reveals that the behavior of commodity prices is a random phenomenon which is a result of combined market forces and other conditions (demand and supply, weather conditions, import-export control, warehousing) which come from several countries which are the major suppliers and users of these commodities.
3. TSB (s = 21), HKSB (s = 19), and TRB (s = 6) all show a distinct seasonal pattern which are related to their delivery periods. Periodicity of 21 for TSB can be explained by its delivery period of six months, or 24 weeks. Speculators are then forced to unload their contracts before maturity date, or face warehousing problems with the delivery of the commodity, thus forcing prices to go down. For HKSB, the seasonality of 19 can be similarly related to its delivery period of also 24 weeks. It is earlier than the TSB series by two weeks, because actual delivery period for HKSB is 22 weeks, or two weeks before the end of the stipulated delivery month. For TRB, the seasonality of 6 (being a factor of 24) is still consistent with its delivery period of 24 weeks, but has been reduced due to the high volatility of price movements of TRB, a favorite commodity for all types of investors.
4. The ARIMA model equation finally selected produces the minimum mean square error forecasts at origin N,but variance of the forecasts increases as lead time departs from this origin. Percentage forecast errors computed on the three commodities vary from less than 1% to about 7% where values are forecast to a maximum of five weeks. Forecasting should therefore be limited to just two to three periods away from the origin.
5. The method of forecast updating leads to improved forecast prices for the commodities, with the computed reductions as low as 1/4 the original percentage error (For HKSB, % error was as low as 0.37%). Every time a new observation comes in, forecast origin is shifted to this period, resulting in reduced error variance. The new forecast value one period ahead is estimated, using the new observation. However, updating should not be used beyond two periods since this may lead to doubtful results. Instead, parameters of the model should be reestimated.

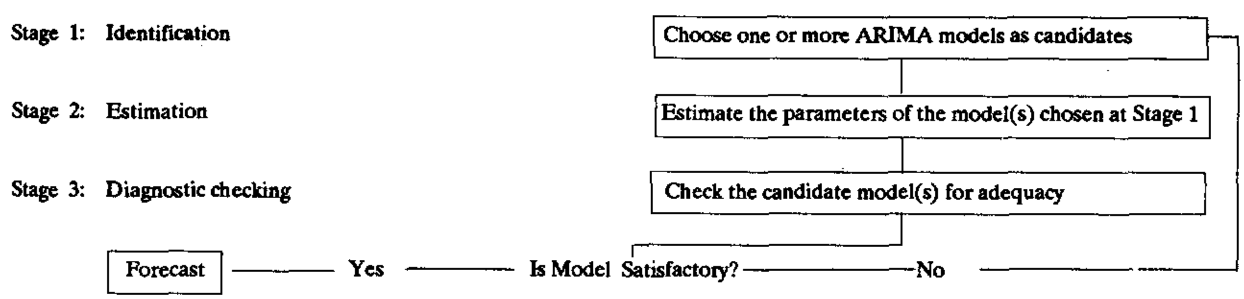


Figure 1. Stages in the Iterative Approach to Model Building as adapted from the B-J Method

## TABLE 1

### EQUATIONS AND CHARACTERISTICS OF THE GENERAL AND COMMON SPECIFIC NONSEASONAL MODELS

| | Model/Equation | Theoretical ACF | Theoretical PACF | Stationery Conditions | Invertibility Conditions |
|---|---|---|---|---|---|
| (1) | Moving Average Model of order q, (MA)(q)<br>$\bar{y}_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\ \varepsilon_t$ | cuts off after lag q | dies down | | |
| | (a) MA(1)<br>$\bar{y}_t = (1 - \theta_1 B)\ \varepsilon_t$ | cuts off after lag 1 | dies down | always stationery | $/\theta_1/ < 1$ |
| | (b) MA(2)<br>$\bar{y}_t = (1 - \theta_1 B - \theta_2 B^2)\ \varepsilon_t$ | cuts off after lag 2 | dies down | always stationery | $\theta_1 + \theta_2 < 1$<br>$\theta_2 - \theta_1 < 1$<br>$/\theta_2/ < 1$ |
| (2) | Autoregressive Model of order p, AR(p)<br>$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\ \bar{y}_t = \varepsilon_t$ | dies down | cuts off after laq p | | |
| | (a) AR(1)<br>$(1 - \phi_1 B)\ \bar{y}_t = \varepsilon_t$ | dies down | cuts off after lag 1 | $/\phi_1/ < 1$ | always invertible |
| | (b) AR(2)<br>$(1 - \phi_1 B - \phi_2 B^2)\ \bar{y}_t = \varepsilon_t$ | dies down | cuts off after lag 2 | $\phi_1 + \phi_2 < 1$<br>$\phi_2 - \phi_1 < 1$<br>$/\phi_1/ < 1$ | always invertible |
| (3) | Mixed AR-MA Model of order (p,q), AR-MA(p,q)<br>$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\ \bar{y}_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\ \varepsilon_t$ | dies down | dies down | | |
| | (a) AR-MA(1,1)<br>$(1 - \phi_1 B)\ \bar{Y}_t = (1 - \theta_1 B)\ \varepsilon_t$ | dies down | dies down | $/\phi_1/ < 1$ | $/\theta_1/ < 1$ |
| (4) | ARIMA Model of order (p,d,q)<br>$(1 - \phi_1 B - \ldots - \phi_p B^p)(1 - B)^d\ \bar{y}_t = (1 - \theta_1 B - \ldots \theta_q B)\ \varepsilon_t$ | | | | |

## BIBLIOGRAPHY

Bowerman, B. and R. O'Donnell. 1979. *Forecasting and Time Series*. PWS Publishers.

Box, George E. P. and Gwilyn M. Jenkins. 1976. *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc.

Cottman, John. 1981. *Time Series Analysis*. Cambridge: Cambridge University Press.

Jiler H., ed. 1980. *Forecasting Commodity Prices*. Com. Research Bureau Inc.

Kaufman, P.J. 1981. *Technical Analysis in Commodities*. New York: Piere.

McCleary, Richard and Richard Hay, Jr. 1980. *Applied Time Series Analysis for the Social Sciences*. London: Sage.

Nelson, Charles R. 1973. *Applied Time Series Analysis for Managerial Forecasting*. Holden-Day, Inc.

Pankratz, Alan. 1983. *Forecasting With Univariate* Box-Jenkins Models: Concepts and Cases. New York: Wiley.

Reilly, Frank K. 1982. *Investments*. Chicago: Dryden Press.