Forecasting Commodity Prices Using ARIMA

by ERIC WEISS, Ph.D.

Back in December 1981, J. Louis Anon wrote an article in Commodities called "Catch Short Term Profits using ARIMA" and touched off a minor revolution in forecasting commodity and stock option prices with his claims that ARIMA could forecast prices within 1.5 percent.

I had written an ARIMA package for statisticians who have an Apple II (or III), which was the only one for a microcomputer. Anon learned of this and referred customers who did not want to use a large mainframe computer (such as the one you can buy time on from his employer, Chase Econometrics) to me. Talking with potential customers, it became clear that the Anon article failed to answer many of their questions. It also became obvious that the original package had to be modified for traders to be able to use easily.

WHAT ARIMA DOES

ARIMA is a forecasting technique which manages to catch trends and changes in trends very well. In other articles in this series, I shall discuss some trading techniques which might be followed to capitalize on these forecasts. To understand ARIMA, it is best to start talking about regressions or curve fitting.

A regression is nothing more than trying to find the equation which best predicts prices. A simple regression would be to use yesterday's price to try to predict today's (whether high, low, or close does not matter now). To do this, we need to look not just at one "today" and one "yesterday", but a whole bunch of them. In other words, we will use Monday's price to predict Tuesday's, Tuesday's to predict Wednesday's and so on. Even better would be to use many weeks' prices (use Friday to predict Monday).

An even fancier regression would use not just the day before's price, but also two days earlier's.

A really sophisticated approach would include other things which are thought to effect the price. Yesterday's silver price might influence today's price of gold. July's contract may help to forecast December's.

Many times we are not sure if silver has an effect on gold, so there are statistical tests which tell us the odds are that there is an impact. For example, there might be only a five percent chance that silver influences gold. If there is no impact, then the factor should be removed to clean things up and to prevent it from causing erroneous forecasts in the future.

In the case where only past day's prices of the same contract are used, this is sometimes called Auto-Regression and provides the AR in ARIMA.

The MA or Moving Average part is a bit misleading since the weights do not have to be equal or even add up to one as is necessary for an average. But the terminology sticks despite this problem so we have to use it.

Say that we have done an autoregression on a gold contract using a Compu Trac file with 100 trading days of data. Since this is all in the past, it is easy for the computer to compare the predictions with the actual prices for each day. The difference is the prediction error, or just error for short.

Now comes a part that gives ARIMA a lot of its power to predict contract prices: let's do a regression to predict the errors using earlier errors! If we can predict the error that we just made in predicting a price, then it is a simple matter to correct for this and come up with an even better prediction.

So we have a series of numbers to forecast using past prices, and another series to correct the forecast using past errors. Not bad for the first pass through the data. ARIMA allows further improvement: it does some calculations to refine the numbers and repeats all of this again. If this is more accurate, the process continues until there is very little improvement in the accuracy of the forecast or the numbers change very little.

So far, the model is ARMA, not ARIMA. The I stands for integrated. This involves a little, but just a little, calculus, so bear with me. There are two major areas of calculus: integral and differential. Basically differential calculus takes an equation and looks at the effect of small changes in the variables on the total value. This is called differentiation. Integration is the reverse of this. Now sometimes there is an overall trend to the series of prices. In this case, ARIMA removes it first to clear the air. It can always be added back in. The easiest way to remove this trend is to look at differences between each day's prices and forecast these differences instead of the original prices. The nice part about ARIMA is that the ARIMA program converts back to prices from differences. The opposite of differencing is integrating, hence the I:

THE USER'S PROBLEM

There is a lot of math in all the calculations described earlier, but math is what your Apple does much more easily than you or I could. What a person does better than the computer is supply judgment. It is a basic principal of the ARIMA approach as developed by Box and Jenkins to keep the model as simple as possible. So the question becomes how many past days to build into the equation and how many past errors to work in. Too few, valuable information will be discarded. Too many, and erroneous or irrelevant information will be used in the forecast. Either one will lead to unpredictable errors in forecasting the future.

Now if we difference the prices when there is no need to, valuable information is stirred up, confused and lost.

THE NEXT ARTICLE

The next article in this series will look at the rules of thumb which statisticians use in determining how many of each type of parameter to use and when to difference.

