

A Cross Correlation-Based Stock Forecasting Model

By: Sungil Kim, Michael E. Baginski

Researchers are continuously seeking to develop and improve stock forecasting models by analyzing the past value of a company and predicting future performance based on past data trends. Prior literature on stock analysis focuses heavily on forecasting a single stock price based on its own past data. This type of analysis is susceptible to stock market volatility and not very effective for intraday trading; in addition, it is difficult to apply a method used for a particular stock market sector to other market sectors. In this study, we present a cross-correlation-based forecasting model using sets of closely related stocks to forecast future stock performance.

Cross-correlating two stocks works as follows: When the price of stock A is related to the price of stock B but there is a time delay of K days, predicting stock B's price based on stock A's price will reflect the future performance of stock A, K days earlier. For highly correlated pairs, the two stocks are assumed to exhibit a similar pattern in the short term. For a long-term investment, an algorithm must be run continuously that "buys" stock B whenever stock A shows a marked increase in price if the correlation is strong and delayed by K days. The algorithm will also include a sell price once the order is filled that reflects the expected increase in value of stock B.

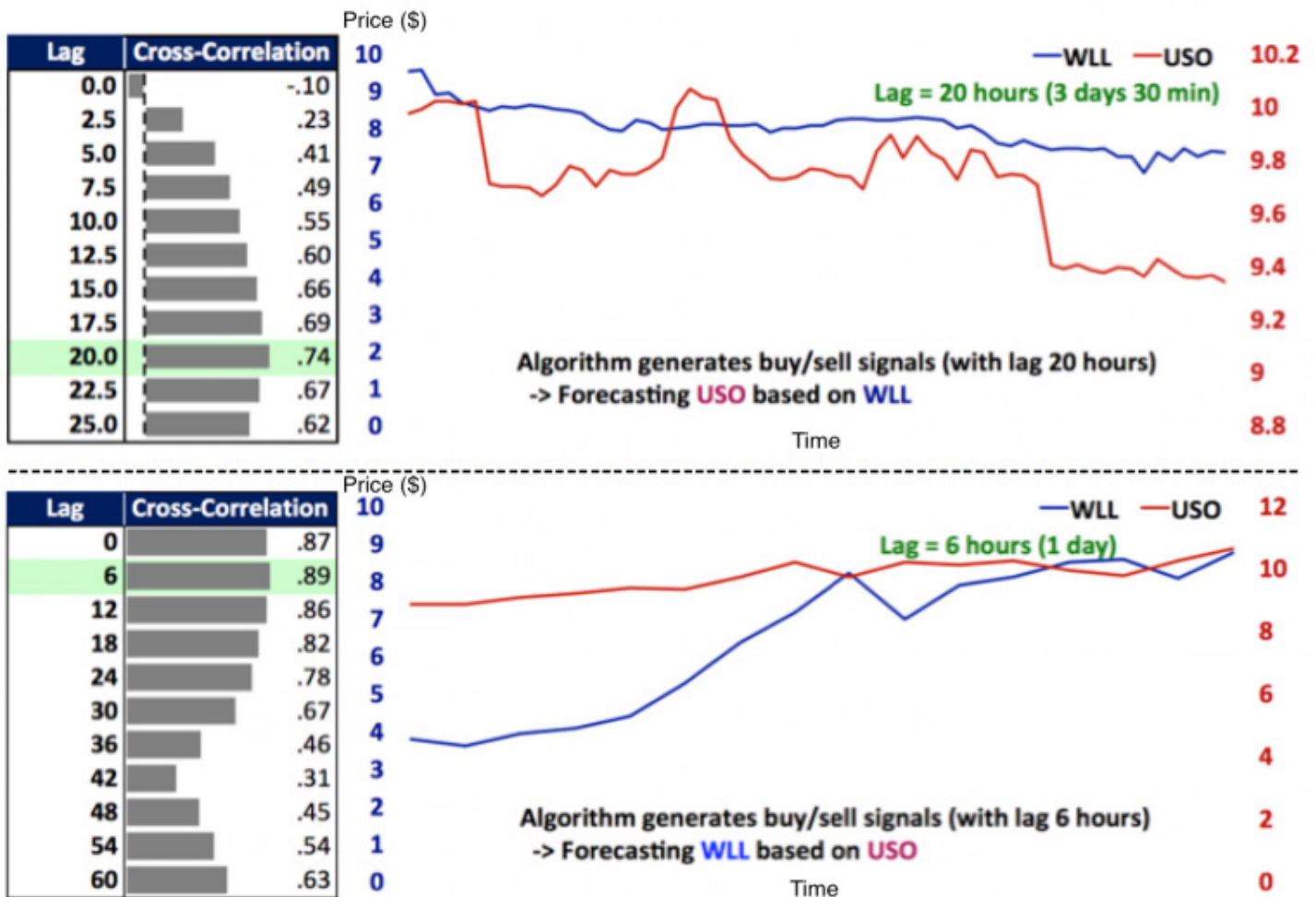


Figure 1. The forecasting model that generates sell signal (top), buy signal (middle), and the summary of result (bottom).

The proposed forecasting model discussed generates buy and sell signals along with corresponding trade dates and takes the following inputs: a pair of stocks, range of dates, correlation coefficient (ρ) threshold, and maximum number of tries. The model first retrieves data from two stocks in a specified range of dates. Then, it calculates the cross-correlation and finds (1) whether the two stocks are strongly correlated ($\rho > \rho_{\text{threshold}}$) and (2) whether the time delay (lag) between the two stocks is not zero. After these two conditions are met, the forecasting model generates a buy or sell signal depending on the performance of one stock that influences the other stock with the lag. In case either of the two conditions fails, the algorithm either adjusts the range of dates or changes the pair of stocks when it reaches the maximum number of tries.

The accuracy of the developed model is measured using U.S. stocks from the energy sector, which is more volatile than the technology sector and other indexes (e.g., S&P 500). We chose the energy sector to measure the resistance to volatility, effectiveness and accuracy, and profit per dollar. In particular, we chose Whiting Petroleum Corporation (WLL) and United States Oil Fund (USO), used data from the previous seven years to compute the cross-correlation, and forecasted for 47 days. Results show that the proposed model accurately forecasts the upward trend 15 out of 17 times (88.2%) and the downward trend 26 out of 30 times (86.7%), for a total of 87.2% (Figure 1). Compared to a previous study¹ that uses a data-mining algorithm with lagged correlation with 67% accuracy, the proposed model is significantly more accurate. Furthermore, the proposed model generates 3.2% profit per dollar over the span of the 47-day forecast interval. This result shows that the developed forecasting model is ideal for high-risk, high-return investments. In addition, the model can be used for intraday trading for a pair of stocks with a lag of less than a day.

This research developed a cross-correlation-based forecasting model and demonstrated that a pair of strongly related stocks shows a similar trend in the near future, unless the lag time is too big (greater than several months). The proposed model provides new insight to researchers, investors, and individuals regarding how cross-correlation can improve the accuracy of forecasting highly volatile stocks.

Statement of Research Advisor:

During the last year, I have directed Sungil Kim's research in using cross-correlation information for classes of stocks to predict future stock performance. The method he developed does work, and the forecasting technique could easily be applied in other arenas such as weather prediction. He has done an excellent job and we are planning on publishing his results shortly.

—Michael E. Baginski, Electrical & Computer Engineering

References:

[1] C. Fonseka and L. Liyanage. "A Data Mining Algorithm to Analyse Stock Market Data using Lagged Correlation". In Information and Automation for Sustainability (ICIAFS), 2008 4th International Conference on, 2008.