

# Knowledge Discovery in Financial Data

James E. Gentle

Department of Computational and Data Sciences

George Mason University

jgentle@gmu.edu

## Abstract

Stock price data have for years been a source of intrigue for economists, traders, and the ordinary person on Main Street. Because humans are very good at pattern recognition, it is natural that they would identify various patterns in sequences of price data. The next step, of course, is to take a particular pattern as predictive of future price movements that correspond to the price movements that followed a similar pattern in the past. This type of investigation is called “technical analysis”, and has provided well-paid employment for thousands of analysts. Although careful analysis has generally discredited the predictive ability of technical analysis, this area of investigation will always retain some appeal for data miners who may identify ever more complex patterns. There are several objectives other than trying to predict future prices that motivate mining of financial data. One, of course, is just better to understand the behavior of this kind of data. An important objective is to be able to identify insider trading and/or market manipulation. All of the major markets have in place programs that attempt to identify unusual activity and alert market oversight agencies. In order for any such program to be effective, reliable models of “normal” price movements are needed, and then, of course, measures of departure from the normal model. There are many interesting stylized characteristics of financial data that provide challenges to any model: heavy tails, asymmetry in rates of return, asymmetry in lagged correlations, clustering of volatility, aggregational normality, quasi long range dependence, and finally, seasonality. An objective of searching for patterns in the data is to develop an adequate stochastic model of normal behavior. No model of price movements can ignore exogenous events, some that relate directly to the activities of the companies for which the stock represents ownership, and others that may have bearing on the overall economy. To relate such events to stock price movement requires, first of all, a reliable database of current information, and then effective means of mining the news items for pieces of relevant information. This area of text data mining and reconciliation with observed price data provides a multitude of challenges, which with the growing integration of global markets, will continue to grow in interest for several years.